

# UC Irvine

## UC Irvine Electronic Theses and Dissertations

### Title

Health State Estimation

### Permalink

<https://escholarship.org/uc/item/75x5j2qf>

### Author

Nag, Nitish

### Publication Date

2020

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,  
IRVINE

Health State Estimation

DISSERTATION

submitted in partial satisfaction of the requirements  
for the degree of

DOCTOR OF PHILOSOPHY

in Computer Science

by

Nitish Nag

Dissertation Committee:  
Professor Ramesh C. Jain, Chair  
Professor Michael J. Carey  
Professor Gopi Meenakshisundaram

2020



# DEDICATION

*To my inspiring advisor Ramesh Jain, exceptional parents Bishwajit and Sesha, my illuminating partner Tabya, and the shoulders of all the giants we stand upon.*

# TABLE OF CONTENTS

	Page
<b>LIST OF FIGURES</b>	<b>vii</b>
<b>LIST OF TABLES</b>	<b>xix</b>
<b>LIST OF ALGORITHMS</b>	<b>xxi</b>
<b>ACKNOWLEDGMENTS</b>	<b>xxii</b>
<b>CURRICULUM VITAE</b>	<b>xxiii</b>
<b>ABSTRACT OF THE DISSERTATION</b>	<b>xxvi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Overview of Health and Medicine . . . . .	7
1.2 Modern Trends in Health Sciences . . . . .	12
1.2.1 Personalized Health . . . . .	12
1.2.2 Predictive Health . . . . .	14
1.2.3 Precision Health . . . . .	15
1.2.4 Participatory Health . . . . .	19
1.2.5 Preventive Health . . . . .	19
1.2.6 Healthcare Systems . . . . .	20
1.2.7 Performance and Wellness . . . . .	21
1.3 Health State Estimation is Critical for Progress . . . . .	23
1.4 Contributions . . . . .	25
1.5 Thesis Outline . . . . .	28
<b>2 Understanding Health Computing</b>	<b>29</b>
2.1 Cybernetics: Inspiration for this Work . . . . .	29
2.2 Defining Health . . . . .	33
2.2.1 Biological . . . . .	39
2.2.2 Functional . . . . .	41
2.2.3 Environmental . . . . .	44
2.2.4 Systems . . . . .	46
2.2.5 Perceptual . . . . .	47
2.2.6 Integrating a Holistic Approach . . . . .	51

2.3	Design Requirements for Health State Estimation . . . . .	52
<b>3</b>	<b>Literature Review</b>	<b>59</b>
3.1	Health Data Acquisition . . . . .	60
3.1.1	Self Insight . . . . .	61
3.1.2	Digital Sensors . . . . .	64
3.1.3	External Data Sources . . . . .	76
3.2	State Estimation for Individual Health . . . . .	78
3.2.1	Quantified Approaches . . . . .	79
3.2.2	Classification Approaches . . . . .	80
3.2.3	Probabilistic Approaches . . . . .	83
3.2.4	Subjective Approaches . . . . .	86
3.3	Personal Human Health Models . . . . .	87
3.3.1	General Human Systems . . . . .	88
3.3.2	Sub-Population Fitting . . . . .	92
3.3.3	Static Individual Parameters . . . . .	93
3.3.4	Towards Dynamic Personal Models . . . . .	94
3.4	Artificial Intelligence and Graphs . . . . .	98
3.4.1	Graph Networks . . . . .	98
3.4.2	Causal Reasoning . . . . .	99
3.4.3	Knowledge Structures . . . . .	100
3.4.4	Systems Thinking . . . . .	105
3.4.5	Microservice Architecture . . . . .	107
3.4.6	Data-Driven Learning . . . . .	107
3.4.7	Robust Artificial Intelligence . . . . .	110
3.5	Unmet Challenges and Scope of Dissertation . . . . .	112
<b>4</b>	<b>Health State Estimation Framework</b>	<b>114</b>
4.1	Health State Estimation General Overview . . . . .	115
4.2	Graph Network Block Structure . . . . .	116
4.2.1	Nested Microservices . . . . .	118
4.2.2	Causal Knowledge Infusion . . . . .	119
4.2.3	Data-Driven Model Updates . . . . .	120
4.2.4	Combinatorial Generalization . . . . .	121
4.3	Semantic Laminae: Interactive Layers . . . . .	122
4.4	Utilities: The Dimensions of Health . . . . .	124
4.5	Biological Blocks: Natural Micro-Services . . . . .	127
4.6	Events: Inputs to Life . . . . .	128
4.7	Initializing the System . . . . .	133
4.8	State Update Method . . . . .	134
4.8.1	Interface Event Retrieval . . . . .	136
4.9	Individual Dynamic Model Updates . . . . .	137
4.10	Revisiting the Design Requirements . . . . .	139

<b>5</b>	<b>Instantiating Personal Health State Estimation</b>	<b>142</b>
5.1	Experimental Overview . . . . .	143
5.2	Methods . . . . .	144
5.2.1	Longitudinal . . . . .	144
5.2.2	N=1 . . . . .	145
5.2.3	Multi-Modal Fusion . . . . .	146
5.2.4	Living-Labs Implementation in the Real World . . . . .	146
5.3	Data . . . . .	147
5.3.1	Collection Tools . . . . .	147
5.3.2	Events . . . . .	153
5.3.3	Biology and Utility Direct Measurements . . . . .	159
5.4	HSE System Initialization . . . . .	162
5.4.1	Translating User Intent . . . . .	162
5.4.2	Retrieval of Relevant Utility State Space . . . . .	163
5.4.3	Populating Biological Blocks . . . . .	165
5.5	Health State Update . . . . .	173
5.5.1	Time and Event Based HSE Updates . . . . .	173
5.5.2	Observation Based HSE Updates . . . . .	177
5.5.3	Propagating Updates . . . . .	177
5.6	Personal Model Update . . . . .	190
5.6.1	Knowledge Update . . . . .	190
5.6.2	Parallel Observations . . . . .	194
5.6.3	Cross-Modal . . . . .	197
5.7	Other Visualizations of HSE . . . . .	201
5.7.1	Multi-Person View . . . . .	202
<b>6</b>	<b>Personal Health Navigation</b>	<b>206</b>
6.1	General Health State Space . . . . .	209
6.2	Personal Health State Space . . . . .	210
6.3	State Transition Network . . . . .	211
6.4	Knowledge Layers and Regions of Interest . . . . .	212
6.5	Current State Estimation . . . . .	213
6.6	Goal Decomposition . . . . .	214
6.6.1	Timescales . . . . .	215
6.6.2	Recurring Goals versus Achievement Goals . . . . .	216
6.6.3	Synergistic Goals versus Competing Goals . . . . .	216
6.6.4	Identifying User Motivations for Goals . . . . .	217
6.7	Guidance . . . . .	217
6.7.1	Routing . . . . .	219
6.7.2	Actions and Cybernetic Control . . . . .	220
6.7.3	Recommendation . . . . .	225

<b>7</b>	<b>Conclusions and Future Work</b>	<b>236</b>
7.1	Technical Challenges . . . . .	237
7.1.1	Building Rich Knowledge Atlases . . . . .	237
7.1.2	Hardware . . . . .	238
7.1.3	Computing . . . . .	239
7.1.4	Accuracy and Precision . . . . .	239
7.1.5	N=1 Statistics . . . . .	240
7.1.6	Event Mining . . . . .	240
7.2	Real-World Adoption . . . . .	241
7.2.1	Engagement . . . . .	241
7.2.2	User Autonomy, Authority, and Ownership . . . . .	242
7.2.3	Privacy . . . . .	243
7.2.4	Security . . . . .	243
7.2.5	Health Economics . . . . .	244
7.2.6	Standards and Interoperability . . . . .	244
7.2.7	Research Community . . . . .	245
7.3	Final Words . . . . .	246
	<b>Bibliography</b>	<b>247</b>
	<b>Appendices</b>	<b>271</b>
A	Abbreviations . . . . .	271
B	Resources . . . . .	274



# LIST OF FIGURES

	Page
1.1 <b>Crude Mortality Rates for All Causes.</b> In the latter half of the twentieth century, the non-infectious disease continues to be a significant cause of mortality, reflecting the need for new strategies in health sciences. . . . .	2
1.2 <b>The Health Continuum.</b> Attention to health status occurs when there is a detectable signal. Ancient civilizations could only use subjective symptoms and signs as an indicator of the health state. Today, we have episodic measurements of a few criteria to understand our health state, such as blood pressure or lipids on an annual basis. Beyond those measurements, we still primarily rely on symptoms as the primary signal. Both of these signals are sparse, and most individuals do not even collect symptom data. The next revolution is to have continuous sensing of the health state so we can always have attention to improving health through the means of ubiquitous computing and modeling at the individual scale. . . . .	5
1.3 <b>Death Rates.</b> This graph shows the death rates in the 20th Century United States [172]. . . . .	10
1.4 <b>Personalized Heart Models.</b> Patients clinical magnetic resonance imaging data constructs personalized three-dimensional computer models of post-infarction hearts. These are assessed for the propensity of each model to develop arrhythmia. In a proof-of-concept retrospective study, the virtual heart test significantly outperformed several existing clinical metrics in predicting future arrhythmic events. The non-invasive personalized virtual heart risk assessment is a form of highly specialized individual health state modelling[11].	14
1.5 <b>Precision in Biological Understanding.</b> Selective COX-2 inhibitors like celecoxib are a type of Non-Steroidal Anti Inflammatory Drugs (NSAID) that directly targets COX-2, an enzyme responsible for inflammation and pain. Targeting selectivity for COX-2 reduces the risk of peptic ulceration and is the main feature of the drug celecoxib. This type of precision arises from more in-depth biological understanding. . . . .	16
1.6 <b>Visualization of Contributions.</b> . . . . .	27

2.1	<b>Cyber-Physical Connection.</b> Cybernetics connects physical reality with computing information for control. In the real world, sensors measure various aspects of the world and translate these observations into data. The system must take the data and produce an estimate of the system state. Based on this computed estimation and a control goal, the system produces guidance for the real world. The actions taken in the real world may or may not be the same as the guidance due to disturbance, noise, or other factors. Thus, we take continuous measurements to see how the system state has been perturbed based on the last iteration of guidance. System components can be a human entity such as a doctor or a computing machine. . . . .	30
2.2	<b>Basic Cybernetic Loop.</b> Cybernetic control occurs through feedback loops. The loop begins with measurements from one or more sources, compiled into a computational representation of the system state. This estimated state is compared to the goal state. The delta between the goal and current state is used to give guidance to operate the controllers appropriately. These controllers partake in actions in the real-world to perturb the true system. Disturbances may also perturb the system from the environment or stochastic inputs. At the next time interval, this loop cycles again with a fresh set of measurements.	32
2.3	<b>Blood Pressure Characterization.</b> Doctors use a classification approach with threshold values to have predetermined categories of defining health. The nature of health operates on a continuum basis, as shown in red. The state of resting blood pressure for each individual is always changing over time dynamically. Regardless of the approach, the individual is unable to sense their state position at any time, highlighting the need for continuous sensing.	35
2.4	<b>Biological health is at the level of inanimate molecules interacting to provide the organism with the fundamental components of life.</b> Metabolic pathways provide the methods by which energy and materials can be synthesized to support these components. Here is an example of biological health at various abstraction levels. We zoom into the heart. There is the tissue of the heart, each heart cell (myocardocyte), the metabolism network in this cell, a particular central pathway called the Tri-Carboxylic Acid Cycle (TCA Cycle), Adenosine Tri-Phosphate (ATP) which is an energy currency in biological systems produced by the TCA Cycle, and the oxygen atom within this ATP molecule. . . . .	41
2.5	<b>Aristotle’s Four Causes.</b> Aristotle’s Four Causes refers to the various lenses we can understand why something exists. Aristotle argues that the final cause is by far the most important since without it, the other causes may not have a direct purpose to exist. . . . .	43
2.6	<b>Dunn’s Health Grid Modeling: External environment and its relationship to wellness.</b> Dunn’s model of health highlights the strong relationship between a favorable environment and achieving peak wellness. If a good environment or good health exist exclusively, one cannot achieve an optimal level of health. . . . .	46

2.7	<b>Maslow’s Hierarchy of Needs.</b> There are base physiologic and safety needs (food, shelter, security) as a fundamental requirement, psychological needs such as a sense of belonging and esteem, and finally needs for self-fulfillment such as reaching one’s full potential. . . . .	49
2.8	<b>Mind - Body Connection: How Stress Impacts Physiology.</b> How a person perceives their environment has a direct impact on their physiology. When an individual perceives stress, the evolutionary encoded response of increasing blood pressure occurs to deliver high-performance blood delivery for a “fight or flight” response. This can be both a positive or negative impact on the health state, depending on the context. For example, a skier negotiating a racecourse may have stress from turning at high speeds. This stress allows the muscles to function at high forces and serves as a positive acute stressor. Blood pressure returns to normal after the race. In contrast, a professor who is worried about getting funding approved may have this stress effect chronically rather than acutely. In this context, the stress response may damage parts of the vessels and organs. Thus stress is a highly context-dependent component of computing the health state. . . . .	50
3.1	<b>Sensor categories for acquiring health data.</b> . . . . .	60
3.2	<b>Multi-Omics.</b> “Oma” means mass in Latin. Each layer of omic data is a massive array of data parameters regarding the specific layer. The advent of omic data mechanisms allows for high-resolution understanding of the biological state at various abstraction levels. Figure adapted from [144]. . . . .	71
3.3	<b>Personal Exposome.</b> The exposome consists of both biological agents we are exposed to, and physical components such as light, sound, and chemicals. These exposures are related to a certain location and time. Figure adapted from [137]. . . . .	72
3.4	<b>Estimation Approaches are Interconnected.</b> . . . . .	78
3.5	<b>State Space for Quantified Health.</b> Here are four charts showing how hormonal fluctuation occurs during menstruation. At each moment, an ovulating female has a point on this state space they exist. Regions specific the limits to which the state can exist and what the state means [268]. . . . .	79
3.6	<b>Fletcher Curve.</b> Lung function is measured through a quantitative mean of FEV1, which is the Forced Expired Volume a person can exhale in 1 second. This parameter is then placed upon the map of which regions and thresholds are relevant for understanding health state. [259]. . . . .	82
3.7	<b>Layers of personal data modeling.</b> . . . . .	87
3.8	<b>Modeling the Heart.</b> This figure shows the left coronary circulation model subjected to constriction of one of the main branches leading to blocked blood flow [203]. . . . .	89
3.9	<b>Health Belief Model.</b> This model predicts the probability of a person taking action that would benefit their health. The model focuses on understanding what people believe in and where they see benefits in certain actions. [295]. .	90

3.10	<b>Behavior Models.</b> Rasmussen’s behavior model outlines behavior in three different levels: Skill-based behaviour represents human behaviour that takes place without conscious control. Rule-based behaviour involves task execution controlled by rules. In unfamiliar situations in which there are no rules or procedures, behavior is knowledge-based [288]. . . . .	91
3.11	<b>Retul Bike Fit Personal Model.</b> Individual measurements are collected and applied to a highly customized bike fit. . . . .	94
3.12	<b>Personal Regulome.</b> This consists of the complete set of switches for all genes within an individual. Here is a rendition of chromosomes 18-20 of an individual whereby certain events either turn on or turn off the gene expression of specific genes within the individual. This regulome has specificity down to the cell level and is a key pillar in developing dynamic individual models. . . .	96
3.13	<b>Workflow of Human Brain Network.</b> Connections can be made using both functional and material causes for establishing the graph network. . . .	99
3.14	<b>The Heart Network.</b> The merging of omics data and biological domain knowledge is used to create graph networks, highlighting causal relationships.	100
3.15	<b>Cytoscape.</b> Cytoscape is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating with gene expression profiles and other state data. [212]. . . . .	102
3.16	<b>Heterogeneous Networks in Biology.</b> Hetnets contain edges and nodes that have type specificity as part of their properties. Using these types as distinctions, various visualizations, queries, and algorithms can be applied. This particular HetNet maps relationships between compounds and diseases [117]. . . . .	102
3.17	<b>Multiscale Cardiology.</b> There are multiple layers of consideration for practicing precision cardiology including integration of knowledge, experiments, data, environment, and person’s lifestyle. Computing techniques to extract understanding of how actions can be best taken to help the cardiovascular state are still needed [138]. . . . .	104
3.18	<b>Systems Thinking.</b> By considering an entire ecosystem of components, system thinking aims to understand the behavior of complex entities. This approach includes understanding how sub-entities are connected and related, how effects can circularly impact the system, what emergent properties arise from the system as a whole through synthesizing the system into a single architecture. <i>Adapted from Leyla Acaroglu.</i> . . . . .	105
3.19	<b>Cardiovascular Function.</b> Blood flow is shown here moving through stocks and flows through the STELLA interface. This model can only focus on a single dynamic variable within these containers [233]. . . . .	106
3.20	<b>Neural Networks.</b> Given an input layer of data, the neural network learns patterns and features to match an output layer. . . . .	110

4.1	<b>Health State Estimation Flow.</b> Events drive biological changes in the user’s body. These biological components determine utility state dimension values. The combination of these utility dimensions provide the health state, akin to the GPS position. The user’s position is visualized to view the current state within a certain semantic context defined by the user intent. . . . .	116
4.2	<b>Graph Network Block Unit and Update Process.</b> Each GNB describes a directed multi-graph, including multiple edges in the same direction or opposite directions, as shown between node B and C. Edges can also be to a self node as shown for Node A. The properties characterizing the entire multi-graph are summarized as global attributes within U. Step 1 shows the initialization of a GNB. In the following steps, red highlights the updates. Step 2 shows the edge update. Step 3 shows the node update. Step 4 shows the global update, a visualization of algorithm 1. . . . .	117
4.3	<b>Nested Microservices.</b> The GNB structure global attributes give rise to a microservice. These microservices are structured in a nested manner that can interact with each other in complex networks. This allows for the relevant level of abstraction to communicate in any system that wishes use the GNB. <i>Figure adapted from Meyer et al. [179].</i> . . . . .	119
4.4	<b>Knowledge to Graph.</b> Here, we translate cardiac output equations to the graph structure. Cardiac output refers to the volume of blood that the heart can pump per minute. Once we have the relationship in graph format in conjunction with data streams from either cheap continuous sensors like a weight scale or watch, or a more costly sensor like an ultrasound, we can update the relationship to more accurately reflect the model of the individual.	120
4.5	<b>User Interface.</b> A user can browse through various domains of interest, or search for a specific domain of interest in order to populate the a lamina. . . .	123
4.6	<b>Semantic Laminae.</b> For the system to initialize, the user must specify what goals or domain of interest they have, which is referenced against a domain knowledge database (DB) curated by experts of various fields. These experts outline the relevant utility dimensions and have map layers that contain ROIs. These map layers may also contain the connections between regions and the layers of information on how to move between states in the region. These connections will be expanded upon further in Chapter 6. . . . .	124
4.7	<b>Basic Utility Taxonomy.</b> These nodes show the nested microservice architecture of the utility dimensions. The movement node is expanded to reveal the next layer of microservices. This layer can be further expanded to more detailed child nodes. This is a simple overview of the utility GNB at large. . .	125
4.8	<b>Nested Biological Microservices.</b> Human biology is represented as GNB nodes at various layers of abstraction within the framework. This figure translates the cardiovascular system knowledge layers in Figure 2.4 into the GNB structure. The four chosen abstraction layers for this work are the system level, organ level, tissue level, and molecular level. . . . .	127

4.9	<b>Major Factors that Change Health.</b> This figure displays the burden of health factors in the USA for the 20 years of 1990 to 2010 [185]. The top three items that change the health state of an individual in the United States are lifestyle events, followed by environmental and neurological events, validating the necessity of tracking events for estimating health state. . . . .	128
4.10	<b>Personal Events: Time Scale of Intervals.</b> The temporal scale of events that happen to a person have a large range that can span over 15 orders of magnitude. . . . .	130
4.11	<b>Event Processing Framework.</b> . . . . .	131
4.12	<b>Initialization Steps.</b> These are the four steps in initializing the graph structure for the individual. <b>1.</b> First, we query the user for their intent. <b>2.</b> Using this query, we retrieve the appropriate laminae they can visualize. <b>3.</b> These laminae require a certain set of utility dimensions to be displayed. These are populated next. In order to know the user position on these dimensions, the appropriate biological blocks need assessment. <b>4.</b> The fourth step populates these appropriate blocks to monitor. . . . .	134
4.13	<b>State Update Methods. 1.</b> The system initiates from the user intent. <b>2.</b> If there is no data from the event or stream DB reported, self-edges update nodes which then propagate along the directed edges to give new utility values. <b>3.</b> If interface events are detected from the event stream, they update the relevant biological states, which then propagate. <b>4.</b> If there are observed biological attributes directly from sensors, they update the respective biological nodes. <b>5.</b> If there are observed utility attributes directly from sensors, they update the respective utility nodes. <b>6.</b> The health state position is reported on the lamina of interest, and/or can be viewed as raw utility dimensions, and/or can be viewed as biological attributes. . . . .	136
4.14	<b>Learning Situations.</b> Relationships between nodes update in both inter-layer edges or intra-layer edges. <b>Situation A.</b> When new knowledge is given to the system, the appropriate edges or nodes are updated. <b>Situation B.</b> Event-mining can discover relationships between temporal events within the user personicle. These events can reflect an update to a biological node that models behavior or any other biological component. <b>Situation C.</b> When two nodes that are connected are both measured in parallel, the relationship between the two nodes can be characterized to match the incoming data. <b>Situation D.</b> When many source nodes are related to a target node, we can use all the nodes to learn how they impact the target node. . . . .	139
5.1	<b>Chapter 5 Experimental Overview.</b> The first part of Chapter 5 is initializing the system for the test subject in both breadth and depth. The second phase demonstrates the state estimation in action. The final set of experiments demonstrate three approaches to update the personal GNB model.	143

5.2	<b>Central Hardware Devices.</b> These are the three main central devices to which other sensors link via local wireless connections of either Bluetooth, WiFi, or ANT+ protocols. The devices themselves include their sensors. The notebook computer contains a microphone, front camera, and HCI capture software. The mobile phone includes a light sensor, microphone, accelerometer, gyroscope, barometer, front, and rear camera. The wearable watch includes an optical sensor, accelerometer, gyroscope, barometer, and GPS. . . .	149
5.3	<b>Activity Sensor Hardware.</b> . . . . .	149
5.4	<b>Biological Measurement Hardware.</b> . . . . .	149
5.5	<b>Manual Measurement Tools.</b> These tools were used to capture basic physical measurements. . . . .	150
5.6	<b>Clinical Hardware.</b> These machines are used as examples of reference values to compare with the accuracy of lower cost sensors. . . . .	150
5.7	<b>IoT Devices.</b> These sensors are used to monitor the local exposome of the individual including light, sound, temperature, humidity, and smoke. . . . .	150
5.8	<b>Custom Gas Sensor IoT Device.</b> This sensor IoT system was built with multiple gas sensors connected to an Arduino micro-controller that links to a Raspberry Pi computer. The computer sends the signal via a web Rest-API that allows for a minute by minute monitoring of the gases in the local environment. . . . .	151
5.9	<b>Exercise Event Understanding.</b> Single exercise event streams. . . . .	155
5.10	<b>Exercise Event Understanding.</b> Single exercise event summary with location data, heart rate zone charts, and power data zone charts. . . . .	156
5.11	<b>Public Geospatial Dataset Integration.</b> Many aspects of the environment, including ones that induce stress, can be elucidated from domain literature and publicly available GIS datasets. Related to pollution, we gather air, audio, and light as factors that impact the health state. Related to demographic information, we gather crime, income, and cardiac deaths to estimate socioeconomic stress [193]. . . . .	159
5.12	<b>Computer Vision for Biological Measurements.</b> Images can give insight into a person’s health state. In this figure, we use facial images for three purposes. First, we detect landmarks on the person’s face to identify their height-to-width ratio, lending insight into their baseline testosterone production. Second, we use color detection to see how the user’s skin color is changing based on each month. Skin color is classified using the Fitzpatrick type used routinely by dermatology physicians [232, 236]. We see on average that spring time, the user has lighter skin than the rest of the year. Third, we use edge detectors to characterize the wrinkling of the skin to understand collagen health.	161
5.13	<b>Cycling World Lamina.</b> This interface displays many different types of cycling of interest to the user. The user can select a specific type, zoom in and out, or view the whole lamina at once. . . . .	163

5.14	<b>Utilities of Human Health.</b> In this work, we have defined five basic clusters of human utility. Given the selection of the user, we focus on the utility blocks that are connected to power production. In this work, we focus on the endurance capacity of critical power production. These nodes are highlighted in red. . . . .	164
5.15	<b>Translating Regions-of-Interest to Utility State Space.</b> On the left, we see all different types of cycling that a user may be interested in participating in a cycling lamina. On the right, these types of cycling are overlaid on the utility dimensions of power production versus duration. This state space of power given duration is the CP curve. . . . .	165
5.16	<b>Populating Biological Blocks from Utility Blocks.</b> The CP utility state space (background) exists through 4 main biological systems (foreground). The first is the ATP-CP system (orange and yellow), the second is the anaerobic glycolysis (blue), the third is aerobic glycolysis (gray), and the fourth is lipolysis (red). . . . .	166
5.17	<b>Aerobic Glycolysis Domain Knowledge.</b> This figure shows the flow of oxygen (O <sub>2</sub> ) from the air to the muscles used in glucose oxidation, also illustrating how the nodes are connected in the GNB. The processes are well-characterized through the scientific literature, and in some instances, can be described mathematically. These processes are captured in the edges of the GNB. . . . .	167
5.18	<b>Nested Graph Network Blocks within Aerobic Glycolysis.</b> There are many biological components nested within aerobic glycolysis. We focus on the details of gene expression inside the left ventricle in this experiment set. These nodes are highlighted in red. . . . .	168
5.19	<b>Translating Domain Knowledge to Graphs.</b> Mitochondrial biogenesis pathways are taken from literature to populate the molecular GNB layers of the HSE system for our test subject. This GNB is stored as a node within the muscle tissue nodes. . . . .	170
5.20	<b>Cardiovascular Graph Network Block.</b> This network shows how events impact the cardiovascular state down to the gene expression level. The graph network of biology is taken from compiling open-source data from graph knowledge database Cytoscape and other literature. Consolidation of this into a unified graph was done by the author. Green arrows denote an activating effect on the target node. Red arrows denote an inhibitory effect of the source node to the target node. . . . .	171
5.21	<b>Graph Network Block Structure Summary.</b> The total overview of the 4 main system components instantiated for the test user. Red circles denote areas of focus for the experiments carried out in the next chapter. . . . .	172
5.22	<b>Exercise Interface Events.</b> For the years of 2018-2019, CTL is given in purple, GOVSS is given in green, HPA Load by day is given in red. These specific interface events are used to determine impact to the CRF health state.	175



5.23	<b>Circadian Rhythm of Interface Events.</b> This figure shows different interface events retrieved from the test subject data. Different days of the year are represented as concentric circles, and various sectors of the circle represent different times of the day. Events in a day are represented as colored arcs on that circle. Volume and pressure overload events that occur in the left ventricle are retrieved as inputs towards changing the health state. This figure also shows the environmental interface events. High PM2.5 intake events are recognized over one week. These are the instances where per minute PM2.5 intake was greater than 0.7 $\mu\text{g}$ . Low blood oxygen events are recognized over one year, where blood oxygen saturation goes below 95%. The highlighted sectors roughly represent the time between sunrise and sunset. Thus we can see how different events overlap with circadian patterns. . . . .	175
5.24	<b>Interface Event Summary.</b> Total Exercise impact, sleep quality impact and stress impact by day for years 2018-2019 in the test subject. Exercise events are shown in quartiles showing periods of higher impact in the upper quartile, and minimal impact in the lower quartile. High variance within sleep and stress occurs daily. . . . .	176
5.25	<b>GNB Update Cycle.</b> This illustration shows the time-lapse of steps within the update cycle for a GNB system. Step 1 colors the layers of the HSE GNB structure from the initialization. Step 2 shows how the value updates are given for inputs (green) consisting of an exercise CTL of 40, age of 30, weight of 58kg, and resting heart rate of 48. The estimated nodes are blue. . . . .	178
5.26	<b>Single-Dimension Health State.</b> Here is a single dimension of one hour CP value for the subject over the course of time. . . . .	179
5.27	<b>Multi-Dimension Health State.</b> Here we see all 18,000 utilities of CP at six month intervals. . . . .	180
5.28	<b>Location Impact on Health State.</b> Here we can see how four different pollutants impact a single bike ride activity. Blue is low exposure, red is moderate exposure, and yellow signifies high exposure to the particular pollutant during the unique time-points and location during the bike ride. The EPA defines these levels. The levels can be used to inform the user, health experts, civil engineering, and recommendation systems of health state impact to aid in planning. . . . .	182
5.29	<b>Interface Events for Physiological and Pathological Changes.</b> This figure demonstrates the difference in response between physiological and pathological heart changes resulting from various types of events. Volume overload is when the heart needs to pump at a high volume flow rate. This causes physiological hypertrophy which strengthens the stroke volume of the heart. Pressure overload causes the heart to push against higher forces which increases wall thickness, therefore reducing stroke volume and reducing cardiovascular function. . . . .	184
5.30	<b>Daily changes in Cardiovascular Biological Health.</b> This figure demonstrates the balance between healthy and pathological heart function on a day-by-day basis computed through the graph structure in Figure 5.32. . . . .	185

5.31	<b>Daily Changes in Left Ventricular Stroke Volume.</b> This figure demonstrates the balance between healthy and pathological heart function on a day-by-day basis computed through the graph structure in Figure 5.32. . . . .	186
5.32	<b>Health State Graph Updates.</b> This figure demonstrates the estimated signaling pathway nodes for cardiovascular health change at the molecular gene expression level. An average for each month is shown in the year 2019. Legend is given in Figure 5.20. . . . .	188
5.33	<b>Updating Model of Event to Biological Node.</b> This figure demonstrates the difference in the model between the original knowledge-based edge weights and after the knowledge update from the literature. High Intensity (HI) exercise was thought to be the primary causal pathway to increased mitochondrial biogenesis in the muscle and heart. New literature shows that Low Intensity (LI) exercise does this much more than HI exercise. Using this update, the edge weights on the right are changed. . . . .	191
5.34	<b>Genetic Update of Model of Event to Biological Node.</b> This figure demonstrates the difference in the model between using the most updated general knowledge and knowledge, specifically about the user’s genetics. Because this user has two different types of ACTN3 gene, the impact of both HI Exercise and LI Exercise is higher. . . . .	192
5.35	<b>Parallel Data Model Learning.</b> The above two graphs show heart rate data related to power data in two different activities. During activity A, we can see three distinct linear bands of relationships in red, magenta, and green. These clusters are separated due to a phenomenon called aerobic decoupling. We can check the causal agent of these multi clusters in the ride data by referencing the aerobic decoupling node. In ride A, it is 40.3 percent, versus Ride B is 0.9 percent. For this reason, the graph system can causally understand that activity B is a better sample. This allows for learning the relationship with higher quality. . . . .	196
5.36	<b>Heart Rate and Power Relationship.</b> This box plot shows the relationship between heart rate and power for the test subject from various instances within a one month period. . . . .	196
5.37	<b>Memory Parameter Optimization.</b> For each of the three sensor modes, we tested the model to see if there was an optimal amount of days to use as a memory for the CRF prediction. Any model memory of 10-100 days gave similarly accurate learning results, demonstrating there was no optimal learning memory for the model. . . . .	199
5.38	<b>Cross-Modal Sensor Fusion.</b> By using multiple nodes in our learning model, we can see a direct improvement in the CRF utility estimation, which shows the value of having a multi-modal sensor fusion approach. . . . .	200
5.39	<b>Personal Mobile Dashboard View.</b> The personal mobile dashboard view is a concept design mock-up of how a user may access their health state. Screen 1 we show how health state can fuel recommendations. Inspired by jet engine dashboards, screen 2 can give a live snapshot of the health state. Screen 3 provides a comprehensive list of all bio-variables and health states being tracked, with a ranking system to provide relevant results at the top. . .	202

5.40	<b>Patient Dashboard View.</b> The patient dashboard view is a heat map of bio-variables and summary scores that affect each subject. Variables are normalized to a specific color for each box. The variables are graded on a scale ranging from blue being most healthy and red being least healthy. This type of visualization integrates cross-modal data in a manner that a clinician, hospital, public health agency, or any domain expert can use to monitor the health of a patient panel or population. Clicking on a particular box would pull up further insights and details. This highlighted box in this image shows the noise pollution insight for patient 22. . . . .	204
5.41	<b>Total Health State Estimation System.</b> Here we can see all the blocks of Chapter 5 in a system overview. . . . .	205
6.1	<b>Navigation Vision for Health.</b> Navigation for health has four key components: Measure, Estimate, Guide, Action. Step 1: Take measurements in the real world through a variety of sensors. Step 2: Additional measurements may be needed if they are not sufficient for a reasonable estimate. Step 3: From these measurements, we can produce an Estimate of the current health state. Step 4: Guide the user based on a reference goal and the user’s model. Step 5: Take actions in the real world through a variety of mechanisms. Step 6: Assisted mechanisms may contribute to the executing of these actions. Step 7: Retake measurements to understand how the person’s health state has changed, and this updates the personal model perpetually. . . . .	208
6.2	<b>Flight Envelope.</b> This graph shows the area of total possible states in which the plane is able to fly given its unique parameters. . . . .	210
6.3	<b>General Health State Space.</b> The state space includes all the possible states that all humans can exist, with several dimensions. In order to interact with this state space effectively, a user may provide a particular intent, which can reduce the dimensionality to the relevant regions to visualize the general health state space. . . . .	211
6.4	<b>Personal Health State Space.</b> The PHSS includes all the possible states that a specific human can exist at the current moment and at future time points. This state space is a subset of the general health state space. To convert this space to a computationally efficient world, we may apply a discrete layer on top of this state space, allowing for graph-based approaches for further applications. . . . .	212
6.5	<b>Connected State Space Network.</b> The state space network includes all the possible transitions from one state to another for the specific individual. The connections contain the information of what inputs the matrix will cause a particular state to convert to the next state. The edges here are directed (although not visually shown). These connections are not static and are re-populated, given a reflection of the current situation. . . . .	213

6.6	<b>Knowledge Layers.</b> Layering information on the state space allows for semantic interactions. Domain knowledge specifies these layers from the relevant database. We can gather specific regions of interest from knowledge sources that can label the state space so a user can understand the possibilities for their health in natural language or a relevant visualization. In this example, specific regions of interest are labeled with the colors of green, orange, yellow, and red. Red is an example of an undesirable region, while blue represents regions that are not within the person state space possibilities of the user. The black boxes represent points of interest, which is essentially a more confined region of interest. . . . .	214
6.7	<b>Goal Selection and Route Planning.</b> Given a particular interaction layer, a user can select a goal of choice. We compare this goal to the current state of the user in order to determine routes that can lead to the goal state. Multiple routes may be generated based on optimization criteria that the user or the system can weigh. . . . .	219
6.8	<b>Control and Re-Planning.</b> Once a specific route is selected, the user and system only need to worry about controlling the state transition for the current time iteration. Cybernetic loops are employed to ensure that the state transition is controlled successfully in the presence of disturbances and noise. Once the state has transitioned to a new state, the system re-plans the most optimal route again. Due to changes in resources, health state, or other factors, the route may be different at the next time iteration. Here we see route Y selected at time $k$ , and at the next time iteration, the route plan has changed to Y.2 based on the updated state and resources at time $k+1$ . . . . .	222
6.9	<b>The Long Tail Problem.</b> Usually users only see items that are top rated by other users. This leaves items that are highly relevant for the user far down in the search results, such as items that are healthy. . . . .	230

# LIST OF TABLES

	Page
2.1 <b>Literature Review of Health Definitions.</b> Throughout health research history, various attempts have been made to define health. Here is an overview of these attempts. . . . .	37
3.1 <b>Exposome tools for measurement.</b> Figure adapted from [294]. . . . .	74
4.1 <b>Joint Utility Table.</b> This table is a sample list of human joint utilities. For each utility, there is a set of output attributes that determine the utility value, in this case “Range of Motion.” . . . . .	126
4.2 <b>Event Schema.</b> Event schema uses an interval approach to define the event and associated parameters. Data streams are split into two schemas, one for general data streams, and one specifically for a location that includes a geospatial value of latitude, longitude, and altitude stored as “LOCATION.” .	132
5.1 <b>Quasi Experiments.</b> Quasi experiments cannot definitively establish causality, but provide evidence towards a related mechanism (correlation or causal). Experiments establish proof through demonstration [46]. . . . .	145
5.2 <b>Hardware Sensors.</b> . . . . .	152
5.3 <b>Software Sensors.</b> Software sensor systems used to track the user’s neurological state and lifestyle events. . . . .	153
5.4 <b>Sensor Streams.</b> Sensor streams captured at 1 second intervals during exercise activity. . . . .	157
5.5 <b>Exposome Sensors.</b> There are two forms of exposome data collection within this experiment. The first is GIS matched data, which uses the user’s location to retrieve the nearest public sensors as an estimate of their local environment. The second set of data uses hyper-local in-room sensors to capture data about the immediate vicinity of the user’s environment. . . . .	160
5.6 <b>Organ Global Attributes.</b> Global attributes of major organ system GNBs	167
5.7 <b>Genetics SNP.</b> Genetics SNP profile of subject related to the CRF GNB. . .	192
5.8 <b>Slope optimization of VAM models.</b> . . . . .	200
5.9 <b>VAM Parameter Optimization.</b> For the VAM based model, there is a beneficial effect to choose samples in which the slope gradient of the activity is steeper, which quantifies how much weight to assign to data samples through the quality metric of slope in our learning model. . . . .	200

A.1	Abbreviations . . . . .	271
B.2	Hardware, Software, Knowledge Sources, Data Tools, and Visualization Resources . . . . .	275

# LIST OF ALGORITHMS

	Page
1 Graph Network Block Update . . . . .	121
2 Aerobic Decoupling . . . . .	195

# ACKNOWLEDGMENTS

Professor Ramesh Jain has been the single most influential figure in my academic development, not just as a graduate student, but in my ability to think about life in its entirety. His aspirations of changing the world in a meaningful way have been vital in the development of my thought processes. We both share this same vibration together and thus have been exceptionally synergistic in our dream of making society a truly healthier and happier place. Our relationship has been multi-faceted in persevering towards this aim. I am jovial to know that will continue in the coming years. Ramesh's approach to always have fun, love what you do, and fuel the best intentions reverberate in my mind daily. I aspire to pass forward his level of enthusiasm and optimism to future generations beyond my own life. Through challenges and breakthroughs, Ramesh has always been a genuine best friend and trusted mentor. Every moment I was able to spend with him since the first day I walked into his office is treasured. I look forward to continuing the growth in our relationship long into the future and thank him for his unwavering belief in me.

Next, I would like to thank my thesis committee members, Professor Gopi Meenakshisundaram and Professor Michael J. Carey. Professor Meenakshisundaram has been thoughtful and supportive in shaping this work towards helping humanity. Professor Carey has equally been thoughtful in questioning the technical approach to make this work polished and have greater clarity.

I want to thank my lab mates, Vaibhav Pandey, Preston Putzel, Hari Bhimaraju, Siripen Pongpaichet, Hyungik Oh, Mengfan Tang, and Pranav Agrawal for the stimulating discussions and their support of my work. Additionally, I would like to thank Srikanth Krishnan, Daniel Azzam, Matthew Wade, Mars Wang, and Abhisar Sharma for their collaboration towards advancing ideas in this thesis. I want to extend my appreciation for Vaibhav's close cooperation in a large portion of the research progress presented, and for his companionship as a labmate.

My parents and brother have been the backbone of my upbringing through their unwavering support. Most importantly, they have always given me the freedom and means to pursue my passions freely and to the best of my ability. I hope to embody their virtues as I continue in life. My partner Tabya Sultan has been one of the most significant influences in my total lifespan. She has shaped much of the thoughts in this work. Her strength in believing in my potential is unrivaled, and I kindle this strength in my pursuit of excellence.

Finally, I want to extend my gratefulness for everyone who has impacted me in my life journey to this point. Each one of these people will always be a part of shaping who I am and will be. Thank you.



# CURRICULUM VITAE

Nitish Nag

## EDUCATION

<b>Doctor of Medicine</b> University of California, Irvine (UCI) <i>*expected</i>	<b>2023*</b> <i>Irvine, CA</i>
<b>Doctor of Philosophy in Computer Science</b> University of California, Irvine (UCI)	<b>2020</b> <i>Irvine, CA</i>
<b>Private Pilot</b> California Airways	<b>2013</b> <i>Hayward, CA</i>
<b>Bachelor of Science in Nutrition Physiology and Metabolism</b> University of California, Berkeley (UCB)	<b>2011</b> <i>Berkeley, CA</i>
<b>Bachelor of Arts in Integrative Biology</b> University of California, Berkeley (UCB)	<b>2011</b> <i>Berkeley, CA</i>
<b>Bachelor of Arts in Molecular and Cell Biology: Biochemistry</b> University of California, Berkeley (UCB)	<b>2011</b> <i>Berkeley, CA</i>

## RESEARCH EXPERIENCE

<b>Graduate Student Researcher</b> University of California, Irvine (UCI) <i>*expected</i>	<b>2015–2020</b> <i>Irvine, CA</i>
<b>Scientist</b> Renovel Discoveries Inc.	<b>2011–2013</b> <i>Fremont, CA</i>
<b>Undergraduate Researcher</b> UCB, <i>Hellerstein Metabolism Lab</i>	<b>2009–2011</b> <i>Berkeley, CA</i>
<b>Undergraduate Researcher</b> UCB, <i>Brooks Exercise Physiology Lab</i>	<b>2009–2011</b> <i>Berkeley, CA</i>
<b>Undergraduate Researcher</b> UCB, <i>Sul Nutrition Lab</i>	<b>2007–2009</b> <i>Berkeley, CA</i>
<b>Research Assistant</b> Bexel Pharmaceuticals Inc.	<b>2004–2007</b> <i>Union City, CA</i>

## PUBLICATIONS

**A Novel Epidemiological Approach to Geographically Mapping Dry Eye Disease in the United States through Google Trends** 2020

Daniel Azzam, Nitish Nag, Julia Tran, Lauren Chen, Kaajal Visnagra, Matthew Wade, American Society of Cataract and Refractive Surgery

**Continuous Health Interface Event Retrieval** 2020

Vaibhav Pandey, Nitish Nag, Ramesh Jain, Proceedings of the 2020 ACM on International Conference on Multimedia Retrieval

**A Navigational Approach to Health: Actionable Guidance for Improved Quality of Life** 2019

Nitish Nag, Ramesh Jain, IEEE Computer

**Flavour Enhanced Food Recommendation** 2019

Nitish Nag, Aditya Narendra Rao, Akash Kulhalli, Kushal Samir Mehta, Nishant Bhat-tacharya, Pratul Ramkumar, Aditya Bharadwaj, Dinkar Sitaram, Ramesh Jain, Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management

**HealthMedia'19: 4th International Workshop on Multimedia for Personal Health and Health Care** 2019

Susanne Boll, Jeannie S Lee, Jochen Meyer, Nitish Nag, Noel E O'Connor, Proceedings of the 27th ACM International Conference on Multimedia

**Synchronizing Geospatial Information for Personalized Health Monitoring** 2019

Nitish Nag, Vaibhav Pandey, Likhita Navali, Prateek Mohan, Ramesh Jain, arXiv preprint arXiv:1907.10594

**Cross-Modal Health State Estimation** 2018

Nitish Nag, Vaibhav Pandey, Preston J Putzel, Hari Bhimaraju, Srikanth Krishnan, Ramesh Jain, 2018 ACM Multimedia Conference on Multimedia Conference

**Ubiquitous Event Mining to Enhance Personal Health** 2018

Vaibhav Pandey, Nitish Nag, Ramesh Jain, Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers

- Surface Type Estimation from GPS Tracked Bicycle Activities** **2018**  
 Nitish Nag, Vaibhav Pandey, Aishwarya Manjunath, Avinash Vaka, Ramesh Jain, arXiv preprint arXiv:1809.09745
- Intrinsic and Extrinsic Motivation Modeling Essential for Multi-Modal Health Recommender Systems** **2018**  
 Nitish Nag, Mathias Lux, Ramesh Jain, arXiv preprint arXiv:1808.06467
- Endogenous and Exogenous Multi-Modal Layers in Context Aware Recommendation Systems for Health** **2018**  
 Nitish Nag, Vaibhav Pandey, Ramesh Jain, arXiv preprint arXiv:1808.06468
- Live Personalized Nutrition Recommendation Engine** **2017**  
 Nitish Nag, Vaibhav Pandey, Ramesh Jain, Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care
- Pocket Dietitian: Automated Healthy Dish Recommendations by Location** **2017**  
 Nitish Nag, Vaibhav Pandey, Abhisar Sharma, Jonathan Lam, Runyi Wang, Ramesh Jain, International Conference on Image Analysis and Processing
- Health Multimedia: Lifestyle Recommendations Based on Diverse Observations** **2017**  
 Nitish Nag, Vaibhav Pandey, Ramesh Jain, Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval
- Cybernetic Health** **2017**  
 Nitish Nag, Vaibhav Pandey, Hyungik Oh, Ramesh Jain, arXiv preprint arXiv:1705.08514
- Fibroblast Growth Factor 21 Is Not Required for the Reductions in Circulating Insulin-Like Growth Factor-1 or Global Cell Proliferation Rates in Response to Moderate Calorie Restriction in Adult Mice** **2014**  
 Airlia CS Thompson, Matthew D Bruss, Nitish Nag, Alexei Kharitononkov, Andrew C Adams, Marc K Hellerstein, PloS one

# ABSTRACT OF THE DISSERTATION

Health State Estimation

By

Nitish Nag

Doctor of Philosophy in Computer Science

University of California, Irvine, 2020

Professor Ramesh C. Jain, Chair

Life's most valuable asset is health. Continuously understanding the state of our health and modeling how it evolves is essential if we wish to improve it. Given the opportunity that people live with more data about their life today than any other time in history, the challenge rests in interweaving this data with the growing body of knowledge to compute and model the health state of an individual continually.

This dissertation presents an approach to build a personal model and dynamically estimate the health state of an individual by fusing multi-modal data and domain knowledge. The system is stitched together from four essential abstraction elements: 1. the events in our life, 2. the layers of our biological systems (from molecular to an organism), 3. the functional utilities that arise from biological underpinnings, and 4. how we interact with these utilities in the reality of daily life. Connecting these four elements via graph network blocks forms the backbone by which we instantiate a digital twin of an individual. Edges and nodes in this graph structure are then regularly updated with learning techniques as data is continuously digested. Experiments demonstrate the use of dense and heterogeneous real-world data from a variety of personal and environmental sensors to monitor individual cardiovascular health state.

State estimation and individual modeling is the fundamental basis for many revolutionary efforts in health. Continuous access to the individual's state allows a departure from disease-oriented approaches, to a total health continuum paradigm. Precision in predicting health requires a detailed understanding of an individual and their trajectory. By encasing this state estimation within a navigational approach, a user can use a systematic guidance framework to plan actions for each moment to transition their current state towards a desired one. This work concludes by presenting this framework of combining the health state and personal graph model to perpetually plan and assist us in living life towards our goals.

# Chapter 1

## Introduction

It is health that is real wealth and not pieces of gold and silver.

---

*Mahatma Gandhi*

Vibrant health lays the foundation of high-quality living. If we define quality as a function of positive experiences over time, then we want to optimize those experiences and extend their duration. Any life experiences rely upon a subset of general abilities to move, think, sense, and maintain homeostasis of our biology. These abilities arise from a complex interactive network of molecular machines that are always in flux. Maintaining and expanding the capacity of these abilities paves the way for enhancing health and, therefore, quality of life. When these abilities suffer from dysfunction, we realize we must take care of them. Rather than address health when a person has already fallen ill, the challenge for the next century and beyond is to ensure long lives of high performing health for each individual, thus enabling people to flourish in their life goals.

Traditionally, health has been viewed with the lens of avoiding sickness rather than pushing towards health optimization. There are exceptions with athletics and military, where

improved health status directly equates towards the success in the field. Why have the health sciences generally placed a large number of resources in research and application towards avoiding sickness rather than optimizing health? Based on the historical trend of communicable disease, it was *was* a great need.

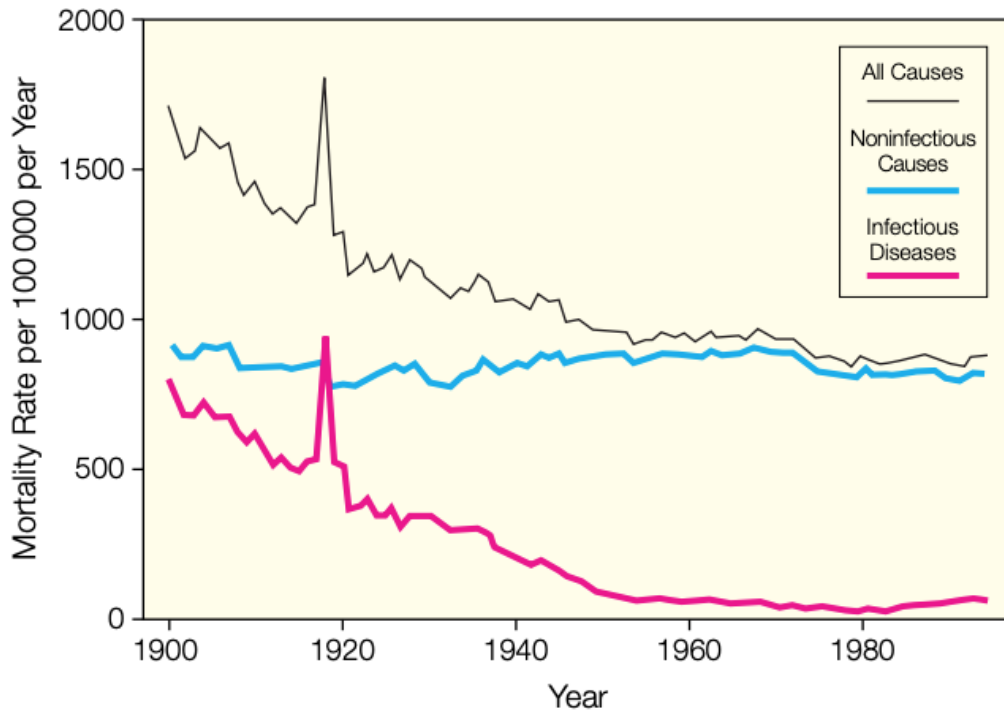


Figure 1.1: **Crude Mortality Rates for All Causes.** In the latter half of the twentieth century, the non-infectious disease continues to be a significant cause of mortality, reflecting the need for new strategies in health sciences.

Until recently in human history, a significant contributor to mortality and morbidity was infectious disease and injury. Progress in health sciences during the last century resulted in a significant decrease in communicable diseases, as shown in Figure 1.1. Antibiotics reflected one side of the coin, by treating patients in a reactive manner, once an infection had already become a health concern. On the flip side, advances in preventing infections through population-level data offered insight into ways to improve public hygiene and contain infectious agents, while vaccines biologically prevented infections. Continuing progress in battling infectious diseases today remains a valiant effort in sciences and medicine, but

it cannot scale as a general approach for all human health. Infection is mostly an episodic problem, in which diagnoses are made after a patient is under attack and arrives at a medical facility. Treatment is usually prescribed based on evidence-based rules to solve the problem, and monitoring of the patient no longer occurs. The process described is an **episodic** approach towards an individual's health. Episodic refers to *occasional measurement and awareness of health state* and actions to change the state towards a better one.

Globally chronic diseases, which include cardiovascular disease (CVD), cancers, chronic respiratory disease, and diabetes, have emerged as the primary contributor to degrading quality of health. We see this including impoverished areas of the world, with 80 percent of chronic disease burden occurring in developing countries [2]. The WHO explains that all regions of the world have a similar relationship between the major modifiable risk factors and the main chronic diseases [211]. Modern medicine relies mostly on reactive health systems that take care of individuals only after they become sick due to the established paradigm of episodic medicine. This means that healthcare professionals are systematically responding to acute health problems and are symptom-focused. The overall health is not a priority within healthcare systems [98]. Prescribing medicine is the hallowed strategy medical experts have taken toward treating disease, just as they approached infectious disease. It is necessary to move beyond this episodic health model to tackle chronic disease and significantly improve quality of life. When comparing chronic disease to infectious disease, there are fundamental differences in how the situation arises. There is no single event that leads to the disease, rather slow changes in the state of the body. Risk factors, including raised blood pressure and glucose levels, abnormal blood lipids, and weight (overweight and obesity), express the causes of chronic diseases. Major modifiable risk factors alongside non-modifiable risk factors (including age and genetics) are the root causes of these diseases over time [211]. If we are to apply a control input to keep people on a healthy trajectory via feedback loops, we must be able to estimate the health state continuously [292]. Control systems inherently need to know the difference between the current state versus the goal state. This approach



is taken in acute moments in the hospital after entering an undesirable health state, such as during anesthesia or Intensive Care Unit (ICU) monitoring. Once again, these are episodic moments in a person's life. Hence there is a clear distinction between episodic versus life-long continuous estimation problems in health.

Health is a dynamic state that is continually changing based on a specific individual's biology, environment, and lifestyle. The state space of health is complex and multidimensional, but for the sake of simplicity, in Figure 1.2, we describe a one-dimensional continuum ranging from optimal health to death. Everyone is somewhere on this continuum at every moment, but the quality of life increases if there is a shift towards the optimal (left or green side). Technology that senses a user's position on the health continuum can drastically improve human health. In prehistoric times, only symptoms (perceptual indications) guided people's actions. Ancient civilizations started to document medical conditions with basic human sensory abilities, such as how people visually appeared, their pulse rhythm, or coughing sounds. In clinical language, symptoms (what a person feels) or signs (what is observable by another person) are unfortunately latent signals of an already dysfunctional biological state. Today, advanced methods detect changes in health, but only in a professional health-care setting after patient visits due to severe enough symptoms or on an annual basis. Reducing the lag time in recognizing health state changes requires transitioning from the current measurement approach to a new high-resolution multi-modal continuous-sensing paradigm.

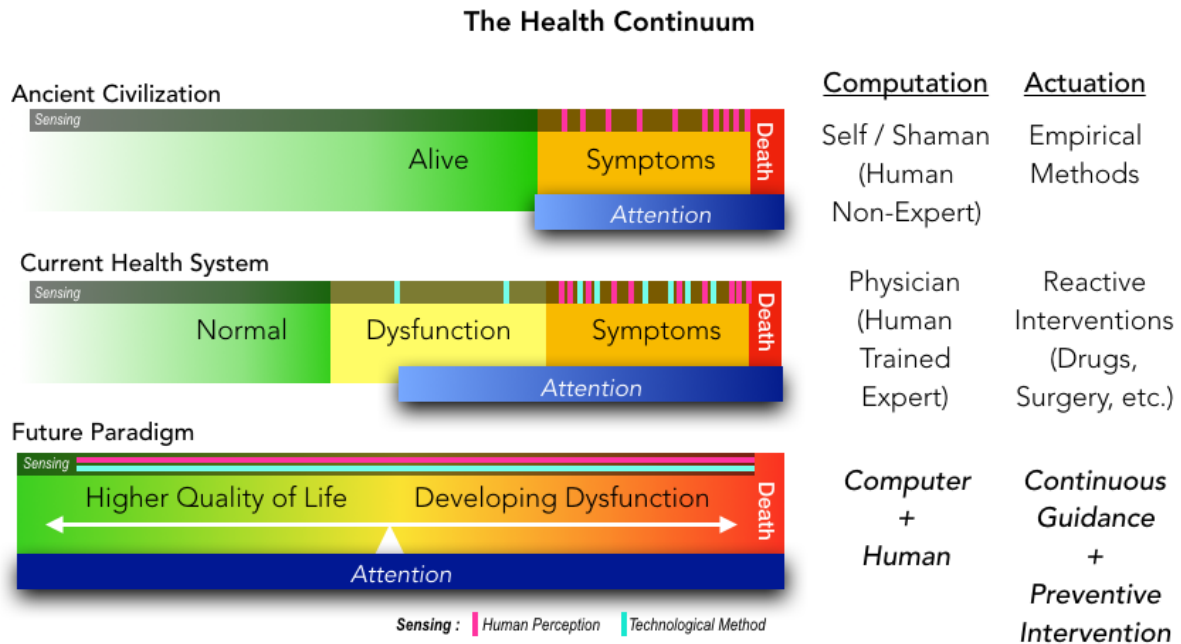


Figure 1.2: **The Health Continuum.** Attention to health status occurs when there is a detectable signal. Ancient civilizations could only use subjective symptoms and signs as an indicator of the health state. Today, we have episodic measurements of a few criteria to understand our health state, such as blood pressure or lipids on an annual basis. Beyond those measurements, we still primarily rely on symptoms as the primary signal. Both of these signals are sparse, and most individuals do not even collect symptom data. The next revolution is to have continuous sensing of the health state so we can always have attention to improving health through the means of ubiquitous computing and modeling at the individual scale.

A general characteristic of humans and computing systems is the intelligibility and clarity of understanding a situation, allowing for the best long term, planned decisions for a future outcome. In today's society, we see this clearly in economics. The tangible and quantitative nature of financial capital allows humans or computers to make the decisions they see fit for their intentions and plan for a quality of life they aspire to have. The difference between finances and health is that live quantitative health status is virtually invisible to us, especially when we are healthy. In financial matters, people can quantitatively see how much they are saving or spending at any given moment. Market data updates every second across the globe. If people become aware of their biology, they will see how daily life is affecting their health

and make informed decisions accordingly. Health decisions are too important to be episodic. They must be an intrinsic part of daily life, giving us the motivation behind proposing the state estimation and navigation perspective in this work.

Ultimately, individuals, clinicians, and in general cybernetic systems (Figure 1.2), make decisions based on determining the health state of the individual. Because people are unable to feel their health change over time from the multitude of factors affecting them, we need to develop methods to quantify and report health status using continuously collected multi-modal data sources. If we can track changes before biological dysfunction, we may be able to correct course, prevent, or delay the onset of chronic diseases. We can use the same principle to change that health state towards a more optimal direction if the user wishes.

Computing efforts regarding state estimation in other fields have advanced society drastically in the last several decades. For example, the Global Positioning System (GPS) launched in 1973 by the United States military for location state estimation. It was opened to the public in 1983 by a bill signed by President Ronald Reagan with limited functionality and was subsequently offered with full, accurate functionality by a bill signed in 2000 by President Bill Clinton [175]. In the last 20 years, this allowed for systems to take a user's location to provide them with services such as understanding their position on a map, tagging photos, providing navigation, tracking their movements, hailing a taxi, or ordering pizza to the home doorstep. The state space of location included four dimensions of latitude, longitude, altitude, and time. Our work aims to lay the groundwork for similar tasks in the context of health.

So how can we compute the health state continuously? The primary fuel arises from the diverse data streams produced about an individual daily. With the rapidly increasing availability of low-cost sensors in the last decade, there has been an explosion in the amount of continuously collected multi-modal data. We see the relevance in the field of health with the advent of wearable, Internet of Things (IoT), and ambient sensors, human-computer

interactions, images, video, audio, and other digital sensors. Many of these low-cost sensors produce continuous streams of data usable in a wide range of scenarios. The challenge rests in leveraging this data to provide perpetual insight into the health of individuals. While this allows measurement of the user state continuously, the downside of these sensors is that measurements can be very noisy, cause information overload, produce non-actionable metrics, or are not directly related to the attribute we want to measure. These limitations have so far proven to be a barrier in using low-cost sensors for real-world health state estimation (HSE) and decision making for health. Furthermore, enabling data assimilation will require a causal understanding of how sensor information relates to health status. To transform the data into usable information, existing medical and biological scientific domain knowledge must be used.

There are many challenges in providing a high-resolution continuous and individualized state estimation for health. This dissertation aims to establish the preliminary groundwork to address issues for the general health state concept and methods by which to apply it at an individual scale.

In the next section, we will look at the thought process transition from traditional medicine to state-of-the-art approaches towards improving health and state estimation, giving insight as to what components are emerging needs in computing the health state.

## **1.1 Overview of Health and Medicine**

The history of significant trends are important for us to understand health today and its future developments, especially in understanding patterns that cause progress. From ancient times until today, humans pay most attention to their health with the lens of healing from illness. From a traditional point-of-view, the health of the individual only became a concern

when negative impact could be perceived, causing the development of systems that operate in the following steps. First, the individual acts as the primary sensor that their health state is not the same as earlier. Second, a doctor or domain expert would conduct further tests to gain an understanding of the causal reason behind the change in the health state. Third, the doctor or health expert would prescribe medication or procedure that would hopefully return the individual to their former health state. After this point, the individual is no longer monitored until then next issue arises. The impacts on the health state occur due to several critical situations. Situations that are significant impacts on health are listed below. Each situation has a particular state estimation challenge and a unique set of actions to keep health and homeostasis.

1. **Infections** have historically been a threat to health. The resulting invasion from different species battles with natural immune defenses, medical agents such as antibiotics, or physical removal.
2. **Cancer** occurs due to errors in the genetic information of a cell. There is a natural rate of error for DNA replication enzymes, which causes errors during each division of a cell [101], a component of natural life and the process by which evolution occurs. We, therefore, see age as a factor for cancer. Additional errors in the DNA structure happen by insults from carcinogens, inflammation, or stress.
3. **Physiological function** of the person determines their operation from cellular to organismal level. This includes the pumping of the heart to supply blood flow, the strength of bones to provide support, the absorption of nutrition in the digestive system, and the capacity to see. Many chronic diseases are a result of physiological dysfunction.

4. **Injury** can occur from either chronic issues or acute events. Overuse or improper use may eventually result in injury. Acute injuries are caused by unintentional accidents and or intentional combat. The severity can range from mild to traumatic.
5. **Neurological and mental health** is a complex topic and traverses a broad range of sciences. Inherently there are some components of this aspect that are objective and others that are subjective, leading to unique challenges in scientific progress.
6. **Developmental** disorders are caused by various factors determined by cellular growth from fertilization until adulthood. As examples, these can be encoded in genetic errors (chromosomal errors at birth), cell signaling biology (receptor dysfunction), or deficiency during growth (low bone mass from low dietary calcium).

Before the twentieth century, infectious disease was a key determining factor for mortality. With emerging infections, finding a cure, or reactively treating the disease that had already occurred became commonplace. This is why health was primarily determined by the absence of illness or disease [12]. The progress in the dramatic reduction of infectious disease came through significant advances in state detection and actionable steps. The state detection improved through the use of culturing the infection and microscopy to identify the pathogen. Culturing amplifies the signal from the infectious agent. We use these techniques even today, although further advances in gene sequencing and other assays promise to speed diagnosis time and precision. Actionable steps to address the infection are usually antibiotics or antiviral medications.

Adversarial pathogens became less of a fixture in medical research over the last century, and chronic diseases became the center point of health challenges. These chronic diseases are primarily a result of physiological function change. Currently, six in 10 adults in the United States have a chronic disease; 4 in 10 adults have two or more. This makes chronic disease the leading cause of death and disability [196]. The major chronic disease that dominated

epidemiological data since 1940s was and continues to be cardiovascular diseases as shown in Figure 1.3, with 647,000 Americans dying from heart disease each year (or 1 in 4 deaths) [42]. Heart disease is, therefore, the leading cause of death in the United States. These above situations are mostly modeled using large-scale population data by aggregating individuals to produce statistically significant outcomes due to the lack of high-resolution personal data. Thus the state estimation for chronic disease is based on a population model that predicts the health outcomes of an individual. Measurements on the individual are taken generally on an annual basis to update the state. Medications are given to individuals if the state changes to a new classification, given population model thresholds. Overall for chronic disease, patient state estimation is conducted through general modeling, updated on an episodic basis.

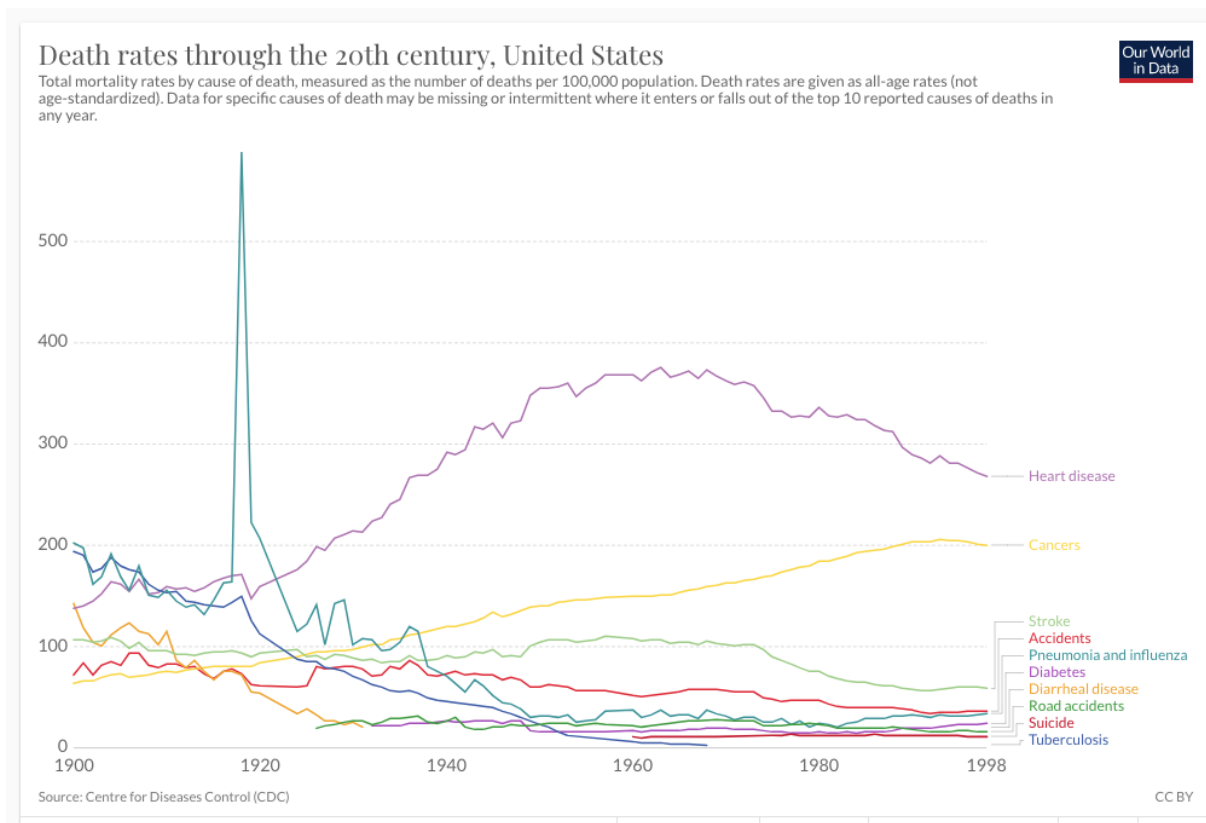


Figure 1.3: **Death Rates.** This graph shows the death rates in the 20th Century United States [172].

While genetics and mutations play a role, CVD and a significant portion of cancers are largely a function of external factors that can be addressed by factors extrinsic to the biological risk, such as lifestyle and the environment. Cancer develops from both random intrinsic DNA replication errors (or “bad luck”) and from factors not related to genetics. Cancer causation is multi-factorial and the modification of risk factors is critical in cancer prevention [305]. If external and environmental risk factors dominate, then lifestyle choice, which reduces risk exposure, is crucial for this reduction. One-third of cancers are attributable to environmental factors or inherited predispositions [264]. The approach to chronic diseases has been similar to infectious diseases in that treatment has been a reaction to symptoms. Introductions to medication continue; however, the rate of disease decrease has stabilized, as shown in Figure 1.3. This motivates the need for more robust methods to address these diseases than sole reliance on medication developments and reactive approaches for addressing when an individual is already in a state of disease. Furthermore, tracking the lifestyle and environmental events give substantial clues into the evolving health state of the person in addition to medications or procedures that assist in reducing risk for disease. In summary, the actionable items to address the chronic disease are widespread and occurring at every moment of the day.

Health science disciplines usually still focus on these reactive approaches today. Integrating the knowledge from this is essential, as the illness lens highlights regions of our health state that are undesirable. This thesis work will primarily focus on physiological function concerning cardiovascular function and disease, as it is the primary determinant of major chronic diseases.



## 1.2 Modern Trends in Health Sciences

There is a need for better state estimation and actionable steps for future progress. As reactive methods for addressing disease is no longer sufficient for a global decrease in rates of chronic disease, we examine new models. These approaches require a true understanding of the individual state of health and predicting and preventing entry to poor health state regions. Personalized methods also acknowledge that individual state must be uniquely observed, requiring high quality and quantity of data about an individual.

Projections show that medicine and the health sciences will move from a reactive to a proactive discipline over the next few decades a discipline that is predictive, preventive, precise, personalized, and participatory [120, 237]. There must be a holistic approach to disease, emerging technologies, and analytical tools to ensure proactive advancements. Holistic approaches refer to the total connected network of the individual's health. Emerging technologies include the proliferation of sensors and computing power, and analytical tools refer to the manipulation of large amounts of information to produce meaningful and actionable insight. Other advances we will also review include developments the growing wellness industry which are currently outside the scope of the healthcare industry. Wellness focuses on moving a patient towards the optimal state of health, as shown in the future health continuum in Figure 1.2. In contrast, healthcare focuses on moving the patient from the red area of dysfunctional health to the level of mediocre health. We will take a deeper dive into these new approaches as they will be an integral part of the design requirements of this research.

### 1.2.1 Personalized Health

*Personalized health refers to building individual models to represent the health state, and tailoring inputs for a specific individual.* Personalization includes two key sub-models. The

first is understanding of an individual’s preferences and life situation, encapsulating the subjective attributes and autonomy of the user. On the other hand personalization for the biological machinery is an objective model. As an example, in Figure 1.4, the individual heart electrical patterns are modeled [11]. Data from “omic” sources are incorporated to understand the complexity of an individual’s biology, which includes genetics, genomics, microbiome, proteomics, and metabolomics [201]. The word “oma” refers to a mass in Latin, and in the context of omic data refers to the large mass of information. There must be a combination of both of these models (subjective and objective) for many use cases. Examples include food preferences for a diet change and exercise preferences for modulating cardiovascular health.

Personalized medicine, which is more specific in the treatment of disease, is used in the context of focusing on a person’s unique molecular profile to help them be less susceptible to disease. Genetics plays a large role in personalized medicine to understand one’s unique traits and shape the biological model through this knowledge. The Human Genome Project had an impact on producing treatments that could be better tailored to the individual. Reading the sequence of the DNA will allow one to predict how an individual will respond to a particular drug and to receive earlier diagnoses and risk assessments for more appropriate treatment [278]. Objective personalization alone is not enough, as the person must be part of the whole loop [57]. The integration of contextual components relevant to the user’s wishes and preferences is just as important. As mentioned earlier, augmenting the data and model with subjective data from the user also has unique challenges. From this model, we can create predictions, precisely guide the individual, and prevent entry into undesirable regions of the health state space.

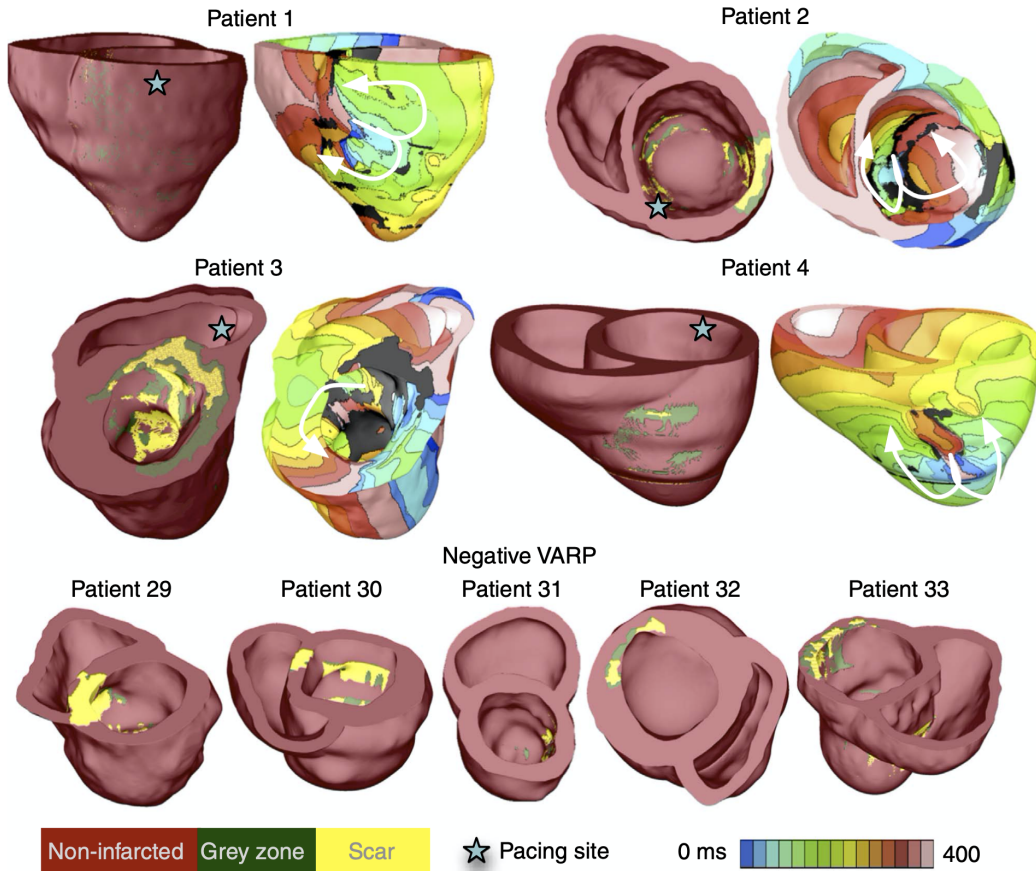


Figure 1.4: **Personalized Heart Models.** Patients clinical magnetic resonance imaging data constructs personalized three-dimensional computer models of post-infarction hearts. These are assessed for the propensity of each model to develop arrhythmia. In a proof-of-concept retrospective study, the virtual heart test significantly outperformed several existing clinical metrics in predicting future arrhythmic events. The non-invasive personalized virtual heart risk assessment is a form of highly specialized individual health state modelling[11].

## 1.2.2 Predictive Health

*Prediction refers to estimating the probability of a future event.* Improvements in this require the above personal model. Predictive medicine is a branch of medicine that focuses on risk identification for a disease, such as a future unwanted biological event, including a heart attack. The goal is to identify patients at risk that could then enable prevention or early treatment. This can be done by either single or multiple analyses as markers for future disease [16]. Risk calculation of particular diseases and obtaining characteristics of

individuals could help medical professions to predict a patient's risk for an outcome of interest and also determine what treatment may be most effective. The Emory Predictive Health and Society Strategic Initiative and its Center for Health Discovery and Well Being define health in the context of the entire human experience. They identify measurable variables that describe and predict a healthy state and to use that knowledge to design health-focused interventions that are affordable and effective [32]. Predictive health utilizes data sources from a diverse array including genomics and epidemiology in order to best predict the risks for an individual. From a computational perspective, this requires data mining to find patterns that pull in knowledge from population and individual components, while also employing learning techniques for data-driven pattern discovery.

### 1.2.3 Precision Health

*Precision health refers to more accurate approaches to understanding the health state and optimizing inputs to produce the ideal state.* For example, while doctors may inquire about family history of disease to understand hereditary risk, a more precise approach would be to sequence one's DNA in order to analyze the genetics data for inherited disease risk. This greater detail in understanding allows for a more targeted approach in the estimation and input selection. Precision will contribute to improving HSE and inputs through various avenues that we will discuss below.

- **Biological** understanding with more precision, allows for better state estimation and input approaches. At an organ level, each individual has unique biology. In the case of cardiovascular function, models shown in Figure 1.4 are also precise, in that they try to pinpoint precisely the nature of the cardiovascular electrical activity [11]. At a molecular level, the details of signaling pathways and cellular processes allow for better health and medicine. A classic example in medicine is in the Cyclooxygenase (COX) inhibitor

drug class. Ibuprofen is a drug that inhibits all COX enzymes, which help in reducing pain but also cause problems for COX enzymes, which are part of general functions in the gut. The classification of Ibuprofen is thus a non-specific COX inhibitor. The discovery of COX-1 and COX-2 specific sub-types gave rise to a new class of medications that are specific for COX-2 (the component responsible for pain), allowing for more the development of more advanced pain medications (Figure 1.5). This improvement in the precision of molecular targets continues to be an active area of science. Lastly, genetics greatly influences precise and personal biological understanding, which we will review next.

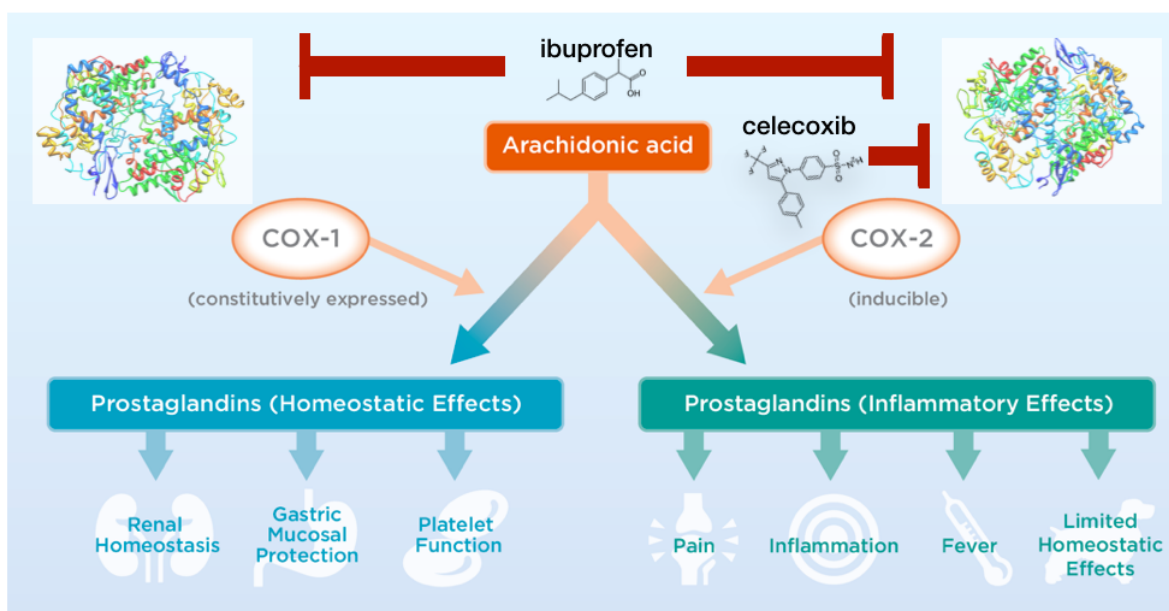


Figure 1.5: **Precision in Biological Understanding.** Selective COX-2 inhibitors like celecoxib are a type of Non-Steroidal Anti Inflammatory Drugs (NSAID) that directly targets COX-2, an enzyme responsible for inflammation and pain. Targeting selectivity for COX-2 reduces the risk of peptic ulceration and is the main feature of the drug celecoxib. This type of precision arises from more in-depth biological understanding.

- **Genetics** plays a fundamental role in having precise information for molecular and cellular modeling and targeting for an individual. Even more cutting edge in this field is the use of immune therapy, in which a person's own white blood cells are trained to precisely identify and destroy cancer cells in the body rather than delivering poisons

that impact both healthy and cancer tissue [299]. Furthermore, targeting specific antibodies can more precisely treat an autoimmune disease. These are examples of biological targeting using genetics. Risk modelling and predictive efforts may also use genetics. For example, the BRCA gene is a determinant of breast cancer risk. Patients having high-risk BRCA genes may elect to have mastectomies or get screened more frequently, given this genetic risk. In summary, genetics allows for many aspects of state estimation or technologies to provide a more accurate outcome for an individual.

- **Location** precision on the body is a general concept that may improve many areas of health. The traditional crude approach is to give a general pain killer medication that is distributed to the whole body via the blood and has adverse side effects due to the lack of a specific target. A precision approach would understand where the inflammation is in the knee and deliver a medication targeted to only the knee tissue and molecular pathways of relevance to the disease pathology. For example, a cream or injection is more precise than an oral pill. Topical NSAIDs may be better for localized pain near the surface of the skin, such as pain in the hands and knees, while oral NSAIDs may be better for pain deep within the body. Similarly, location for surgery can be more precise using laparoscopic methods, rather than an invasive open procedure. Precision in surgical methods also enables providers to be more accurate in their procedures, making fewer mistakes and damage to healthy tissue. Laser cataract surgery is needed to remove a cataract. In this procedure, a small probe is inserted into the eye, emitting ultrasound waves to break up the cataract into pieces. In laser surgery procedures, lasers are used to make incisions to create a more precise opening to the lens capsule [43].
- **Dosage** of inputs, such as drugs, can be more precise to the needs of an individual. Context, the personal model, and the components mentioned above characterize precise needs. When an individual is in high altitude, there is an adverse effect on coagulation

in patients taking Warfarin [262]. If the dosage of the medication could be controlled depending on the location of the individual, this could be more precise in controlling the coagulation more appropriately. Cytochrome P450 is an enzyme that metabolizes the breakdown of drugs; individuals have different rates of breakdown depending on their specific enzyme type. Knowing the exact sensitivity and breakdown methods will better allow for precise drug dosage. Extending beyond drugs, we can understand there is a precise quantity of inputs that produce the optimal outcome of the individual's health state. Knowing the precise amount for each input (such as food, exercise, drugs, environment, sleep) is a precision challenge.

Precision medicine is a specific subset of precision health focused on optimizing the inputs to better alter the health state. Medicine is an input that is purposefully designed to change the health state in a certain circumstance. The NIH defines precision medicine by how genetics, environment, and lifestyle can help determine the best approach to prevent or treat disease [202]. More medical institutions such as Stanford Medicine are using precision medicine via unique biology and life circumstances for diagnostics [230]. Former President Barack Obama launched the Precision Medicine Initiative to accelerate these advancements, emphasizing how precision medicine will enable health care providers to tailor treatments to better suit patients through the use of gene sequencing and lifestyle factors. To get there, we need data about the patient collected by health professionals and the patients themselves [125]. From a computational perspective, this requires an in-depth database for capturing the relevant domain knowledge that matches precisely with each individual's context and personal model. Capturing this high resolution of data and having the individual take action on the best inputs requires them to be a part of the cybernetic loop.

## 1.2.4 Participatory Health

*Participatory health refers to an individual directly engaging with understanding their health state and providing inputs for change.* Traditionally healthcare providers were the key individuals who understood patient health state and decided how to alter it. The providers and patients generally held all information had little interaction with their health state (other than through following directions on medications). However, more recently, healthcare providers and patients are working as partners in health by using communication tools to increase active patient participation in medical decisions. This approach uses physical and physiological tracking devices, “easy-to-use software that records one’s data, analyzes them, interprets them graphically, and allows one to share and compare them with others. This is expected to enable more reliable self-monitoring and more self-management of health care” [170]. Ultimately, participatory medicine refers to the patient taking actions that change their health state, rather than depending upon the healthcare team to address the changing health state of an individual. From a computational perspective, this requires development in human-computer-interaction (HCI) for engagement, integrating into relevant services for the user, and learning patterns from the event streams, which increase the probability of the user completing specific actions.

## 1.2.5 Preventive Health

*Prevention refers to altering the probability of a future event / or avoid region of interest.* Preventive health behavior has been described in 1966 by Kasl and Cobb, as “any activity undertaken by a person who believes himself to be healthy for the purpose of preventive disease or detecting disease in an asymptomatic stage, in contrast with illness behavior for an individual who is ill, defining the stage of health and discovering remedy” [142]. Today, preventive medicine is a medical specialty recognized by the American Board of Medical



Specialties (ABMS), focusing on promoting health and well-being and preventing disease, disability, and death. Physicians in the preventive medicine specialty practice in the subspecialty areas of public health and general preventive medicine, occupational medicine, and aerospace medicine [3]. While officially, preventive medicine is practiced medically with these specific guidelines, the field of preventive medicine continues to ameliorate individual risk for disease by altering the probability of future risk. It represents a health state as a compilation of risks and seeks actions that reduce this risk based on evidence. Based on the predictive component, preventive measures are taken, such as the common practice of reducing risk factors for heart disease. These risk factors include lowering cholesterol, blood pressure, and other modifiable components. Knowing the heart health state, however, requires much more depth. From a computational perspective, this requires mapping the regions which are undesirable in the health state space, understanding the user health state, and providing inputs that increase the distance to the undesirable regions.

### **1.2.6 Healthcare Systems**

Healthcare defined in this work includes all services and products that are within the scope of health insurance coverage. The economic size of this field globally is 7.3 trillion US Dollars [89]. This field is evolving to encompass new frontiers that will impact individual health. The healthcare industry is attempting to address the complexity of individual health via technological advancements such as digital therapeutics, pharmacy supply chain, virtual primary care, artificial intelligence, genomics, and advancements in consumer health and wellness. There are also developments for providers including diagnostics, and drug research and development to ultimately enhance their services [39]. Trends in the healthcare industry show that consumers have access to more personalized healthcare options that provide a concierge level of service and are expanding convenience through services like video conferencing. There is also greater transparency of healthcare services, given the level of engagement individuals

are now having with the health system. All together, healthcare is growing in domains for the individual, providers, and systems at large. The primary computing advancements for HSE will be in using more advanced diagnostics that are too expensive or exclusive for the patient to procure on their own.

### **1.2.7 Performance and Wellness**

A large developing area within modern approaches to holistic health is through the modes of “wellness,” the consideration of inputs of lifestyle, the environment, and social and perceptual factors. Together, these bring about a more robust consideration of health in daily living and behavioral considerations, and an improvement in health performance, rather than an emphasis on eliminating or reducing disease. The economic size of this field globally is 4.5 trillion US Dollars [89]. The wellness industry is targeting improvement in various components of an individual’s life, as outlined by CBIInsights [40]. Many of the services provided involve tracking and developing methods for changing the health state. High growth areas which include areas for technological advancements for individual health include:

- Personalized nutrition such as food and beverages, vitamins/supplements, and active nutrition.
- Beauty and personal care.
- Fitness and active living, including gym access applications, activity tracking, and measurement of physiological changes.
- Mental wellness including behavioral change programs, educational platforms for mindfulness/meditation.

- Corporate wellness as provided by an individual’s employer, that provides comprehensive platforms for interactive with health such as by measuring, tracking, understanding and making changes to one’s state in multiple domains of their health.
- Sleep technologies to track, change behaviors, and understand the best methods for sleep improvement specific to an individual.
- E-commerce for the sale of online health specific products.
- Feminine care to address women’s unique health needs.
- Travel/hospitality to provide modes of experiencing wellness via retreats and getaways.
- Clothing and fit unique for one’s active performance.
- Functional health or primary care medical practice with a holistic approach, combining nutrition, lifestyle, and medical testing.

New forms of wellness demonstrate the convergence of personalized medicine with digital health, artificial intelligence, systems biology, social networks, big data analysis, and precision medicine. This field is known as *scientific wellness*. “The idea of scientific wellness is a quantitative approach that includes improving the health of individuals, creating personalized treatments, reversing disease transitions and reducing costs distinct from the current wellness trend focusing primarily on behaviors such as diet and lifestyle” [254]. People will then be able to use data and tools to modify habits for diet, exercise, sleep, and more to optimize personal health.

Looking historically and at current advancements, progress in health depends on estimating the health state in order to realize improvements in the treatment of disease or helping individuals live well. Thus the state estimation and health measurements over time are critical for scientific progress.

### 1.3 Health State Estimation is Critical for Progress

Even though the World Health Organization (WHO) for 74 years has defined a preliminary health state, there continue to be inferior methods by which to understand the individual health state. The first page of Google Scholar with the search term “health state estimation” (as of January 1st 2020) primarily displays results in the health state of batteries, drilling, machines, ball bearing, and aircraft engines. The challenge remains in developing a generalized framework for estimating the health state of humans. Health is estimated in most detail when we are in our worst state in the Intensive Care Unit (ICU) at a hospital. Humans doctors with minimal historical knowledge or elementary quantitative modeling of the individual are making judgments about the state of health for their health decisions. Using this judgment, they provide episodic actions to re-direct the health state to an acceptable level, not an optimal level.

Methods for measurement have been developed through time to estimate particular components of health more accurately and more often. Microscopy and the identification of microorganisms detected specific pathogens, highlighting the presence of what pathogens determined disease. The Framingham Heart Study, which started in 1948, has reviewed a diverse population over time to analyze patterns, trends, and outcomes that may apply to the general population. One finding was the discovery of cholesterol and its impact on CVD. From measurements at a population level, the relationship between cholesterol and the progression of CVD was established [130]. Thus, cholesterol now has become an indicator of CVD and cardiorespiratory fitness (CRF), as described by the American Heart Association (AHA) [234, 152].

The more health estimation moves towards real-time, it will open doors for controlling health more precisely. For example, the first glucose monitor was introduced in the 1970s and then began the self-monitoring of blood glucose. Advancements in continuous glucose monitoring,

which read blood glucose monitors more frequently, gave more precise information to manage diabetes. The accuracy of the sensors has been improving over time, and the frequency for measurements has increased [118]. Quantified Self, a movement with the motto self-knowledge through numbers embodies the importance of continuous health measurements. The movement aids individuals in using different technologies to collect data on their daily living (i.e., food intake, physiology) to analyze and improve their health state. Sensors are also more commonly used to collect genetic, environmental, and phenotype information and storage in the personal health record [170]. Objective Self takes the quantified self concept towards a more computationally robust representation of the individual by establishing a basic model about the individual [135].

What we need is a means by which we can continuously know our health state at all times in high temporal and spatial resolution. A parallel example is accessing GPS location on-demand, anytime. In a total health state space, an individual can see where they are in their state space and choose their location within the state space. Disease is just a subset of this state space of regions that are undesirable, similar to not driving a car into the middle of the ocean. What one cares about are the *places to be*, and the health state required to be there. Ultimately, why is the HSE needed?

*Measurement is the first step that leads to control and eventually to improvement.*

*If you cant measure something, you cant understand it. If you cant understand it, you cant control it. If you cant control it, you cant improve it.*

- H. James Harrington

## 1.4 Contributions

This dissertation contributes towards the advancement of dynamically computing the health state of an individual throughout every moment in life. The evolving state of the individual depends on the integration of both domain knowledge and data-driven discovery frameworks that facilitate computation of how data streams and event streams are related. The connected nature of life is modeled computationally through graph-based abstraction and combinatorial generalization. This state estimation paradigm allows individuals or entities to use the continually computed state as a foundation to build systems that can perturb the state towards a chosen goal. An application of this is done through a cybernetic and navigational approach. The focus will be through the physiological function lens, especially of CVD and CRF. Below the contributions are distilled into two main subgroups of general concepts and personalization concepts, visualized in Figure 1.6.

The general concepts include, but are not limited to:

1. ***Precisely defining health.*** Because health is defined in many different terms, we focus on the definition at the scale of an individual. A precise definition sets the best stage for any computation and measurement.
2. ***Establishing the dimensions of health.*** The complex nature of health requires high-level abstraction to represent critical features that make up the health of the individual. In this work, we use utilities as abstracted microservices to represent the dimensions of health.
3. ***Representing the general state space of health.*** The state space of health represents all the possibilities that can exist for human health and provides a foundation map base layer that allows embedding of other layers.

4. ***Structuring the model of inputs to the health state.*** Each moment in life causes some small change in the health state of the individual. Capturing these moments and translating them to inputs into the health state give the next time iteration of the health state.
5. ***Integration of domain knowledge and data-driven approaches.*** In order for the estimation and modelling to be done in an understandable way, depending on a single top-down or bottom-up approach will not suffice. Each has advantages and disadvantages. In this work, we use a combinatorial generalization artificial intelligence technique of graph network blocks (GNB) to meld the two approaches together.

The personalized concepts include, but are not limited to:

1. ***Creating a personal health state space relevant to the individual.*** The true health state space that is reachable for a single person is a subset of the general state space. Multiple factors such as genetics and age, carve out the state space for the individual.
2. ***Current health state of the individual within their personal state space.*** Knowing the current state is the primary objective of this dissertation. Using combinatorial generalization, we can establish a global set of attributes to give a current state estimate of the individual. Computing is done by instantiating a graph of an individual's utility and biological microservices, and estimating how the microservices change. These microservices in aggregate determine the global set of health state attributes.
3. ***Predicting the trajectory of the state at future time points with a personal model.*** By using the GNB, we can simulate future time points of inputs and produce their outputs as the predicted future health state.

4. ***Learning the input effect on a given state for a particular individual, editing the personal model.*** Given a large quantity of data produced by an individual, we use modern learning techniques to produce a model at the N=1 level, allowing us to more accurately model. We observe how inputs to a node in the graph produce results in the real world setting by modifying domain knowledge-based connections to suit a particular person.
  
5. ***Given a desired direction or goal, reducing the state space dimensions and presenting the current state in the context of the user intent.*** People interact with their health at a high level of abstraction and also have abstract goals. By decomposing this level of user interaction into the relevant dimensions, we can provide a user interface to understand their health state.
  
6. ***Providing a structure for health navigation.*** By understanding what inputs produce certain state changes in the individual, we present a method by which perpetual planning, used to guide an individual to alter their current health state towards their desired state. The above goal decomposition guides which inputs are most relevant and useful for the individual and presents each moment as a cybernetic control loop for the navigation system to control each step.

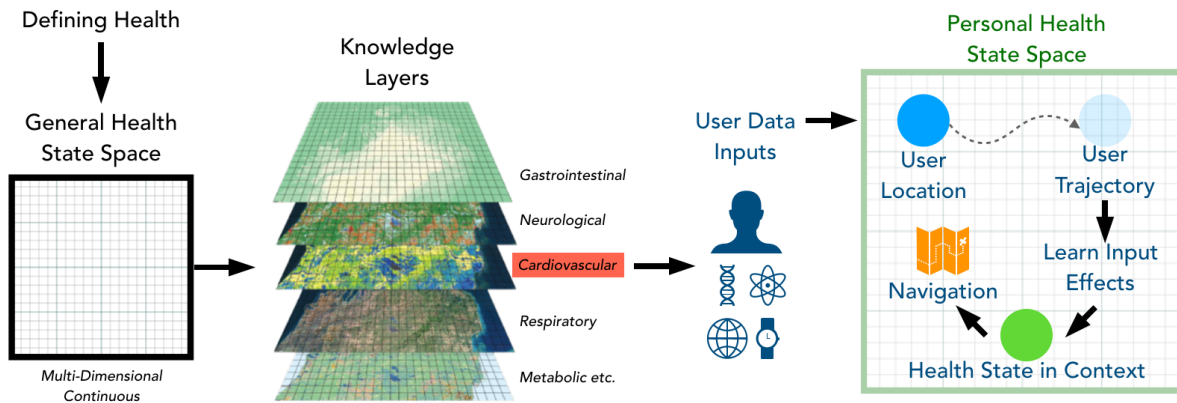


Figure 1.6: Visualization of Contributions.



## 1.5 Thesis Outline

Chapter 2 introduces the driving concepts of cybernetics and navigation for health computing and the flow by which the health state of each individual is dynamically changing. It reviews the literature to define a comprehensive view of “health,” how this health is continually changing, and how certain design elements are essential to produce continuous personal state estimation. Chapter 3 dives into a specific literature review of sensing, estimation, and modeling techniques used in this work. It concludes by presenting the scope of techniques for this dissertation. Chapter 4 presents the HSE framework through the detailed assembly of the design requirements into a cohesive system. Chapter 5 describes the experiments that illustrate HSE at an individual scale using multi-modal data and knowledge integration. Chapter 6 discusses the application of HSE by describing the methods of personal health navigation systems and cybernetics. Chapter 7 discusses the future opportunities in this field and related fields of work and concludes this dissertation.

# Chapter 2

## Understanding Health Computing

To live effectively is to live with  
adequate information.

---

*Norbert Wiener*

### 2.1 Cybernetics: Inspiration for this Work

In the simplest terms, cybernetics is about setting goals and devising action sequences to accomplish and maintain those goals in the presence of noise and disturbances. The Greek origins refer to the ‘art of steering’, and the term was introduced into the literature by Norbert Wiener, a mathematician at the Massachusetts Institute of Technology, USA [292]. His seminal publication “Cybernetics, Control and Communication in the Animal and Machine” described what we know today as control theory for dynamic systems. Cybernetic principles are a universal mathematical theme in both natural and artificial machines. The availability of sensors that estimate the system state from observation enables communication and control. This state is perpetually compared to the goal state so the system can perform the

best actions to maintain the desired goal state. In parallel, biological science also discovered control mechanisms as “homeostasis” via a series of physiology scientists, including Claude Bernard, Walter Cannon, and Joseph Barcroft [96]. Homeostasis referred to the stable internal environment at any given moment. It was first described in French literature originally in 1854 as *milieu intérieur* by Claude Bernard in [25]. Norbert Wiener’s work in cybernetics generalized these principles beyond biology. There are four fundamental principles in connecting the information from the physical world to the virtual “cyber” world, as shown in Figure 2.1.

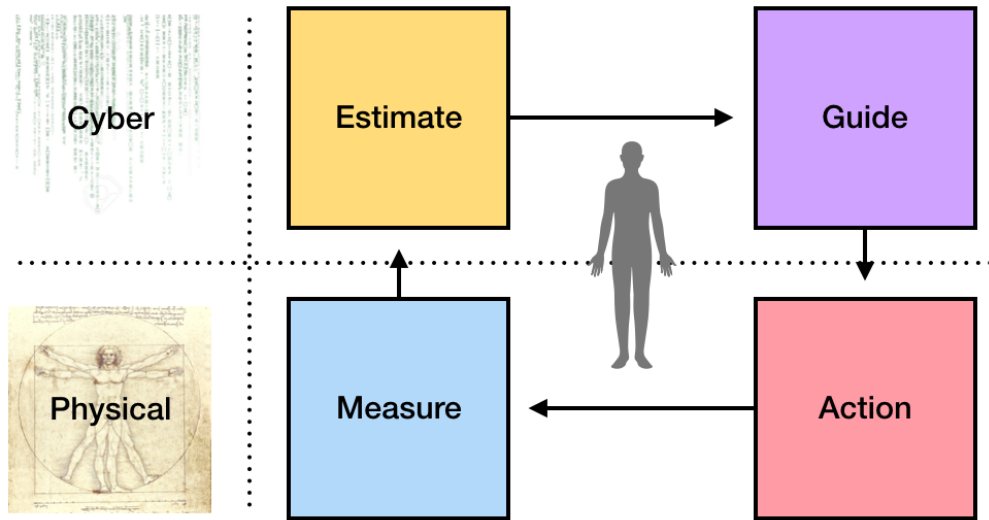


Figure 2.1: **Cyber-Physical Connection.** Cybernetics connects physical reality with computing information for control. In the real world, sensors measure various aspects of the world and translate these observations into data. The system must take the data and produce an estimate of the system state. Based on this computed estimation and a control goal, the system produces guidance for the real world. The actions taken in the real world may or may not be the same as the guidance due to disturbance, noise, or other factors. Thus, we take continuous measurements to see how the system state has been perturbed based on the last iteration of guidance. System components can be a human entity such as a doctor or a computing machine.

The mathematical model at the core of classic cybernetic systems theory states that:

$$X[k + 1] = A[k]X[k] + B[k]U[k] \tag{2.1}$$

$$Y[k] = C[k]X[k] + D[k]U[k] \tag{2.2}$$

Where  $X$ ,  $U$ , and  $Y$  are the system true state, inputs, and measured output vectors respectively.  $A$ ,  $B$ ,  $C$ , and  $D$  are matrices that provide the appropriate transformation of these variables at a given time  $k$ . The state system describes human health. The previous state and the inputs into the system play a role in determining health at time  $k+1$ . Inputs into the human cybernetic system are anything that changes gene expression or physical actions in the body (from molecular interactions to coarse action). Thus a person is continuously exposed to these inputs. The inputs beyond the control of an individual are external disturbances, and the rest are controllable inputs  $u$ . The true health state of an individual at a time  $k$  is represented by  $X$ . The health state is difficult to obtain and is always estimated. What we do get are the observable output variables. The state estimation challenge is in interpreting the observable data to understand the true underlying state. If we solely focus on this, the challenge of state estimation is in matrix  $C$ , with observable components as  $Y$  and our unknown state as  $X$ . This representation is visualized in Figure 2.2.

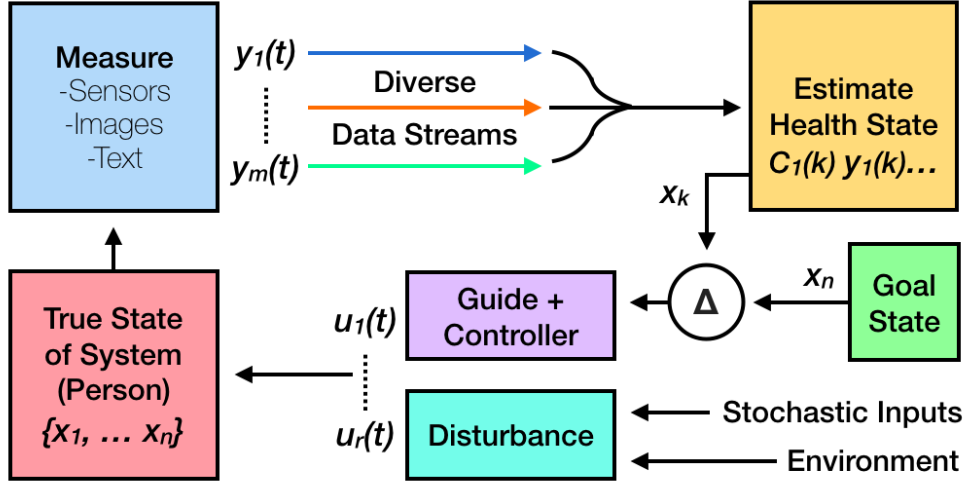


Figure 2.2: **Basic Cybernetic Loop.** Cybernetic control occurs through feedback loops. The loop begins with measurements from one or more sources, compiled into a computational representation of the system state. This estimated state is compared to the goal state. The delta between the goal and current state is used to give guidance to operate the controllers appropriately. These controllers partake in actions in the real-world to perturb the true system. Disturbances may also perturb the system from the environment or stochastic inputs. At the next time interval, this loop cycles again with a fresh set of measurements.

Two subsets of inputs determine the true health state of an individual. The first set of inputs is beyond the control of the individual, such as circumstances of where an individual is born and how they are raised, the genetic composition and biological processes occurring within the body, resources available, and the ambient environment. These inputs are usually not within the ability of an individual to manipulate. We consider these as noise or disturbances to the system, as shown in Figure 2.2 The other significant subset of inputs into the system that determines health is traditionally known as “modifiable risk factors.” Modifiable refers to an ability to decide the control input for the individual. This usually encompasses every decision moment the person lives but can extend to any situation in which a person has a choice (i.e., choosing to take a medication, buying a home in a particular neighborhood). Lifestyle modifications particularly are currently seen as the most promising approach for inputs that lead to healthier living [134]. This is how we will model controllable inputs in the above equation as  $u$ .

“No one notices when things go right.” This simple homily underlies the fact that 95 percent of the US healthcare economy is for direct medical care, and only 5 percent for health improvement [30]. This allocation of resources occurs due to the ongoing challenge of estimating the health state of individuals who seem to be healthy. The challenges arise due to 1) poor sensing ability to develop adverse outcomes with current human perception and methods, and 2) the long lead time for developing chronic disease. By the time current clinical measurements such as cholesterol, blood pressure, or glucose metabolism are beyond the normal range, the user has already been in a dysfunctional health state for quite some time. Clinical researchers refer to this as the *prodromal* state. Capturing the change in health state always (before significant dysfunction begins) is paramount to keeping people healthy states and preventing them from slipping into a diseased state.

The target of this thesis work aims at structuring data and knowledge so we can have continuous HSE. It is essential to consider the representation of the system state, as all computing components of the loop depend on this representation. In general, representation of the state and state space in addition to searching through this state space are two key fundamental challenges in any intelligent system [313]. In simple control systems such as thermostats or even complex systems such as military jets, the representation of the system is quite well understood [69]. A thorough investigation of *defining what the true health state is* must be the first step towards building a robust and universal computational health state representation.

## 2.2 Defining Health

There is a multitude of definitions of health. We will first review two general approaches to health, the first being continuous versus classification approaches and the second being objective and subjective approaches. Following this, we take the various lenses of health definitions

and amalgamate central principles of each. These lenses include biological, environmental, functional, and perceptual views on defining health. In order to build a comprehensive and robust definition of health, we must consider all aspects for a complete understanding. This work will primarily focus on the integration of these definitions into a single unified approach at the individual level definition of health.

## **Classification vs Continuum**

Before 1948, the WHO defined health as the absence of disease or infirmity [301]. The original definition indicated that health is static or unilateral (disease or no disease). This method is a classification approach for having a definitive understanding of health. Medical systems and approaches in modern-day continue to classify health based on disease, conditions, and various metrics to define the state of one's health. For example, coronary heart disease, high blood pressure, and heart failure are all cardiovascular diseases that are considered diseases based on a range or threshold. These thresholds are recorded in the International Classification for Disease, which is used as the universal basis for the health field [209]. It has the word *classification* in the middle of the name. If an individual does not meet the threshold, they do not have the disease. This has large implications. As an example, the American Heart Association formed ranges for cardiovascular risk, and medical professionals take into consideration ranges to determine if an individual is normal, elevated, or at risk, as shown in the blood pressure diagram in Figure 2.3. The thresholds are usually pre-determined for the entire population, assuming in the model that each person is identical to another and represented by a static system.

On the other hand, some classification approaches rely on the presence or absence of an attribute. A few examples include how an individual can either be positive or negative for an infection, have had a pregnancy or not, has an artificial limb or not, or has a chromosomal set of XX or XY. In summary, the classification approach to health is a combination of two

subsets, the first subset of classifications use thresholds, and a second subset is similar to Boolean or categorical attributes [285].

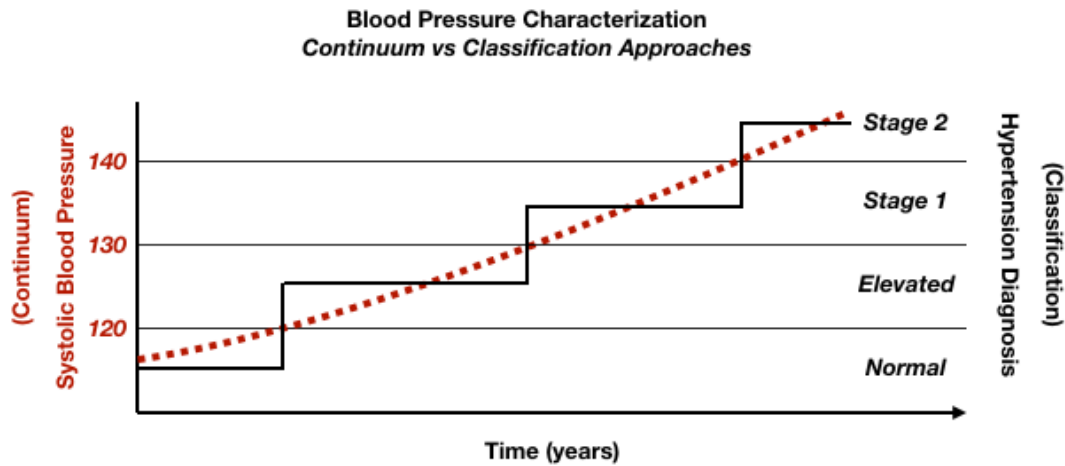


Figure 2.3: **Blood Pressure Characterization.** Doctors use a classification approach with threshold values to have predetermined categories of defining health. The nature of health operates on a continuum basis, as shown in red. The state of resting blood pressure for each individual is always changing over time dynamically. Regardless of the approach, the individual is unable to sense their state position at any time, highlighting the need for continuous sensing.

In 1948, the WHO updated the definition of health to encompass health as a “state of complete physical, mental and social well-being” [301]. This shifted the thought towards defining health on a spectrum beyond disease and takes into consideration the full physical, mental, and social state of an individual. The WHO did not go further to quantitatively define what this means. In 1972, a physician named John Travis proposed a concept of health, referred to as the “illness-wellness continuum,” which states that a medical approach relying on the presence or absence of disease to demonstrate health is insufficient [265]. This concept introduces the idea that an individual falls somewhere on the spectrum of dysfunction and wellness. Where they are at each moment would be their health state on this single axis. The states in which dysfunction is perceivable is reflected by signs, symptoms, and disability. This approach to health is the “health continuum,” also described in chapter 1 Figure 1.2. The health continuum implies that individuals would wish to strive for a higher level of



optimal living through the means of health, not just as a unilateral classification of being labeled with a disease. This modern outlook of health acknowledges its dynamic nature and subject to constant change within an axis of illness and wellness. “Both health and illness are dynamic processes, and each person is located on a graduated scale or continuous spectrum (continuum) ranging from wellness and optimal functioning in every aspect of one’s life, at one end, to illness culminating in death, at the other” [256]. However, it is important to note that the continuum may be interpreted as static if observing a single point in time or location within the spectrum. Regardless, this concept proposes a single dimension concept of state and paves the way for a continuous and quantitative approach.

Clinically, resting blood pressure illustrates the continuum approach. People are unable to feel the biological change as it happens over time, and thus do not even know they are being affected. Second, the blood pressures vary due to a complex impact from both daily actions and environmental exposures, not just a single source or pathogen. Third, these diseases are not truly categorical in nature but are declines in physiologic function over time. In the example of hypertension, the medical practice uses cutoff thresholds to decide when to change the labeled blood pressure status of an individual, when in reality, the average pressure is increasing over time, as shown in Figure 2.3. The thresholds provide a basic region mapping of the blood pressure health state, but this does is not based on a precise model for each individual, instead of a population model.

## **Objective vs Subjective**

In the literature, health has an expanded definition through the introduction of objective and subjective components. Objective components are quantifiable and externally visible, and subjective components are within an individuals internal state of mind. The objective states have four components: the **biological** ability to exist, the **environmental** impact and handling of this impact on an individual, and the ability or capacity to **function**. The

<b>Author</b>	<b>Biological</b>	<b>Environmental</b>	<b>Functional</b>	<b>Perceptual</b>
Dubois	X	X	X	X
Dunn		X		
Galen	X	X		X
Hippocrates	X	X		
Lalonde Report	X	X	X	
Ohlansky	X		X	
Institute of Medicine	X			X
Parsons			X	X
Sartorius	X	X	X	X
Simon et al.			X	X
Spencer		X		
Stokes et al.	X	X	X	X
Travis			X	X
World Health Organization	X		X	X
Nag	X	X	X	X

Table 2.1: **Literature Review of Health Definitions.** Throughout health research history, various attempts have been made to define health. Here is an overview of these attempts.

subjective component generally is understood through the individuals **perception** of their quality of life.

Various reports integrate objective and subjective components as a critical part of defining the scope of health:

a) Two prominent figures had established definitions of health used in modern medicine today. Hippocrates of the fifth century BCE, considered to be the Father of Modern Medicine, established the relationship between environmental and personal cleanliness and the origin of disease. He had classified health based on disease, resulting from internal physical imbalances. In the Roman Empire, Galen expanded on Hippocrates definition by stating that bodily fluids also determined temperament and personality, [17]. These combined outlined health as the idea of physical, environmental, and behavioral factors.

b) The updated WHO definition to expand the realm of health beyond disease and include mental and social well-being: Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity [301].

c) The 1974 Lalonde Report from Canada introduces the concept of “health field concept,” in which identified determinants of health is through lifestyle (personal decisions), environment, human biology (all components of the body influenced by genetics), and health care systems [90].

d) The Croatian Medical Journal describes health as three types: 1) absence of disease, 2) a state allowing the individual to cope with demands of daily life adequately, and 3) a state of balance, an equilibrium that an individual has establishes within himself and between himself and his social and physical environment [241].

e) The Center for Educational Development Health (CEDH) at Boston University and the Association of Teachers of Preventive Medicine Foundation (ATPMF) define health as a state characterized by anatomic integrity, including “performing personally valued family, work, and community roles; ability to deal with physical, biological and social stress; a feeling of well-being; and a freedom from the risk of disease and untimely death” [252]. Furthermore, the health status estimates of individual’s health from “one or more anatomic, functional, adaptive, and subjective indices.”

f) A qualitative study by Simon et al. understands health through an individual’s self-assessment, which is largely linked to one’s physical health, functional capacity, health behavior, and other psychological components. Health also encompasses other aspects including health comparison, health transcendence, externally focused, non-reflective, social role activities, and social relationships [248].

g) U.S. Institute of Medicine Committee on Health and Behavior: Research, Practice, and Policy in Washington, D.C. describes the need to integrate all the above approaches. “Ac-

ceptance of the fact that stress is linked to cardiovascular disease or to other health problems has become commonplace. However, research also reveals many reciprocal links among the central nervous system, which recognizes and records experiences; the endocrine system, which produces hormones that govern many body functions; and the immune system, which organizes responses to infections and other challenges.” The Committee illustrates how perception and biological health are tightly coupled, hence the call for an integrative approach to defining health [132].

The above approaches are compared in Figure 2.1. In the following sections within this chapter, we will first explore the individual components of objective health in biological, environmental, functional aspects. The subjective and social components of health will be discussed as a perceptual component. It is also important to note that there are large health systems that are part of the ecosystem that contributes to health for the individual.

### **2.2.1 Biological**

*Biological health is the sustenance of life through the homeostasis of natural systems. Working biological health supports the basic living functions of life. The literature consensus of what these tenets are: Organization of Structure, Metabolism, Homeostasis, Growth, Reproduction, Response to Stimuli, and Evolution. The only main contention in general literature of living biology is in the case of viruses. Viruses do not conduct their metabolism and thus do not maintain any homeostatic state. They also cannot reproduce on their own.*

Human biological health has been quantified by looking at lifespan or by the calculation of healthy aging and life expectancy [34]. The 2010 Global Burden of Disease Study provides quantifiable health metrics for calculating healthy life expectancy. Lozano et al. estimated life tables and annual number of death for 187 countries from 1970 to 2010. The estimation included number and rate of deaths by age and sex to determine the burden of disease in order

to constrain estimates of cause-specific mortality [161]. A look at lifespan is only a primary attribute of biological existence. The biological understanding of a human is understood more deeply by the genetic composition and biological processes occurring within the body. Illness or disease negatively impacts the overall homeostasis of biological processes and therefore reduces the lifespan.

With stress from its surroundings, an individual must remain stable internally for its biological success. This stress is the particular reason behind the “response to stimuli” factor of biological health. “Under normal conditions, the external environment changes constantly in an unpredictable manner. In order to survive and to continue to function effectively the organism must make adaptive responses to the modifications. It must, as well as possible, repair destructive tissue damage and restore its own internal environment” [70]. The ability to adapt and hold resilience is therefore critical for survival when introduced to external stresses or threats, such as disease. Biological health can, therefore, include the definition as “the perfect, continuing adjustment of an organism within its biology to its environment” [306].

Biological health operates through the complicated yet harmonious interaction of non-living molecules to maintain the steady-state of living functions. A narrow slice of this complex interaction network can be visualized through the basic chart of metabolism, as seen in Figure 2.4. The health of this network to sustain itself in a stable state can be considered the generalized approach to biological health. A full biological health state framework would require a system that can incorporate all the abstraction levels from atomic to the whole organism, as shown in Figure 2.4, and the ability for computation at any level of choice.

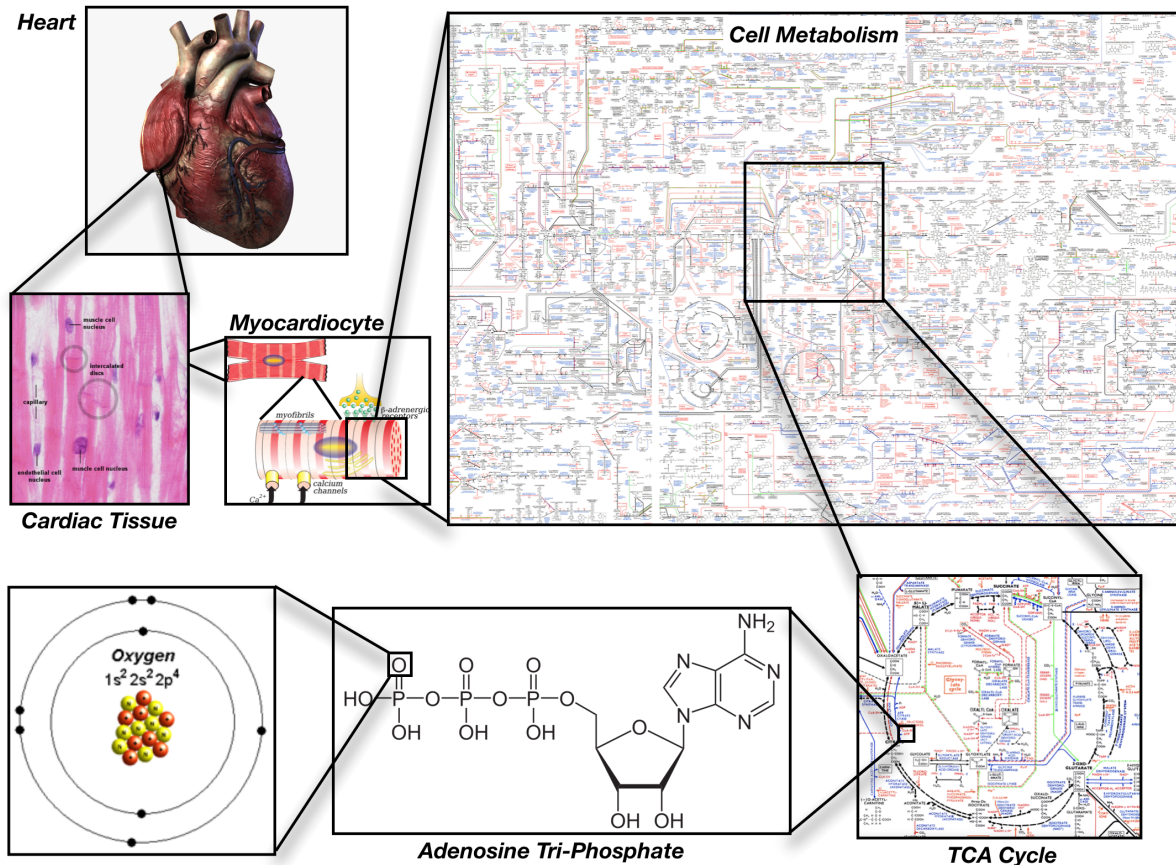


Figure 2.4: **Biological health is at the level of inanimate molecules interacting to provide the organism with the fundamental components of life.** Metabolic pathways provide the methods by which energy and materials can be synthesized to support these components. Here is an example of biological health at various abstraction levels. We zoom into the heart. There is the tissue of the heart, each heart cell (myocardocyte), the metabolism network in this cell, a particular central pathway called the Tri-Carboxylic Acid Cycle (TCA Cycle), Adenosine Tri-Phosphate (ATP) which is an energy currency in biological systems produced by the TCA Cycle, and the oxygen atom within this ATP molecule.

## 2.2.2 Functional

*Functional health is defined as the performance capacity to provide a service or utility.* This definition applies to any abstraction level on the biological spectrum. At a whole human level, the function may be how well a person can run. At an organ level, the function can be how well the heart circulates blood, and at a molecular level, it may be how well a cell

regulates a specific hormone such as insulin. Functional health assumes that health is a resource that allows individuals to be able to use their bodies and minds to perform specific tasks. The function abstracts away biology from health. A biological bone and a titanium bone both provide a service of structure to the body. A natural pancreas and artificial digital pancreas both provide a service of glucose regulation. A biological sinoatrial node in the heart and an electronic pacemaker both provide heartbeat regulation as a service. Functional health fully encompasses biological health, but can extend far beyond biological systems for defining health.

Aristotle's "Four Causes" is a realm of thought that explores the reason for existence [81]. This concept can help answer "why" our biological health truly matters. In Aristotle's telos philosophy, the end goal or purpose of anything is what function comes about. We can use this in making a parallel to how health is ultimately a utility function. Figure 2.5 visually describes his classical example of the "cause" of a table. Material cause refers to the physical matter that makes an object. Wood makes up a table, and molecules make up a human. Formal cause refers to the design of an entity. A table may have a simple design, and a human has an intricate design of interacting parts. Efficient cause refers to how an object comes into existence. Carpentry makes the table and reproduction and cell growth from a single cell (zygote) makes up a human. Final cause, also known as telos, refers to the reason for existence. A table exists to dine upon, and a human exists to exhibit certain capacities such as movement and thinking.

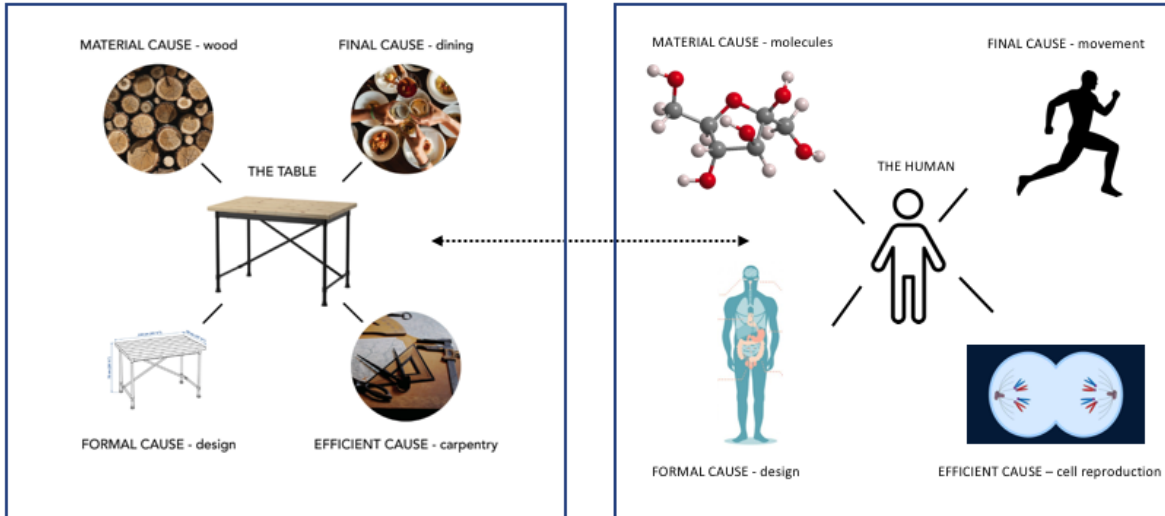


Figure 2.5: **Aristotle’s Four Causes.** Aristotle’s Four Causes refers to the various lenses we can understand why something exists. Aristotle argues that the final cause is by far the most important since without it, the other causes may not have a direct purpose to exist.

Both Aristotle and modern-day organizations, like WHO, have contributed to the evolution of defining health. The WHO identified health as a “resource for everyday life, not the objective of living.” Health includes social and personal resources, as well as physical capabilities. in the first International Conference for Health Promotion in 1986 [297] Furthermore, it encompasses “the *ability* to lead a socially and economically productive life.” Health is also viewed as the *ability* to perform valued social roles, emphasizing functional views and the concept of health to include personal beliefs and values by a society [252]. This means that a loss of a leg is grieved just as much because this means a loss of mobility, and thus, a loss of self. If an individual were to receive a prosthetic for a leg as a replacement, the mobility would be maintained and fully functional. However, the biological basis would alter to a mechanical one. The same can be said for porcelain teeth-replacing biological teeth for cutting food. However, to an individual, there is functional value due to the mobility or cutting and, thus, would no longer be a loss of self.



While the maximum lifespan may have increased, there may not be enough emphasis on the healthspan of the population. While there is a longer lifespan by reduction of child mortality rates, elimination of diseases through modern medicine, the human biological design is not for long-term use (i.e., maintaining healthy functioning knees, heart) [208]. This increased lifespan now becomes an issue of function; to address the functional component there must be an emphasis on the healthspan and the quality of life, not just the length. The WHO has quantified the Disability Adjusted Life Year (DALY) as one lost year of healthy life due to disability or poor health [210]. DALYs for a disease or health condition are calculated as the sum of the Years of Life Lost (YLL) due to premature mortality in the population and the Years Lost due to Disability (YLD) for people living with the health condition or its consequences. This metric indicates that disability or poor health is a loss of function that ultimately leads to a shorter lifespan. Individuals may also define themselves as healthy because they are functioning as well as expected in their age and gender group of the general population. Functional health is, therefore, interwoven with perceptual health.

With living comes the capability to respond to stressful situations. There is a functional capacity to cope and be resilient to the demands of daily living [241]. From a cybernetic perspective, this is the stability of the system to perturbation to the external environment. The diminishing function of an individual in their ability to perform tasks that they would have otherwise been able to can be directly related to disease or illness, much of which can be caused by the surroundings.

### **2.2.3 Environmental**

*Environmental health encompasses external factors outside of an individual that impact the state within the individual.* The environment plays a significant role in the exposures that an individual has over time, and thus is inevitably a major component of one's health. Dubois

explains that “health and happiness cannot be absolute and permanent values, however careful the social and medical planning. Evolutionary fitness requires never-ending efforts of adaptation to the total environment, which is ever-changing” [258]. Because of this continuous change in an individual’s environment, this theory supports the health continuum method in which health is dynamic. While the environment is continually changing, and thus human health is continually changing, humans attempt to create equilibrium within themselves and the social and physical environment via self-maintenance systems [241]. This theory supports that health relies on an adaptation mechanism to the external environment to produce homeostasis at various layers of biology and function.

The Health Grid Model was developed by Dunn et al. to show how external factors impact the quality of life [71]. The graph below shows peak performance as an optimal state of health based on the individuals age and makeup but highly linked to a favorable and unfavorable environment. Dunn called for a more integrated approach to human health via the integration of environmental factors. There may be favorable and unfavorable environments that may positively and negatively impact health. Individuals may be exposed to various risk factors and inputs to their health through the environment (i.e., pollution, water sanitation, chemicals, occupational risks, climate change, radiation, agricultural practices). Individuals may also have exposures to external risk factors through their social environment (i.e., family, friends, society, decisions based on social circumstances). The social environment that an individual surrounds themselves with has a significant impact on daily lifestyle factors, for example, food choices and physical activities.

Understanding an individual’s environment furthermore gives the livability of the location of how healthy a particular area may be for the individual. The Economist Intelligence Unit’s Where-to-be-born Index and Mercer’s Quality of Living reports are measuring the livability of a given city or nation [178, 73]. This environmental component is a combination of subjective measurements such as life satisfaction surveys and objective determinants such as divorce

rates, safety, and infrastructure. These studies review the livability for populations at large but directly links to predicting an individual’s health state based on a chosen environment. A part of this local environment includes any system built in regards to health.

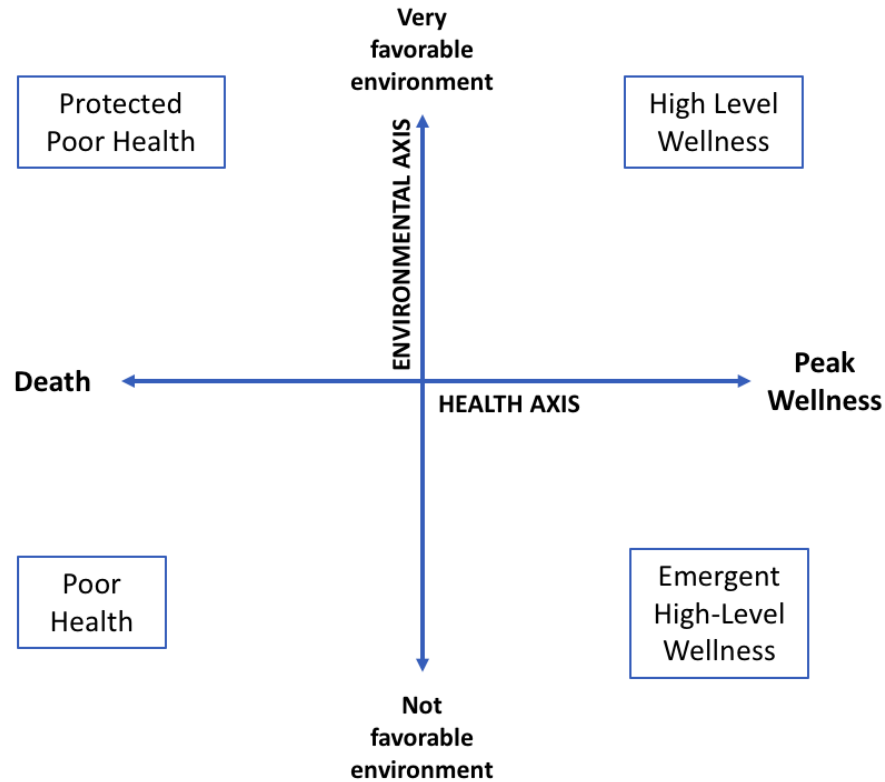


Figure 2.6: **Dunn’s Health Grid Modeling: External environment and its relationship to wellness.** Dunn’s model of health highlights the strong relationship between a favorable environment and achieving peak wellness. If a good environment or good health exist exclusively, one cannot achieve an optimal level of health.

## 2.2.4 Systems

*Systems health is the overarching structure for providing health to individuals at a population level.* Systems for providing health services largely contribute to the outcome of health for individuals. The effectiveness of these systems is tracked by reviewing populations at large. Health systems play an economic role in providing a positive return to individuals. Williamson et al. conceptualize health as a type of capital that invested in many individuals

and institutions for positive health returns that reflect higher individual capacity. Health cannot be exchanged or sold for goods and services. Instead, “as a type of capital, health is a stock of biopsychosocial resources that people can draw on to participate in society [297]. A modern approach to health, such as value-based healthcare, puts what patients value at the center of healthcare. It helps ensure that they receive the care that can provide them with outcomes they think are essential and that limited resources are focused on high-value interventions. In order to do this, there is a need for flexible definitions of health, personalized and tailored to patient values [88]. Another system approach that is becoming more prevalent is the continuum of care. The continuum of care concept involves an integrated system that guides and tracks a patient over time through a comprehensive array of health services spanning all levels of intensity of care. The array of services includes promotion, prevention, treatment, and recovery as an individual goes through the process of receiving healthcare services [237]. While systems are of relevance in defining health, the economic and population level of health will not be part of the scope of this work. Instead, the focus will be on computing how a system may impact an individual as part of their environment.

### **2.2.5 Perceptual**

*Perceptual health is an individual’s outlook on their health state and judgment of their quality of life.* The view of one’s self has become a large component of subjective determinants of health, including one’s emotional state, social connectedness, intellectual stimulation, self-awareness, and development of lifestyle behaviors. With perception is an individual’s ability to think and feel in the way of good health, as understood by their internal state. In addition to “anatomic, functional and adaptive dimensions, much of health is simply about feeling good” [252].

Health is of relevance to the individual concerning their defined goals and value in living. Aristotle's Telos philosophy indicates there is an end or purpose, in "teleology" or the study of purpose within biology. Telos, which means "end", "purpose" or "goal" as shown in Figure 2.5. Dubois in the *Mirage of Health* has also defined "good health as the condition best suited for each individual to reach his or her personal and social goals" [258]. The concept of personalized goals and values in living becomes critical to the state of health because it links to how an individual perceives the state of their life. Individuals act and function in alignment with their own values. Stokes has described Health as with the incorporation of the "ability to perform personally valued family, work, and community roles" and dealing with psychological and social stress [252].

Kimball et al. at the University of Michigan, along with prominent thought leaders such as Janet Yellen and Daniel Kahneman, put together a concept of *Utility and Happiness* where they connect how the perception of living happily relates to the utility for the individual [145]. They have separated short-term happiness as "elation" due to relative change in utility and long-run happiness as the "baseline mood" for the utility. Thaler's work in connecting behavioral psychology with economics shows how the "baseline mood" (i.e., influenced by sleep, exercise, eating habits, genes, time with friends) is linked to the true health utility [145]. Both of these examples show how the quality of life is related to psychology bridged with economic principles of utility to an individual.

Increasingly described in the literature are understanding ones self, motivations, and goals as a foundation for health. The health continuum also highlights that awareness of self, education, self-growth, and adopting healthy lifestyles are critical in moving towards the wellness end of the spectrum and reaching a more optimal level of health. In addition to values, self-perception is another measure for differentiating a healthy state from a diseased state and how this can be dependent on an individuals outlook on the situation and how they feel internally (or subjective well-being) [31]. Physical and psychological wellness of an individual

plays a role in the health continuum in the constant strive for improvement. The domains of wellness can include social (lifestyle behaviors), emotional (stress, resilience), physical (activities), spiritual (meaning, purpose), occupational, intellectual components [131]. The combination of these domains creates a wellness state that incorporates both the mind and the body.

Maslows “Hierarchy of Needs” indicate that humans are built upon physiologic needs as a baseline, but are ultimately striving for self-actualization. This approach sees physiologic needs as the core need to be satisfied before needing safety, love/belonging, esteem, and self-actualization. This work is quite prevalent in the literature in psychological sciences.

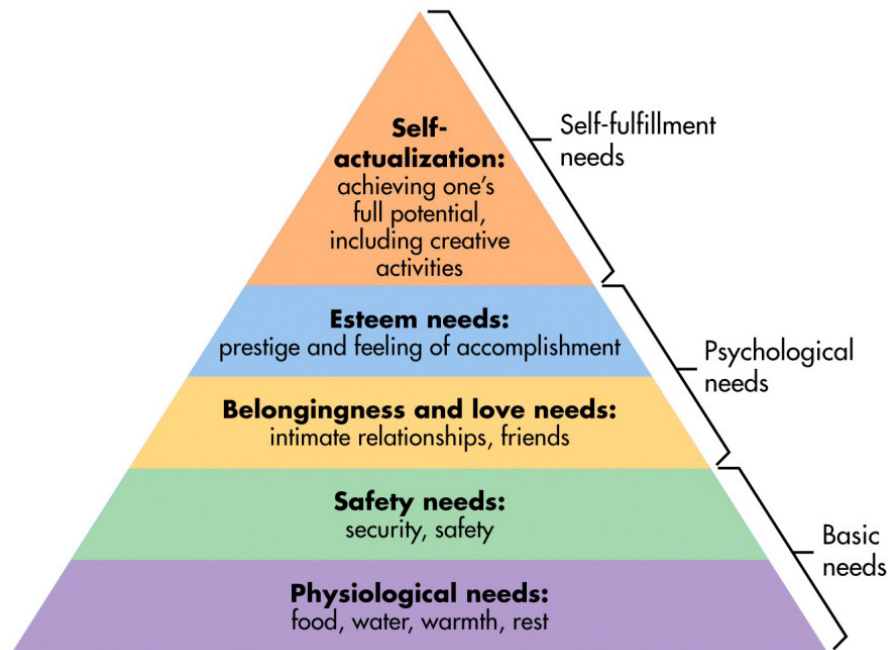


Figure 2.7: **Maslow’s Hierarchy of Needs.** There are base physiologic and safety needs (food, shelter, security) as a fundamental requirement, psychological needs such as a sense of belonging and esteem, and finally needs for self-fulfillment such as reaching one’s full potential.

The Iceberg model of health and illness, as presented by Travis in the illness-wellness continuum [265], describes the state of health to be built upon the foundation of self-understanding, psychological motivations, and lifestyle and behavior. For both Travis and Maslow, the pre-

sensation of health state regards perceptual components as integral considerations that layer on top of high-functioning biological machinery. In this perspective, biological health directly impacts the perceptual health. On the flip side, perceptual health can directly impact biological health. As shown in Figure 2.8, mental states directly impact the biological components of the entire rest of the body. Thus, there is a need to build a network representation of these two views with a bi-directional relationship. The body impacts the mind, and the mind impacts the body.

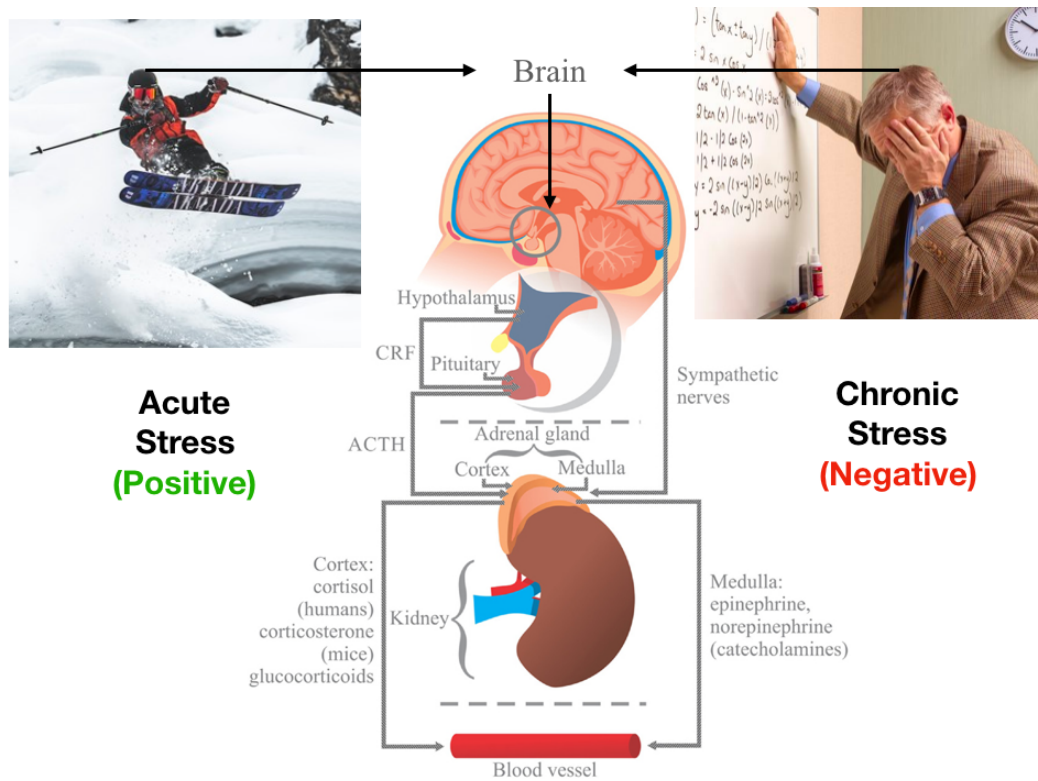


Figure 2.8: **Mind - Body Connection: How Stress Impacts Physiology.** How a person perceives their environment has a direct impact on their physiology. When an individual perceives stress, the evolutionary encoded response of increasing blood pressure occurs to deliver high-performance blood delivery for a “fight or flight” response. This can be both a positive or negative impact on the health state, depending on the context. For example, a skier negotiating a racecourse may have stress from turning at high speeds. This stress allows the muscles to function at high forces and serves as a positive acute stressor. Blood pressure returns to normal after the race. In contrast, a professor who is worried about getting funding approved may have this stress effect chronically rather than acutely. In this context, the stress response may damage parts of the vessels and organs. Thus stress is a highly context-dependent component of computing the health state.

We can also quantify the perception of how an individual perceives themselves and their quality of life (QoL). QoL is a term for the quality of various life domains and is measured as a standard that considers what a good life is for an individual. Many of the factors determining the quality of life may be driven by values and goals and encompassing of emotional, physical, and social well-being. The factors ultimately determine ones life satisfaction or how an individual feels about his or her life. Measurements can be seen globally in the World Happiness Report of 2019, which explores how happy citizens consider themselves to be, which directly impacts their emotional and social well-being.

Ultimately, all the above definitions of health mention important components of health for an individual. We must consider these definitions in a generalized computing framework for computing the health state. These components are also highly interdependent, and this must be clear in the representation. The next phase of research describes how we can pull these components together in a common framework.

### **2.2.6 Integrating a Holistic Approach**

As an individual, we want to consider all aspects of our health when making life choices, a holistic approach. For this reason, this thesis uses an integrated approach to the definition of health by integrating the four components as described previously: biological, environmental (including systems), functional, and perceptual. All four components contribute to the quality of living and a total health state computing framework. We will utilize the approach of the health continuum view and recognize that health is evolving and dynamic by nature on this continuum. This method still allows for classification approaches through thresholds and categories. There is also a nesting of various layers that build a single individual model. Nesting may occur for various domains such as biology, as shown in Figure 2.4. Depending on the abstraction level from a molecular level to a whole person level, various connected



components rely on the functioning of the lower-level components. Measurements may come at various layers of abstraction, and the user or domain expert such as a doctor or intelligent agent may want to perturb or interact at a particular abstraction layer. A holistic model must accommodate this ability to zoom in and out in a fluid manner. Coupling these concepts also requires an understanding of what causal orientation they are related to each other in the real world. The dissertation goal is to bring all components of defining health together holistically in order to best estimate the health state. In the next section, we go into the depth of how this holistic approach translates into detailed design requirements.

## 2.3 Design Requirements for Health State Estimation

After considering the nature of different health definitions in this chapter, and the emerging frontiers in chapter one, we can establish the following design principles for a health estimation framework that can support a variety of complex applications. We will describe these requirements below.

1. **Personalized Modeling.** To predict the future, provide guidance, and understand the preferences and particularities of an individual, researchers must build personal models. These models, which establish the premise that each individual is a unique system, are needed to best estimate the users health state and how various inputs uniquely move them in the state space. Models are accessed for state estimation or providing guidance. The personal model can build knowledge and predict many aspects of an individuals life, such as how an individual reacts to different stimuli under specific conditions or physiological changes from an intervention. To model a person, a combination of long and short term information must interact. Long-term models of an individual can come from the genome or the history of event patterns and behavior. Biological or high-resolution continuous sensors can capture short-term

models. These models are not static; they continuously change with an individual's age, life events, and other life parameters, which makes model-building a dynamic and causal understanding process. Models must have a general representation that can at least address the following health attributes:

*Physiologic function.* Dynamic health attributes include biological health, such as organ function and capacity (CRF, bone strength, and metabolism). Models may have various time and size levels in terms of detail, depending on the application. Short-term models can already predict how electrical activity in the heart may produce arrhythmia (such as in Figure 1.4 while a long term model may predict the impact of long-term lifestyle habits on developing atherosclerosis. Cellular or gross organ-level processes can be modeled separately or in combination, as shown in Figure 2.4. These functions constantly change based on daily factors that include sleep, nutrition, medication, or stress.

*Psychological function.* Extrapolating and modeling mental health will be a complex challenge that requires advanced sensing mechanisms such as affective computing. Mental health components could include various time resolutions of interest, such as overall mental stability or minute-by-minute affect analysis. Evaluating risks for erroneous modeling before field deployment will be essential for the user's safety. Understanding user preferences, influences, and motivations will be essential to close the loop when humans are involved as actuators in a participatory approach. The system uses this information to tailor guidance in a way that makes a healthy lifestyle pleasurable and convenient, enhancing the quality of life. For effective modeling, irrational behavior, social tendencies, and situations in which short term pleasure overrides long-term goals must be included. Risky behaviors may warrant preventive help to reduce harmful events such as drunk driving.

*Risk.* Researchers can model risk assessment for reversible and irreversible physiologic dysfunction, undesired events, or regions of the health state space a person wishes

to avoid. *Disease risk* refers to the health state entering an undesirable region of the state space. An example would be the chance that an individual would develop metabolic syndrome, a condition that medical professionals could reverse with an early-stage intervention. *Damage risk* refers to the risk in which an individual is negatively permanently altered. An example of this would be the likelihood of metabolic syndrome turning into type 1 diabetes, resulting in permanent pancreatic damage. Risk modeling will also include behavior and mental condition models to understand the propensity for self-harm, accidents, or harm to others, as mentioned previously in the psychological function section. For the greatest value, we must prioritize resources and investments based on evaluating the likelihood and consequence of combined risks. Risk models of the individual are directly related to prediction and prevention.

The personal model is essentially a virtual twin of the individual, which allows for modeling the current and future state of the person in various aspects of their biology and psychology. The model's state space must be compatible with continuum and classification approaches.

2. **Predictive and Preventive Capability.** The above model should be able to simulate what the health state will be at various future time points. Given a set of chosen inputs, we can test various scenarios to perturb the health state model to see how the current state changes. These simulations should be able to help predict a trajectory path to compare which input set would produce the most desirable health state for the user. A prediction that falls in an undesirable region can advise preventive actions to steer the state away from the region.
3. **Precision through Understanding and Learning.** A health estimation framework inherently will have errors in the outputs that may not be compared to a ground truth measurable factor, making it challenging to compare various models and judge performance of the model, state estimation, and predictions. The system must have

a method by which to give an estimated error range so the user or computer can understand the quality of the estimation. Precision should improve in the model and estimation with increasing information. The two main approaches for this would be bottom-up data-driven methods and top-down knowledge-driven methods.

4. **Participation with User.** The system must be able to interact with various types of users. These users include the individual themselves, any healthcare professionals, domain and research experts in any respective field, or other computing clients that access the health state for further services provided to the user. Facilitated human interaction with the framework requires representation in a tangible human-understandable form to bridge the gap between the physical and cyber worlds. The HCI should have easy to understand interfaces, a simple way to query the health state and model, and explainable to the human. Developers that wish to create software using the health state and personal model should have secure yet easy access to Software Development Kits (SDK) that can create such HCI and services.
5. **Encompass Holistic Definition of Health.** As we explored above, there are various lenses and dimensions all part of a total health state for an individual. Any complete HSE framework must be able to include these lenses. Biological mechanisms give rise to functional and perceptual experiences and how the environment and systems impact this. Systems can be clumped as part of the environment. Based on the work presented by Kimball et al., we can also view perception through a utility lens, and thus can model this as a functional component of health. Thus, the requirements for the holistic definition must include the biological, functional, and environmental aspects.
6. **Capture Connectedness.** As shown in Figure 2.4 and Figure 2.8, health state is tightly connected to many different parameters in a networked web. This connectedness must be captured to give the best quality estimate of health.

7. **Causal and Explainable.** Given a large number of connections, many of these may be spurious or artifacts of data. Understanding why something is connected and characterizing its true effect is key for a high-quality system. For trust by users, scientists, health experts, and computation outcomes, this explainable aspect is also essential. Many modern artificial intelligence techniques lack in this requirement [103]. Current approaches in using data and machine learning to estimate health are falling short due to their non-understandable black-box nature. Knowing why an individual has a certain health status will be paramount for explaining why a recommendation or treatment is best. Purely data-driven methods can be components in a large system but will have a difficult time explaining the reason for the classification without the incorporation of relevant domain knowledge.
8. **Domain Knowledge Integration.** The best source of understanding causal relationships and connectedness rests within the vast array of domain knowledge already in the literature. Leveraging this domain knowledge is essential, especially as a cold start, to build an explainable system. Any new domain knowledge discovered should also have the ability to integrate. Hard-wiring the domain knowledge is thus not an option. Domain knowledge also is organized at various levels of study (molecular to societal sciences), and the best approach would be to take advantage of the vast knowledge in all these layers.
9. **Zoom Fluidly Between Abstraction Layers.** As mentioned in the previous design requirement, the ability to zoom in and out of abstraction layers should be part of the system. This gives domain knowledge various entry points to participate in the framework and allows for the user or domain expert to understand a layer of interest. Looking at Figure 2.4, the framework should allow jumping between any box layer of interest, while still preserving the connectivity.

10. **Flexible and Modular Architecture.** Because health can be abstracted away from biology through the functional lens, the framework must allow for flexibility through a modular approach. Imagine a patient who receives a kidney transplant. This new kidney has a whole new set of genetics and biology from the patient's original kidney. If the kidney is replaced with a mechanical kidney, this would have an entirely different operational structure. A flexible approach would allow the appropriate module, which modeled the kidney, to swap out for the current one. This ensures that no matter how technology adapts and situations change, the framework continues to be robust for the future.
11. **Diverse Data Fusion and Transformation.** In addition to taking advantage of domain knowledge of all abstraction layers, the system should be able to take advantage of any data produced about the individual. This could include a vast array of sources, types, and frequencies of data and event streams. Any type of data should be able to get processed in a way that is then useful to update the health state or improve the model. By acquiring data from different types of sensors, we can create a more precise understanding of how to model the individual health state. This process needs to collect, aggregate, and process heterogeneous data from multiple sources to infuse into the model. Data transformation methods are needed to convert data to a representation that is suitable for digestion into the personal health model framework.
12. **Perpetually Updating.** As pointed several times in the last two chapters, the health state is in a dynamic flux. With data and domain knowledge, the system needs to continually update the health state of an individual regardless of the situation to give the current state. Similar to how a GPS is always refreshing its alignment with the signal satellites. As an example, the highest performance GPS receivers can update at 50 Hertz (1 Hertz = 1 cycle per second). For the health state, various applications will require a minimum update rate in order to perform.

Using these above requirements for building a state estimation framework, we will dive into the current day literature to see how the research community is approaching these requirements. We will review the various methods for collecting data, approaching state estimation, and constructing a model for individual health.

# Chapter 3

## Literature Review

Science is organized knowledge.

Wisdom is organized life.

---

*Immanuel Kant*

Given the goal of this work is to have a model for computing the individual health state, we explore a literature review of methods for modeling and state representation on human health. The organization of the literature review includes four sections. First, we look at an overview of methods for personal data acquisition to obtain measurements about an individual. Second, we review how to estimate the health state from various measurements. Third, we survey models developed for representing human health. Fourth, we look at relevant techniques by which models can be made intelligent and dynamic. The chapter concludes with challenges that remain in the field of HSE and how this dissertation aims to address these challenges in a novel method.



### 3.1 Health Data Acquisition

To begin any endeavor in estimation, we must first gather measurements in the real world that we can use towards computing the estimate. These measurements are gathered through a sensor. Sensors bridge the computing world with the physical world, as shown in Figure 2.1. The cost and accuracy of sensing mechanisms vary, so data stream continuity and availability will always be different depending on the modality. For example, a wearable heart-rate monitor is an inexpensive and continuous way to measure heart rate but will be less insightful than cardiac perfusion testing with a computerized tomography scanner to evaluate heart attack risk or tissue oxygen delivery. Intuitively, an estimation system that can ingest the most types of data streams will perform better due to capturing a more well-rounded picture of a given situation. This highlights the need for the estimation framework to have a versatile ability to ingest data types. In order to ensure the constructed estimation framework best meets this need, we will review a wide variety of literature demonstrating how personal health data can be acquired through sensors, as summarized in Figure 3.1.

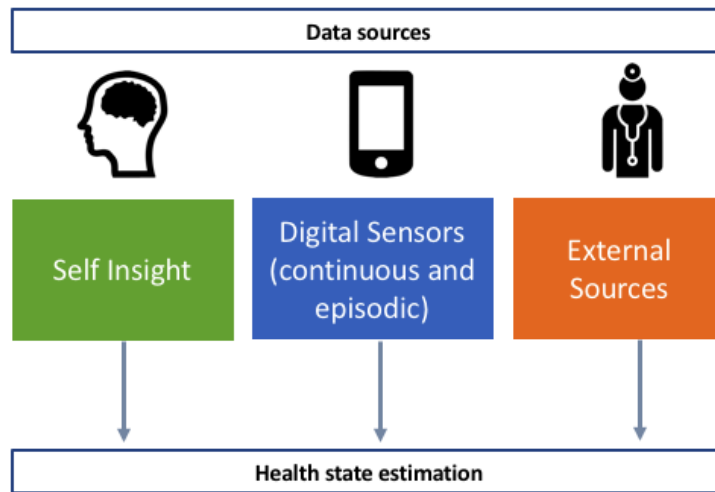


Figure 3.1: Sensor categories for acquiring health data.

### 3.1.1 Self Insight

The most primitive and natural sensor is the person’s biological signals and cognition. Two sets of signals influence our understanding of our own mental and physical health. First, **sensory organs** capture sensory signals about the world external to our skin through sight, hearing, touch (including sensory pain, temperature, texture), taste, and smell. These natural sensors measure various attributes to understand the individual’s physical environment or directly perceivable attributes. Internal to our skin, we then use **internal cognizance** to make interpretations about the external signals and integrate them with signals about our biology. The interpretation occurs through feelings of perceived pain, gut motility, energy, proprioception, and mood, to name a few. Humans use both these abilities to understand their health state and how their environment is impacting them. The ways our natural biological sensors unify, store, quantify, and use these data are not easily compatible with modern computing. Tedious methods that do not capture data continuously, such as questionnaires and surveys, are used to capture these signals in most scientific fields that study these signals. We go over traditional and emerging techniques for extracting self-insight computationally.

**Sensory Organs** capture data streams (sight, hearing, touch, smell, taste) that give us the ability to garner information about the external world. There are several mechanisms to understand insight computationally, capturing a human’s experience with its outside world. These sensors mimic and measure human sensory experience. Here we review sensors developed to mimic the body’s unimpaired natural functionality.

Concerning touch, tactile gloves serve as pressure sensors applied by the hand. High-resolution pressure sensing and mapping enable understanding of human movements, comfort, and ergonomics through visual feedback of hand interactions [216]. Electronic instruments, including the electronic tongue, measure and compare tastes. These instruments can

detect an array of dissolved organic and inorganic compounds like the human taste receptors via reactions for each sensor. They generate electric signals as potentiometric variations, to relate the signal strength to the concentration of these molecules [165]. Microphones can easily capture the audio experiences of an individual and are ubiquitous in today's world through mobile phones, computers, wearables, and IoT devices. Cameras and videos provide visual representations of what a human sees. Google Glass and Snap Glasses are both wearable cameras that are worn as a pair of glasses but can record or observe all the information that an individual sees directly[183]. Lifelogging cameras have thus been in research for several years now [300]. An example of cameras understanding human perception includes the three-dimensional gesture recognition vision systems in which video of the user's gestures are understood for certain meaning and actions [163]. These sensory signals give us detailed information about our experiences. Understanding experiences is one set of clues towards HSE. For example, if an event is relaxing, it may impact the health state positively versus a stressful event impacting the state negatively.

*Perception and Internal Cognizance* Experiences combined with internal signals that interpret health is another avenue by which to collect data about the individual. The self-perception theory indicates that people will know their emotions, attitudes, and other internal states based on their own understanding of their behavior and the circumstances at hand. "If these signals are weak or non-interpretable, the individual becomes an observer and must rely on cues to understand their inner state" [23]. This theory implies that internal sensory signals are learned and understood from making interpretations about their behavioral and physical circumstances.

Furthermore, our internal receptors and neurons enable us to comprehend our sense of position and movement. Proprioception is the sense of position and movement of the limbs, occurring by activity in sensory neurons located in the skin, muscles, and joint tissues. Proprioceptors (sensory receptors) play a pivotal role [100]. As these receptors are sensi-

tive to both tension and pressure, they relay information concerning muscle dynamics to the conscious and subconscious parts of the central nervous system. They provide the brain with proprioceptive information at the subconscious level, about the body's positioning with respect to gravity[223].

Perception data may give us one's view regarding their illness, and thus play a major role in the outcome of disease. The treatment of depression in CVD patients may benefit from a psychological intervention focused on patients' illness representations, self-efficacy beliefs, and their perceived social support [99, 92]. People who are feeling susceptible to an illness, expect to benefit if they change their behavior, and perceive their social environment as encouraging the change, but if they lack a belief that they can indeed change, their efforts are not likely to succeed. "Substantial empirical evidence suggests that self-efficacy beliefs (and the related concept of optimism) are reliable predictors of behavior, and that they mediate the effects of intervention on behavior change, including a number of health-related behaviors" [112].

The importance of identifying internal signals is to note the link between how internal cognizance and sensory experiences directly impact physiology. Two commonly known examples are the stress-response and the gut-brain connection. The common example of cognitive stress impacting physiology is in Figure 2.8 within chapter 2. About 100 trillion bacteria in the gut make neuro-active compounds, including the neurotransmitter serotonin, that regulate our emotions. The neurotransmitter can transmit signals directly to the gastrointestinal system to stimulate or suppress digestion [286]. The "gut-wrenching feeling" that people have historically is a person's stomach, or intestinal distress may cause anxiety or stress. Psycho-social situations directly influence gastrointestinal physiology and symptoms [286].

The methods for measuring or capturing information about sensory signals and internal cognizance is a computing challenge, as the information is difficult to attain continuously and unobtrusively. Currently, methods for capturing information is done through self-tracking,

querying the user through surveys, yet there is no integration at a broader scale of other objective data streams. There is an increasing number of “Quantified-Self” individuals who diligently track many kinds of data about themselves, via data collection tools and self-experimentation methods [49]. There may be several obstacles to this self-tracking, including tracking too much information and not tracking triggers with sufficient context. The current state of the field is to use Ecological-Momentary-Assessment (EMA) to have context-aware triggers that prompt the surveys [246, 72].

Capturing the natural senses with the tools mentioned above helps to feed data towards the estimation, but there is an obvious need to go beyond this. The next section describes data capture methods that go beyond our human senses.

### 3.1.2 Digital Sensors

Throughout history, quantifying health at a higher resolution has always changed what it means to be considered healthy. The stethoscope, blood pressure cuff, and microscope all modified the definition because of their ability to sense a change in the health state beyond human senses. Modern continuous sensors, digital interactions, and biological measurement devices produce vast amounts of data. **Continuous digital sensors** including computers, smartphones, and wearables, are relatively inexpensive in cost, energy, and computing power. The significant progress in this technology makes it widely available for monitoring individual health data streams continuously [167]. While continuous sensing gives us information readily, non-continuous or **episodic sensors** are still crucial for attaining high-resolution information with known levels of accuracy. Non-continuous methods that medical experts use with professional equipment give a deeper understanding and verification of ground truth. These methods include imaging techniques such as functional magnetic resonance imaging (fMRI), electrical signal capture such as electroencephalogram (EEG) and electrokardiogram

(EKG), and other clinical methods that provide one time-slice high-resolution information. The noncontinuous methods give a good baseline of an individual's state. There is a need for merging information from both continuous and episodic sensors for the best quality estimation over time.

Transitioning further into the digital sensor age, we will explore both continuous and episodic digital sensors. The four key sensor themes for continuous sensors include: perceptual, biologic, environmental, and event sensors. **Perceptual quantification** records human sensory information and provides continuous streams of data about emotions, sensory inputs, and brain activity. **Biological sensors** capture physiologic and molecular changes in the biology of the individual. Modern biological sensing mechanisms include omic data arrays, wearables, IoT devices, bio-markers, advanced imaging, and electrical signals. **Environmental sensors** track the external environment to understand a persons surroundings and social situations. **Event sensors** track events or outcomes continuously that can be done virtually or physically. Virtual events include how we use our computers, phones, wearables, and other computing devices into a single collection called a screenome [228]. Physical events include the real-world daily life events of individuals, such as those as described by Nobel laureate Daniel Kahneman These are captured in personal life-logging applications such as personicle [141, 205]. There are various modalities for acquiring information about these themes. These include sensing mechanisms such as imaging, electrical sensing, optical sensing, HCI, audio, radio-frequency, and strain gauges.

**Perceptual quantification** of human sensory information can be observed by methods of observing sensory inputs and brain activity. It is a computational challenge to find methods for quantifying human perception via sensors; however, there are growing mechanisms to compute behavior and perception.

Galvanic skin resistance (GSR) refers to the electrical resistance between electrodes when a weak current passes between them. The electrodes are placed apart and the resistance

will vary based on the emotional state of the subject due to variations in sweat gland fluid production. This sensor can be useful for the skin, as the skin serves as a conductor of electricity in situations when internal or external stimuli are physiologically arousing. “Arousal is widely considered to be one of the two main dimensions of an emotional response. Measuring arousal is not the same as measuring emotion, but is an important component of it. Arousal has been found to be a strong predictor of attention and memory” [180]. Stress sensors have been built based on GSR to measure test stress responses and detect stress states of an individual [274].

The advent of the smartphone and wearable devices has increased the possibilities of embedding sensors such as the accelerometer, GPS, optical sensors, and microphone. The integration of the sensors can collectively analyze patterns to understand an individual’s mental health. We can use passive data collected via devices (such as call history, text messages) or by active interpretation methods, such as analyzing voices in the microphone to measure stress. We can capture data on physical activity and movement during sleep from an accelerometer, optical heart rate, and GPS for a location to determine context. The combination of all of these gives a picture of an individual’s mental state [167].

The detection of emotions is becoming an increasingly growing area of research and development via emotion recognition methods. Corive et al. describes two channels of human interaction, one transmitting explicit messages and other transmitting implicit messages about speakers. The explicit channel is understood by linguistics and technology, but implicit messaging, such as understanding emotions, is not well understood [58]. Techniques by which to process imaging and audio helps establish pattern recognition to uncover human emotions. The techniques may include signals from bio-sensors, such as Haag’s method for training computers to recognize the emotions of the wearers[108]. “Biosensors may be used based on the synchronization between audiovisual stimuli and the corresponding physiolog-

ical responses” [127]. Emotion recognition techniques may also include analyzing speech signals (such as pitch and energy) [148].

In a broader sense of computing intelligence, The Turing test grades the ability of a computer to exhibit intelligent behavior indistinguishable from a real person [267, 245]. Human health services such as a human doctor have a heightened sensory ability to understand the above perceptual aspects about a person, and hence are still sought after to deliver healthcare to people. If a computer can understand these same factors, we may be able to scale health services to more people with better performance. Hence, the above perception and emotion-sensing research and related challenges are vital if computing intelligence aims towards improving human health. Human doctors do not solely use this mode of sensing. They also incorporate objective information about a person’s biology.

**Biological** sensors measure internal changes in biological and physiologic function over time. There currently exists a vast selection of sensors that exist for capturing biological information beyond the scope of this thesis. We focus on devices such as wearables and smartphones that act as biological sensors in a continuous manner [167]. We review a few modalities for biological sensing below:

a) Optical sensors within a wearable device continuously and unobtrusively monitor heart rates over a wearer’s wrist or head [119]. Optical sensors also detect oxygen saturation in the blood, and current research is conducted on blood pressure and glucose monitoring. The optical sensors work by using a reflected light pattern scatter from skin surface contact to estimate then the parameter of choice similar to a radar mechanism.

b) Imaging modalities are useful when screening blood for hemoglobin through cameras to understand a patient’s response to iron supplement treatment [282]. With a light source shining through a person’s finger, they have performed a chromatic analysis to analyze the color of their blood, estimating the hemoglobin level. Here a camera is directly collecting



light, rather than looking for a return signal of scatter. Traditional images of the skin can also be used to detect changes in texture for wrinkles, color for changes in melanin and jaundice [64].

c) Audio modalities such as the microphone can serve as a spirometer to measure lung function. Home spirometry is more widely accepted for its ability to continuously detect pulmonary exacerbation and improve outcomes of chronic lung ailments. Larson et al. presented SpiroSmart, a low-cost mobile phone application that performs spirometry sensing using the built-in microphone [151]. Audio methods also capture lung, heart, and bowel sounds through digital stethoscopes [154].

d) Strain gauge sensors measure the amount of force on the system. These are prevalent in many applications such as power meters, accelerometers, and the old-fashioned weight scale. Strain gauges work on the principle of Hooke's Law as given in Equation 3.1.

$$F = kx \tag{3.1}$$

Where  $F$  is the force applied,  $k$  is the spring constant, and  $x$  is the distance the spring changes. The most classic strain gauge sensor, yet still valuable, is the simple weight scale to measure body weight. These have been combined with electrical impedance sensors to get both body weight and body adipose tissue content [24]. Flexible skin-attachable strain-gauge sensors are an essential component in the development of artificial systems that can mimic the complex characteristics of the human skin. Pang et al. describe a strain sensor that enables the detection of pressure, shear, and torsion. Strain-gauge sensors are used to monitor signals from human heartbeats [213]. We can also measure respiration rate and

volume measurements with wearable strain sensors and are useful for monitoring patients with chronic respiratory diseases [51].

The strain gauge sensor is also used in bicycle ergometers and crank arms to measure the torque applied onto the pedals [60, 27]. Strain gauges can measure general movement activity (such as through wearable accelerometers) and aerobic exercise work capacity as CRF. CRF has a much stronger established relationship with true cardiovascular health [296]. Widespread use of standard wearable accelerometers that measure steps or higher semantic activities like walking, jogging, biking are indicative of activity only, hence the need for using wearables that can give estimates of CRF for better HSE. Wearable devices have estimated energy expenditure through various computational approaches such as deep learning [315], knowledge-based regression [162], and data filtering and segmentation techniques [6]. Energy expenditure provides insight into the total amount of activity performed by an individual but does not provide maximal work output to estimate CRF.

Computational research in CRF prediction began in the 1970s with the formulation of exercise stress scoring metrics based on the newly available chest strap based heart rate monitors [19, 29]. Recently, contextual understanding improved the performance of heart rate based CRF estimation and was further refined by calibrating custom algorithmic parameters for a particular user [8, 7]. Heart rate data incorporates additional features, such as vagal tone (commonly referred to as heart rate variability) and respiratory rate, to provide regression analysis more features for prediction [250]. Improvements in accelerometer-based CRF prediction have been achieved through body placement optimization [215]. The only known research that has attempted the use of multi-modal data for CRF prediction is by Firstbeat Corporation. They use both heart rate and speed information with a proprietary algorithm to filter periods of heart rate that are indicative of steady-state metabolism [56].

e) Omics, including genomic, microbiomic, metabolomic, proteomic, transcriptomic, and epigenomic data, is more widely used in understanding the biological model and state. This

data can be used given a particular scope of the individual that is being reviewed. A genome consists of all genetic material, microbiome (all microbiota or bacteria that live in and on the body), metabolome (all metabolites), transcriptome (all RNA transcripts for gene expression), and epigenome (all epigenetic material or changes to DNA not based on genetics). There are two broad types of omic data: static data and dynamic data. Genetic information is more static due to the fixed measurements that give information on risk and a person's unique model. For example, given the same sun exposure, two different individuals may have different responses to the sun due to their genetics. One may be sunburned, and another may not be. The genetics information gives insight into the relationship between sun exposure and the skin type, and thus the risks associated with genetics. Genetics is, therefore, a system model rather than a determinant for the state. Companies like 23andMe and Helix are focused on providing precise reports for genetics directly for the user via DNA genotyping methods done by mass sequencing technologies for individuals to understand their health model [1, 113]. Dynamic omics data, including data from the metabolome, transcriptome, microbiome, and epigenome, is always in flux with a high rate of change. These give us a systems state with continuous information over time. For example, glucose is a metabolite that is constantly changing. With the transcriptome, each exercise event an individual may partake in rapidly changes one's total RNA and gene expression. The research challenge with dynamic omic data is two-fold. Time-series data is needed to understand how the parameters are changing, and the data still is not available at a low-cost to both user intrusiveness and burden. Omic data research is an active area focused on further understanding of one's precise state [304, 133, 255]. Overall, the use of various omic techniques is referred to as a multi-omics approach, as shown in Figure 3.2.

In general, low-cost continuous sensors, computing methods, and lifestyle integration of cardiovascular health hence provide a unique opportunity towards advances in HSE. A robust framework to integrate vast amounts of temporal data and knowledge from biological sensors in cardiovascular health has yet to be published today.

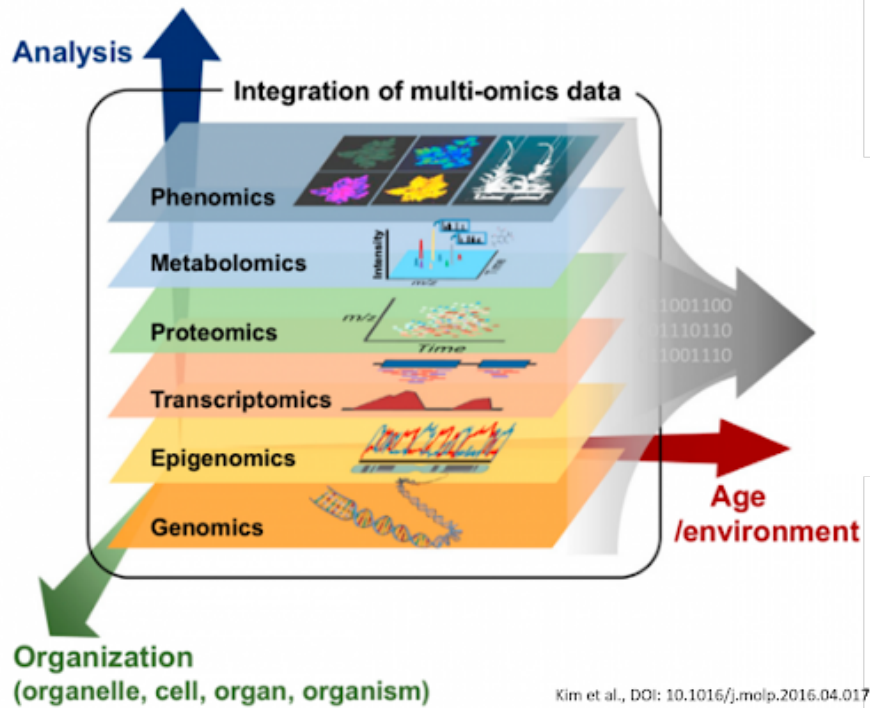


Figure 3.2: **Multi-Omics**. “Oma” means mass in Latin. Each layer of omic data is a massive array of data parameters regarding the specific layer. The advent of omic data mechanisms allows for high-resolution understanding of the biological state at various abstraction levels. Figure adapted from [144].

**Environmental** Environmental sensors allow for understanding the surroundings of an individual and their exposures. Many environmental sensors are measured directly (such as pressure and temperature sensors) or by sensor networks. Wireless sensor networks sense noise pollution in urban areas, and are available continuously. Traditionally noise pollution measurement has been involved and expensive. Santini et al. present a tool known as tiny-LAB, a Matlab-based tool, enabling real-time acquisition, processing, and visualization of data collected in wireless sensor networks [240].

While public data regarding the quality of the environment is readily available, it is not incorporated and tracked at the individual level. An open-source software platform, called EventShop, ingests and assimilates different data streams [86]. It combines different environmental data streams ranging from climate data, air quality, pollen counts, and micro-blogs

(like Instagram and Twitter) to understand how the environment is evolving. Also, the GPS allows us to track an individual’s location data and via geospatial data to understand exposures over time [65].

The concept of generalizing environmental impacts on a person is known as the “exposome,” or one’s exposure history. The United States Centers for Disease Control and Prevention (CDC) defines exposome as the “measure of all the exposures of an individual in a lifetime and how those exposures relate to health” [41]. This includes a longitudinal exposure throughout one’s life since conception. The environmental impact, in conjunction with one’s biology, creates the personal exposome, as shown in Figure 3.3.

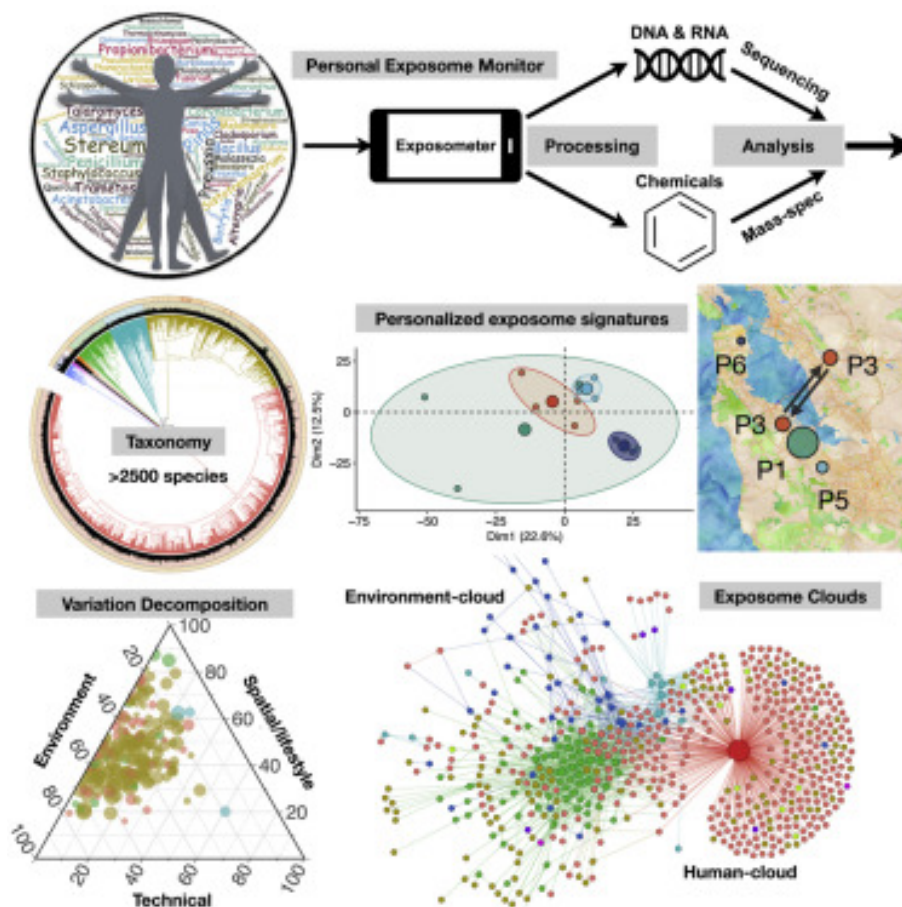


Figure 3.3: **Personal Exposome.** The exposome consists of both biological agents we are exposed to, and physical components such as light, sound, and chemicals. These exposures are related to a certain location and time. Figure adapted from [137].

This concept of personal exposome is noted in several studies [294, 41, 279, 249, 166]. Wild describes the exposome in three domains: 1) general external environment including the urban environment, education, climate factors, social capital, stress, 2) a specific external environment with specific contaminants, radiation, infections, lifestyle factors (e.g. tobacco, alcohol), diet, physical activity, etc., and 3) an internal environment to include internal biological factors such as metabolic factors, hormones, gut microflora, inflammation, oxidative stress [294]. He describes that the exposome requires consideration for both the nature of exposures and their changes over time. Jiang et al. developed a sensitive method to monitor personal airborne biological and chemical exposures. They found that human exposomes are diverse, dynamic, spatiotemporally-driven interaction networks with the potential to impact human health [137]. The exposome allows us to gather information specifically about an individual and be tracking their exposure history and how this may impact their health from a personalized level. Listed below are various tools used for measuring the exposome [294].

Measuring changing environmental pollutants can help model human exposure to micro-environments. Having the ability to measure the local environment of each individual will give insight into how they are affected by unnoticeable factors. For example, air and water pollution increase the risk of diabetes [226]. Long term exposure to particulate matter in the air can activate pathophysiological responses that can induce insulin resistance [314, 159]. A methodology developed and implemented by Dias provides the time-sequence of the exposure events, making possible association of the exposure with the individual activities and delivers main statistics on an individuals air pollution exposure with high spatiotemporal resolution [65]. The data combined with air pollutant concentrations for traffic pollution, estimate the micro-environments given a probabilistic model.

**Events** Digital methods are increasingly capturing events of human life. An event is a signal capturing the occurrence of an action. These daily life actions are what make up the history of each human life and also provide an input that changes the health state.

Approach	Tools
Biomarkers (omics)	
General	Genomics, transcriptomics, proteomics, metabolomics, epigenomics
Targeted	Adductomics, lipidomics, immunomics
Sensor technologies	Environmental pollutants, physical activity, stress, circadian rhythms, location
Imaging	Diet, environment, social interactions
Portable computerized devices	Behavior and experiences (EMA), stress, diet, physical activity
Improved conventional measurements	Job-exposure matrices; dietary recall

Table 3.1: **Exposome tools for measurement.** Figure adapted from [294].

With the rise of smartphone applications capturing events, individuals can easily track daily human activities, nutrition, sleep, and location to give a few examples. Physical activities can be tracked by wearable devices that can measure activity type, location, distance, and other relevant attributes [253]. Photos can easily capture food consumption habits. Image recognition of these photos can help to understand nutrition behaviors. There is growing accuracy in recognition of food photos and large scale images [309] [143]. The “personicle” concept is the most holistic application to collect lifelog events [136]. Life events can be at various scales of time and space. Molecular interactions occur at the atomic size and nanosecond scale, while an event such as “attending college” can last years and span a large area of space. Building a unified data structure to capture any event is essential for a robust data collection system. There have been efforts in pulling a unified view on how to represent these events, which include six key aspects (Structural, Temporal, Causal, Spatial, Experiential, Informational) [289].

Unique to the digital era, time spent with screens is a large part of awake events in the current world. There are growing mechanisms for capturing event data about an individual in this light. The “screenome” reflects the use of digital media and screen time to analyze what people do and spend time with on their screens [228]. Each digital interaction gives insight into a neurological event. Smartphone applications such as Apple ScreenTime and RescueTime tracks habits and time spent on particular applications or time on web [10]. Human interactions with social media capture much of an individual’s time spent (i.e., interests, activities, mental well-being). The interactions give us an understanding of individual

choices, behavior, culture, and habits. To further understand one’s relationship with the screen, there is a proposition for a Human Screenome Project in which “screenomics data should be sifted using a gamut of approaches from deep-dive qualitative analyses to algorithms that mine and classify patterns and structures” [228]. The project proposes reviewing an individual’s screen time to investigate physiological and psychological health states and sociological dynamics of interpersonal relations over time. Further research continues on the “screenome” for understanding screen time and human behaviors [310].

**Episodic sensors** Healthcare practice relies primarily on episodic or non-continuous sensors for dependable data acquisition. Professional-grade equipment generally has a higher expense but also higher accuracy and depth, which provides greater certainty in medical decision making. Professional equipment generally needs an operator for the measurements and may be more challenging to obtain this information unless it is a medical necessity. Although useful at the time taken, data collected in this manner is generally not used for any future steps in the person’s life after the clinical encounter. Professional-grade equipment provides calibrations at a current snapshot of the health state rather than continuous monitoring. Computing methods that are suitable for this include Kalman filtering [106]. Observing the role of accuracy and precision for each sensor is important based on the system goals. While many professional-grade types of equipment have high accuracy, it may be more relevant to hold precision (with continuous measurements), to detect a change in the health state. Episodic measurements combined with continuous measurements are therefore most ideal, increasing the accuracy of the current state with high temporal resolution to changes in the state.

Conventional professional equipment used for collecting individual data includes imaging, microscopy, and blood work. For example, magnetic resonance imaging (MRI) techniques are sensitive to changes in cerebral blood flow and blood oxygenation [149], and cranial magnetic resonance imaging (cMRI) have been used in cardiovascular studies to determine



that sulcal width, ventricular size, and white matter signal intensity change with age, sex, and race [312]. Other equipment such as dual-energy X-ray absorptiometry (DEXA) scanners, allow for one-off measurements of bone density and visceral fats [48].

Several medical procedures to gather data have risks associated that may be adverse to an individual. For example, an angiogram is an invasive medical procedure that holds the risk of radiation exposure from X-rays used, major complications including injury to the catheterized artery, arrhythmia's, allergic reaction to dyes, kidney damage, and more [54]. Thus, in addition to being episodic, there are also non-trivial risks to these high-quality measurements. These expensive sensing mechanisms initialize, calibrate, or validate a health state, and the less expensive sensors, which are more continuous, non-invasive, and accessible, would be available to estimate the current state from the most recently calibrated state.

### 3.1.3 External Data Sources

There may be sources that are external to the individual that may be capturing the individual's health information. This includes domain experts who are capturing measurements or observations on the individual, or external agents that an individual interacts with, such as a business.

**Domain Experts** Domain experts who are trained professionals can have a role in collecting data manually. For example, a physical therapist may use direct measurements and observations for determining an individual's flexibility or with goniometers. Clinicians commonly use norm-referenced, standardized measurements [122]. In the field of psychology, experts question individuals to understand their psychological state and needs, and thus mental health relies on domain experts for detecting and determining psychiatric issues. This system largely relies on domain experts to act as the sensor [266].

**External Platforms** Data acquisition about individuals is also common through platforms operated by external entities and businesses. The platforms designed to collect individual data are for pattern analysis to observe the individual's behavior. Loyalty programs for restaurants, credit cards, and stores collect information on individuals by observing their purchasing behavior. The programs have sophisticated tracking mechanisms between a customer and a brand that can help the program better understand the buyer's habits. Current progress in this field aims to use this information during a live interaction with a customer [61]. Companies may be using searches, clicks, and views on particular topics in order to understand an individual's interest and target advertisements accordingly. This process may not be at the request of the individual but based on the websites an individual visits and items they buy. External data gathering may also allow agents to track habits, behaviors, and patterns on individuals, giving the insight to develop the business [173]. The world's largest consumer-facing technology companies, such as Google, Facebook, Amazon, and Netflix, critically depend on this data acquisition process for their business performance.

The data digitally collected and stored externally is vast and expanding rapidly. As a result, organizations convert this information and knowledge for their objectives, whether it be in "astronomy (e.g., the Sloan Digital Sky Survey of telescopic information), retail sales (e.g., Walmart's expansive number of transactions), search engines (e.g., Google's customization of individual searches based on previous web data), and politics (e.g., a campaign's focus of political advertisements on people most likely to support their candidate based on web searches)" [184]. According to Murdoch, it is inevitable that this data will have a large role in providing insight into the health state, such as data gathering mechanisms to observe broader trends to predict emergency room visits. Google Trends (a free, publicly available aggregator of relevant search terms) has been used to predict surges in flu-related emergency room visits [95]. Similarly, Twitter updates were as accurate as official reports at tracking the spread of cholera in Haiti after the January 2010 earthquake [173].

Data collection through sensing mechanisms remains a considerable research challenge for humans due to the complex nature of understanding and capturing the health state. The above review is just the tip of the iceberg. The next step in the health state computing pipeline is to use these sensor measurements within an estimation approach.

### 3.2 State Estimation for Individual Health

State estimation for health allows us to understand the current health situation for an individual. We will review the various approaches for how health can be estimated: quantified approaches, classification approaches, probabilistic approaches, and subjective approaches. It is important to note that though various approaches may exist, they are highly interconnected, and each one alone may not be sufficient.

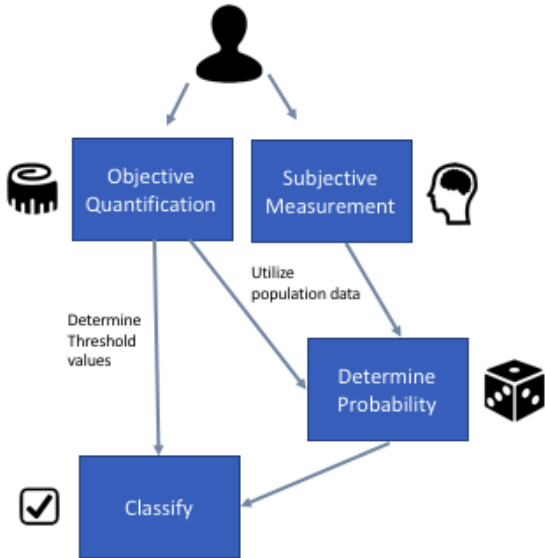


Figure 3.4: Estimation Approaches are Interconnected.

### 3.2.1 Quantified Approaches

Quantified approaches give insight into where an individual is within a state space based on objective measurements. Individuals take blood tests, such as a basic metabolic panel, which is a group of tests that measures different chemicals in the blood, including blood glucose, calcium, electrolytes, and lipoproteins. However, the extrapolation of the information then allows the physician to come to a conclusion or learn what these levels may indicate. A physician can form a conclusion about the readings depending on the parameter of the state space. For example, if HDL cholesterol is 20, that is considered in a poor state, compared to a value of 70, which is a good state. A visualization of these quantified states in the case of hormones is in Figure 3.5.

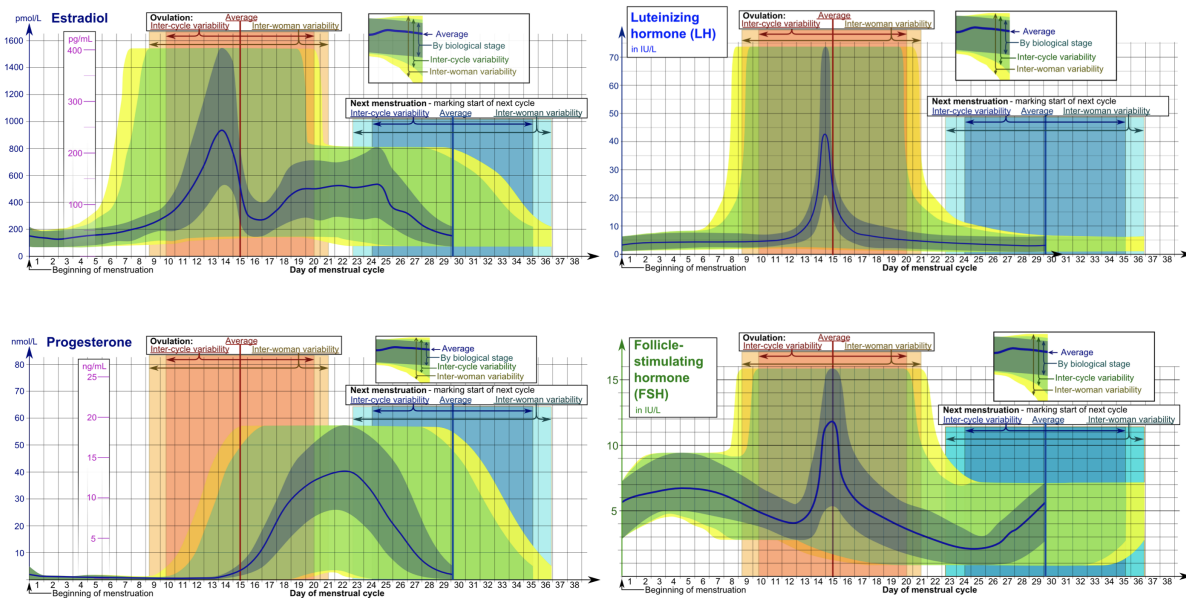


Figure 3.5: **State Space for Quantified Health.** Here are four charts showing how hormonal fluctuation occurs during menstruation. At each moment, an ovulating female has a point on this state space they exist. Regions specific the limits to which the state can exist and what the state means [268].

One's blood alcohol level is determined based on measurements that determine the specific amounts of alcohol in the blood. If the blood alcohol content (BAC) reaches a threshold of

0.08, it indicates legal intoxication, which is a threshold-based classification. Quantification is necessary, but may not be sufficient alone within an interpretation or added meaning of what this measurement indicates. Similarly, eye power is measured by a diopter, a unit of measurement of the optical power of a lens or curved mirror, which is equal to the reciprocal of the focal length measures in meters [66]. The objective measurement of power can be useful when determining thresholds, such as determining whether an individual is legally blind or not. This need for classification may be determined external to the individual (i.e., safety, legal) not necessarily based on the health state of the individual. Thus placing a layer of information on top of the quantification approach provides more meaningfulness to the health state.

### **3.2.2 Classification Approaches**

Classification methods, as we have described in Chapter 2, define health based on disease, conditions, and various measurements to define the state of one's health. Classifications may help in creating standards that may be beyond an individual's specific situation. In the example of the blood alcohol content, the standards of classifications can identify toxicity and thus concerns about safety. Disease classification is internationally recognized through the International Classification for Disease (ICD), Tenth Revision, by the World Health Organization. The ICD-10 code serves as a classifying method to label whether an individual is in a disease state. The ICD allows for systematic recording analysis, interpretation of mortality and morbidity in various locations globally, and disease classification for health problems [209].

Widely used classifications of disease include (1) topographic, by bodily region or system (i.e., vascular disease), (2) anatomic, by organ or tissue (i.e. heart disease), (3) physiological (i.e., metabolic disease), by function or effect, (4) pathological (i.e., inflammatory disease), by the

nature of the disease process, (5) etiologic (causal) (i.e., staphylococcal causing pneumonia), (6) juristic, by speed of advent of death, (7) epidemiological (i.e., typhoid), and (8) statistical (i.e., diet as a cause of atherosclerosis) [231]. For each of these classifications, we see that there may or may not be threshold values that contribute to classification.

*Threshold* Classification approaches are defined by how an individual's health can be labeled. It may be a particular threshold met to define whether an individual has a disease or does not. Thresholds are, therefore, a region of interest within a continuum. For example, in Figure 3.6 showing the Fletcher Curve for lung function decline, the classification of where one falls along the curve can be "onset of symptoms," "disability," or "death" [259]. The thresholds are determined generally by population data. It would not be possible to estimate the health of one individual solely and classify this individual's state unless it was a postmortem analysis. Therefore thresholds are more commonly generated given population data.

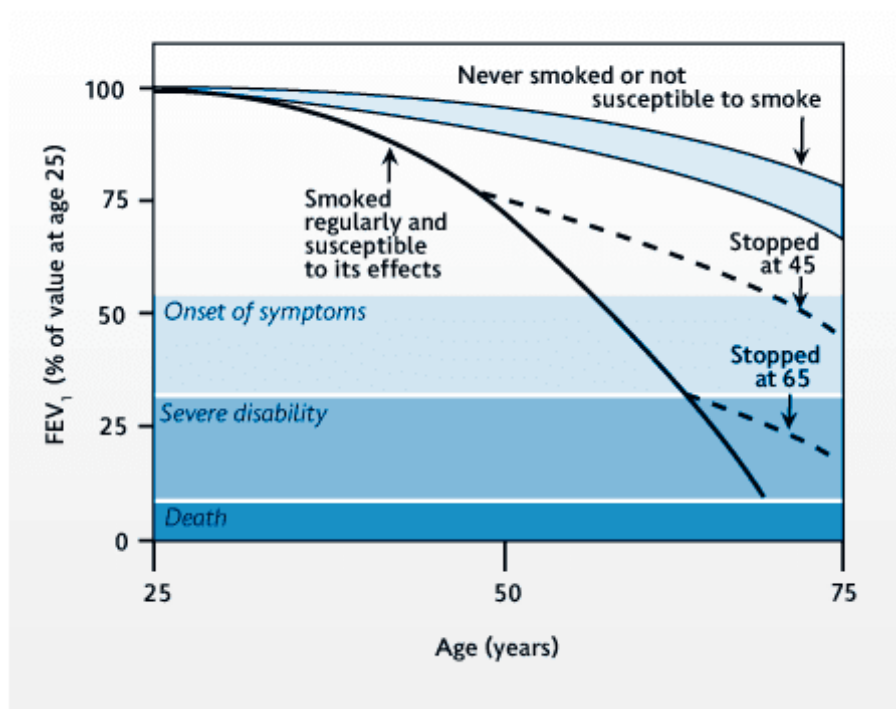


Figure 3.6: **Fletcher Curve**. Lung function is measured through a quantitative mean of FEV<sub>1</sub>, which is the Forced Expired Volume a person can exhale in 1 second. This parameter is then placed upon the map of which regions and thresholds are relevant for understanding health state. [259].

*Categorical* Classification approaches are also important in a categorical sense. For example, in many cases of cancer, a tumor can be defined as metastasized or benign without any particular threshold met. One either has cancer or does not. The same applies to infection; one can either test positive or negative for infection with no thresholds to determine the disease. We also see classifications of sex if an individual either has chromosomes XX or XY. There is no in-between threshold for these classifications.

Classification approaches alone are not sufficient for defining the state of an individual's health if used solely; there will be no observation of state change at a high resolution, and change will only be detected when crossing a threshold. The continuum is essential for recognizing alongside classifications.

### 3.2.3 Probabilistic Approaches

Probabilistic approaches measure the likelihood of a future event. Risk factors are needed to compute the probability of having a disease. There are methods of calculating risk that takes into consideration age, biometrics, and gender to determine chances for having a disease. The information in this approach may be based on the estimation of a population or based on an individual's unique information. Essentially this approach is giving a person a distance between an undesired region on the state space and their current state.

*Population Based* Population-based probabilistic approaches use population samples that are most similar to the individual. It is essentially using other people as a model for one's self. Population health is “the health outcomes of a group of individuals, including the distribution of such outcomes within the group,” including health outcomes, patterns of health determinants, and policies and interventions that are links to outcomes and determinants [146]. Health can be measured at a population level to understand the functioning and well-being of a group that is similar to an individual. Classic factors for calculating population health include geography and demographics information.

In regards to CVD, population-level health estimation is a primary way of understanding behaviors that contribute to cardiovascular disease for individuals. Multifactorial cardiovascular health risks have been investigated in many of the extensive epidemiological studies such as the Framingham Study, which lay the foundation for most modern clinical guidelines by the AHA and American College of Cardiology (ACC). Conclusions from these large studies are useful in current clinical practice through the Atherosclerotic Cardiovascular Disease (ASCVD) calculator [130]. This pooled cohort algorithm was based on linear regression analysis for four separate cohorts of individuals female blacks, male blacks, female whites, male whites. They also take into account age, systolic and diastolic blood pressure, cholesterol, smoking history, diabetes (binary field: yes or no), medication history (hypertension, statins,



aspirin only), and is only applicable for patients in the age range of 40-79. The limitations of this calculator include the requirement of invasive blood data and non-consumer based lab processing. The estimated risks and scores are specific to defined combinations of the risk factors and demonstrate how they vary over a broad spectrum of potential profiles. These determined risks are now primarily used as guidelines for understanding individual risk [290]. The American College of Cardiology takes the inputs of risk factors for ASCVD such as age, gender, blood work (systolic blood pressure, diastolic blood pressure, cholesterol), medical history, and medication status. The calculation will generate the chances by percentage for the likelihood of ASCVD. More information such as genetics and behavior can further enhance the accuracy of the risk calculation [290, 36, 160].

There are many human health risk assessments, models, and tools developed by the Environmental Protection Agency (EPA). For example, the “All Ages Lead Model (AALM) predicts lead concentration in body tissues and organs for a hypothetical individual, based on a simulated lifetime of lead exposure” and “statistical methods extrapolate to a population of similarly exposed individuals.” This model provides risk assessors with a tool for evaluating the sources of lead in a population or potential exposure to lead [76].

There have been several efforts for making long-term population health research using probabilistic approaches more expansive, include the United States Precision Medicine Initiative led by President Barack Obama. The Precision Medicine Initiative is a long-term research endeavor, involving the National Institutes of Health (NIH) and multiple other research centers, which aims to understand how a person’s genetics, environment, and lifestyle can help determine the best approach to prevent or treat disease [125]. Other examples of long-term population health research include Mobile Sensor Data to Knowledge [147], and Alphabet’s Verily division [273]. These approaches seek to understand what population cohort characteristics allow for similarity comparisons between individuals to group them into sub-populations. These sub-populations allow for a more specific estimation of an individual’s

health state. The World Health Organization has conducted long term studies for lung cancer as rates had rapidly increased in the 20th century; sub-populations and geography were closely examined in the epidemiological studies to understand the reasons for the spike in disease [26]. Long term research efforts for acquiring data may take decades before we have data available for meaningful insight, as they largely depend upon outcomes of morbidity and mortality before they become sufficiently robust. For this reason, they are inappropriate to use as the benchmark for state estimation at the individual level.

*Non-population based* Personalized models are used in probabilistic approaches to determine one's health state. Current approaches to identify patients at risk for arrhythmia are of low sensitivity and specificity. Personalized approaches to assess Sudden Cardiac Death (SCD) risk in post-infarction patients have been developed based on cardiac imaging and computational modeling. Personalized three-dimensional computer models constructed post-infarction hearts from patients' clinical magnetic resonance imaging. The propensity of each model to develop arrhythmia was assessed [11], as shown in Figure 1.4 in chapter 1 (personalized heart models).

Personal experience may also account for valuable information in probabilistic approaches. For example, if an individual experiences knee pain when cycling that leads to knee tendinitis, by experience, they can determine if the probability of the same pain occurring to be associated with the tendinitis if they have data collected. These are experience-based probabilities that can help individuals estimate their health state. The personal experience and perception of pain in determining the probability of a health state may also a subjective approach.

### 3.2.4 Subjective Approaches

Subjective approaches for estimating health include observation and perceptual data that can be used as information to understand one's state and also future risks. If any individual holds information such as self history, changes in feelings over time (i.e., pain increase over time), this information can be used in correlation to other objective metrics to generate a conclusion about the current state. This method for gathering knowledge and information depends on a user's intent or experience. We can observe this, for example, in the dynamic nature of recreation experience. Experience patterns capture engagement over time. Seven qualities of hikers' experiences (four mood measures, two satisfaction measures, and landscape scenic beauty) were assessed at twelve moments during a hike in Hull's study. The subjective data components may be relevant for understanding the experience and how this is contributing to the state of the individual [128].

There have been advancements in tools for data collection methods, such as facial recognition, natural language processing (NLP), and surveys that combine pieces to generate an estimation for how an individual is feeling. Facial recognition attempts to estimate emotions and track feelings over time. NLP relates thoughts, language, and patterns of behavior learned through experience to specific outcomes [169, 85]. Basic survey methods for asking individuals about their lifestyle and behavior are also still common forms to gain knowledge about an individual's health trajectory.

All of the above-mentioned estimation approaches have a unique value. The integration of the above approaches is still highly sought after in the research and clinical communities. In order to truly have the integration at an individual level for state estimation, we need to investigate modeling techniques.

### 3.3 Personal Human Health Models

Given the measurements described above and estimation approaches, we need to have a personal model to fulfill some essential design requirements mentioned. A model characterizes the relationship between the inputs and outputs to one's health. The most obvious benefit is the ability to predict, which requires having a model by which to estimate a future state. Thus, we review various approaches in forecasting or modeling long term health. To design a robust model, we must incorporate domain knowledge, taking epidemiological data, and individual data to ultimately make a generalized personal model. We must note that the model is not always static and needs to be updated.

There are several distinct levels we consider for building a personal model. In the first level, researchers have applied general rules directly related to known human domain knowledge (i.e., biology, physiology, behavior) to establish a general human system model. This model approach is most familiar in the form of textbooks. In the second, we incorporate specific knowledge that applies to sub-categories of similar people that we call cohort generalization or sub-population fitting. This layer is the function of the individual at a specific time and place. Clinicians use this method quite often when suggesting a course of action for a particular patient. Third, we consider “static” variables such as genetics and birthdate (fixed individual parameters). Fourth, dynamic individual models use the previously mentioned three layers with constant real-time data to update the model. The visualization of these layers is shown in Figure 3.7.

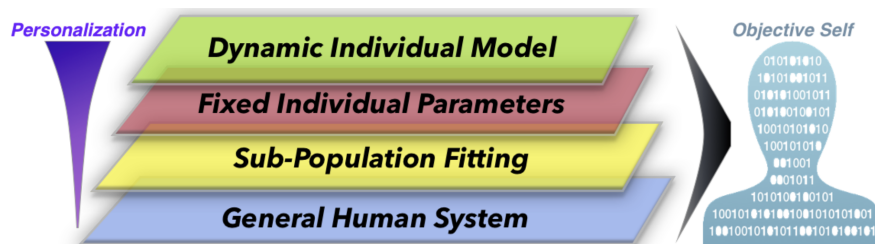


Figure 3.7: Layers of personal data modeling.

### 3.3.1 General Human Systems

We first review models using general rules provided by medical and biological knowledge to represent the human. These are objective models based on well-defined textbook knowledge that gives us the insight to forecast human health. Models help us understand a human's relationship to the environment, its current situation, and plans for the future. We will review models using biological and behavioral domain knowledge.

*Physiological modeling* takes biological knowledge and apply this to our understanding of how a person functions. Biological information regarding the individual is captured in, for example, the Athlete Biological Passport (ABP). This model is based on the monitoring of biomarkers to catch doping with performance-enhancing drugs. The ABP gives a unique mark on an individual's hematological profile as a baseline, and any deterrence in the system, such as through doping, will trigger a physiological change [251]. This model applies a standard set of thresholds to all athletes, such as having a hematocrit under 50 is considered normal.

Researchers have also built models regarding functional biological components, such as a model of the heart for better understanding the functional interactions between cells, organs, and systems, as well as how interactions between them change disease states. “The rapid growth in biological databases; models of cells, tissues, and organs; and the development of powerful computing hardware and algorithms have made it possible to explore functionality in a quantitative manner all the way from the level of genes to the physiological function of whole organs and regulatory systems” [203]. Figure 3.8 is a model determined by several human samples, giving us a representation of the constriction of blood flow within an example heart. This type of model allows for the simulation of various physiological situations.

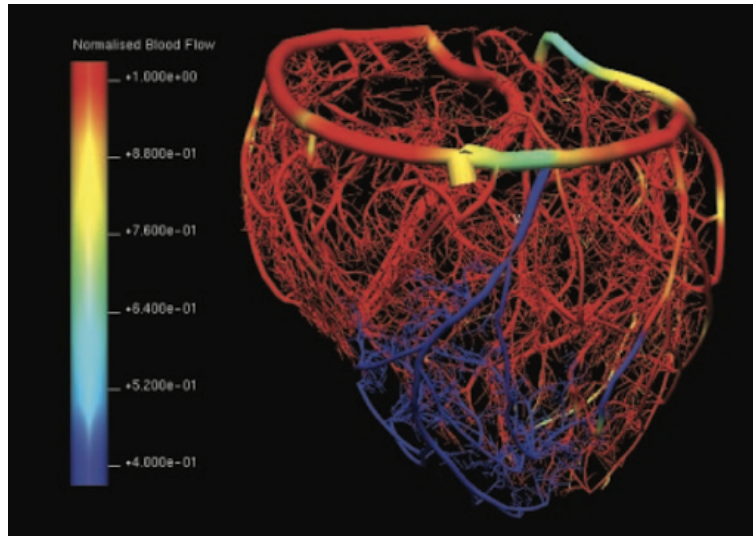


Figure 3.8: **Modeling the Heart.** This figure shows the left coronary circulation model subjected to constriction of one of the main branches leading to blocked blood flow [203].

Computational modeling can be used for biological processes that are occurring within the mind. For example, the neurological process of associative memory has a model of processing. This model demonstrates how the brain handles pieces of information in a serial order. A study conducted sequentially arranged patterns to simulate memory serving as a digital representation of human memory [287].

*Behavioral modeling* allows for the interpretation of human behavior and preferences. Human behavior can be very complex, and models are often simplified within a system or domain. A traditional behavioral model is Pavlov’s classical and operant conditioning, which pairs a neutral stimulus with an unconditioned stimulus to elicit a desired response [53, 97]. Another early theoretical model developed for understanding health behaviors was the health belief model in Figure 3.9, which shows how particular human attributes contribute to perceived severity, benefits, and motivations to take action on health.

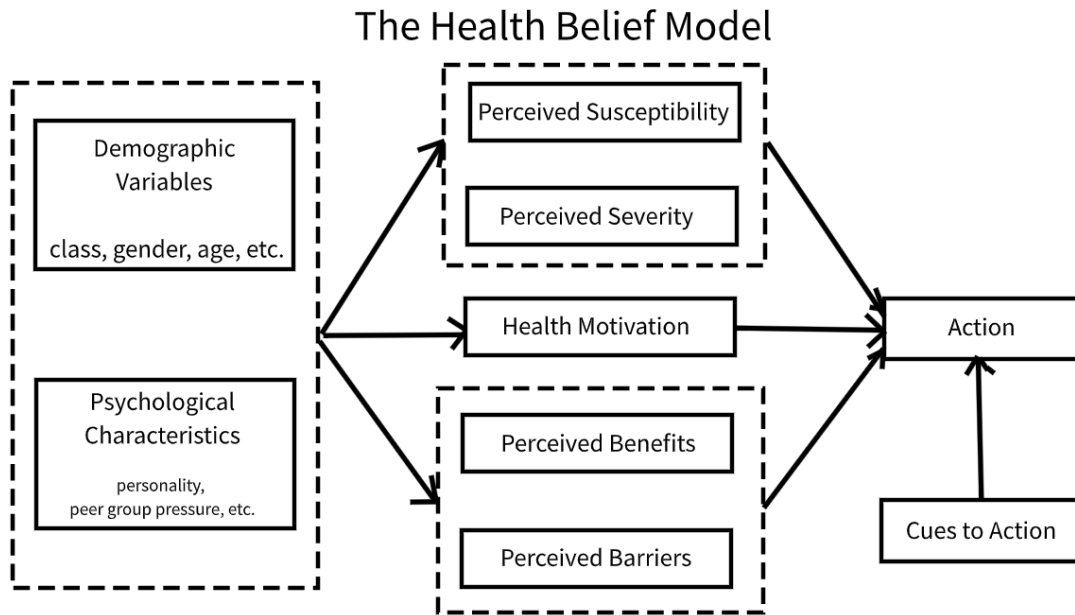


Figure 3.9: **Health Belief Model.** This model predicts the probability of a person taking action that would benefit their health. The model focuses on understanding what people believe in and where they see benefits in certain actions. [295].

A third example of a human behavior model shows that behavior has multiple layers that are interconnected. This research was conducted in the modeling the behavior during training laparoscopic surgeons [288], shown in Figure 3.10.

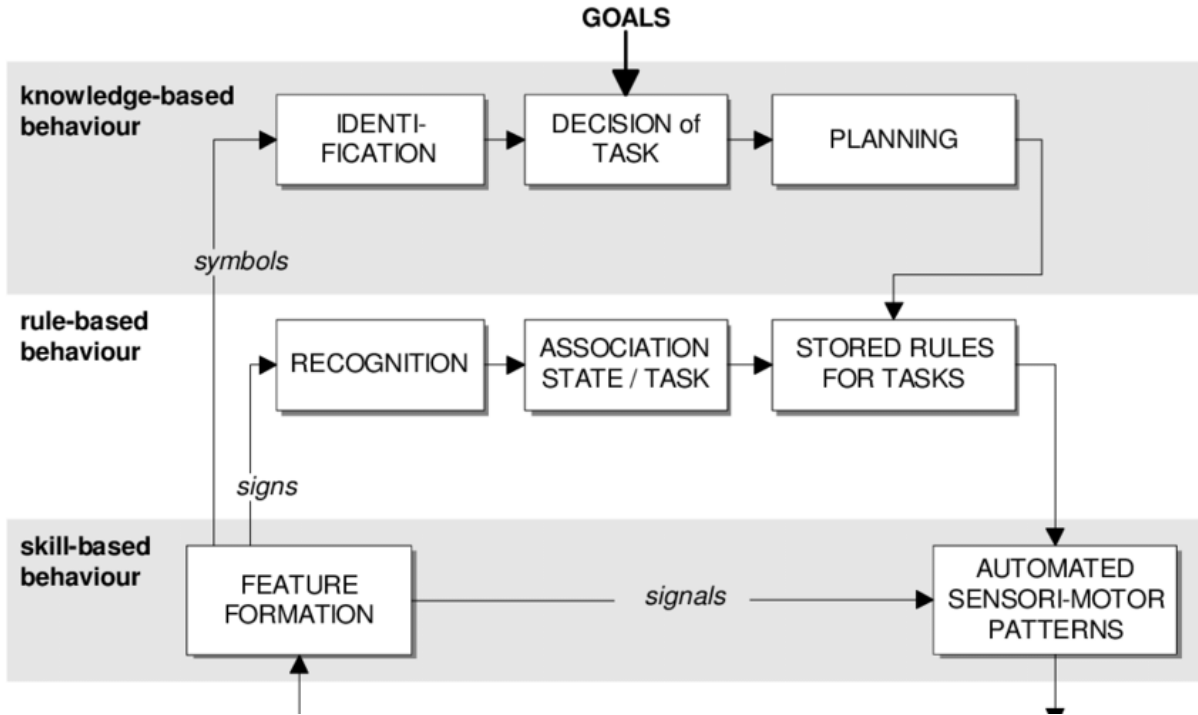


Figure 3.10: **Behavior Models.** Rasmussen’s behavior model outlines behavior in three different levels: Skill-based behaviour represents human behaviour that takes place without conscious control. Rule-based behaviour involves task execution controlled by rules. In unfamiliar situations in which there are no rules or procedures, behavior is knowledge-based [288].

The model of human driving gives a specific representation of behavior. “The model of the driver includes lateral and longitudinal control tasks such as path-following, obstacle avoidance, and headway control, examples of steering and braking activities performed by the human driver” [164]. Macdam et al. recognize that both physical limitations and other attributes that make the human driver unique, characterize human control behavior. Knowing this model predicts the performance of the human over time.

Quantification methods are needed for behavior models to be more comprehensive and precise. Human interactions have been modeled through Bayesian computer vision systems to recognize human behaviors in visual surveillance. Oliver et al. designed a system combining a top-down with bottom-up information in a closed feedback loop, comparing two different state-based learning architectures for modeling behaviors and interactions [207]. There



are also quantification methods applied to human performance modeling to analyze human function and development of systems for optimal user experience. Sebak et al. summarize this well in their review of behavior models describing how “human performance models predict human behavior in a task, domain, or system. However, these models must be based upon and compared against empirical human-in-the-loop data to ensure that the human performance predictions are correct” [242].

Many of these approaches use a connection between various functional abilities or biological components connected through an informal or formal graph structure. We can see this clearly in many of the figures visually. While using these objective measurements to model general health through biology and behavior may be useful, combining this with models based on demographic data gives us further precision into one’s long term health state.

### **3.3.2 Sub-Population Fitting**

Demographic, cohort, or population-based models refine general health models. Cohort generalizations consider population-level findings and apply this data to an individual, based on most similarities to the individual. For example, risk calculators predict cardiac arrest. The Revised Cardiac Risk Index is a cardiac risk stratification tool that allows us to model humans by making taking population data and making calculations based on known unique factors (gender, age, blood panel) a [105]. A defined cohort allows estimation of distributions of prevalence rates of relevant variables, such as risk factors. Advantages of a population-based cohort study arise from the specific health-related parameters for the reference population, which allows the estimation of population risk figures used for planning [257]. This is a typical medical approach to sub-group matching.

Sub-population data is useful for making assumptions about individuals based on others who are similar to them, known as collaborative filtering. This is a typical computer rec-

ommender system approach to sub-group matching. This method makes predictions and recommendations based on feedback information from a community of users. The method assumes that people are similar when they have similar interests and do not assume causal modeling. We see the collaborative filtering techniques commonly used for online platforms, such as Facebook, Google, Netflix, and Amazon, used in recommendations to an individual based on the preferences of others who are similar. The technique applies to healthcare in systems with limited prevalence. Personal healthcare modeling for communications between users in online forums uses this method to monitor the condition of patients and help people with similar conditions interact and exchange experiences [276].

### **3.3.3 Static Individual Parameters**

Individual parameters that are not constantly changing over time and are “fixed” can be used to understand an individual’s model at a given point in time. These parameters may include genetics data, birth information, and their unique medical history, or developmental disorders. “Fixed” is in quotations because all components are subject to change, including genetics, thorough DNA damage, and technologies like CRISPR [55]. Recognizing fixed parameters allows for modeling an individual upon these constraints. For example, the bone length is usually a fixed parameter when an individual reaches adulthood. As the growth plates close, the bone will remain a particular length going forward, providing information on one’s body size and enabling customization for health or shopping for clothing. We have seen these models developed in Retul’s bike fit testing for helping users select appropriate bike models, improving pedaling technique and selecting bike components for one’s specific biomechanics and measurements, as shown in Figure 3.11 [229]. Orthotic specialists use this approach as well. Many multimedia efforts focus on understanding which clothing best suits the fit of an individual’s parameters [13, 82].

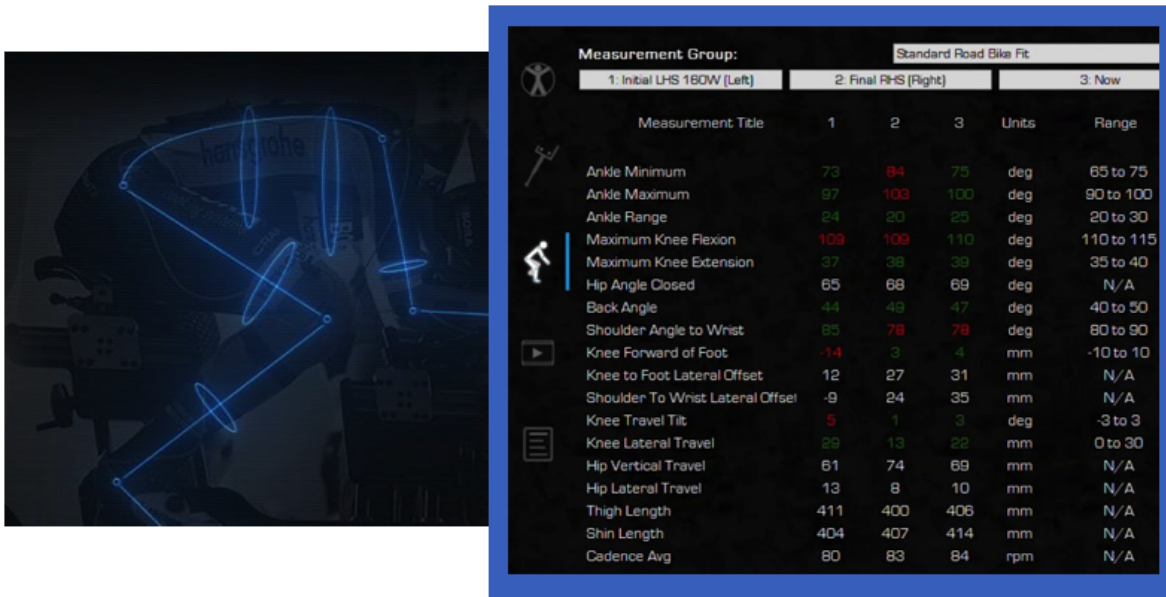


Figure 3.11: **Retul Bike Fit Personal Model.** Individual measurements are collected and applied to a highly customized bike fit.

Other static parameters, including genetics, can be useful to model an individual’s exposures and future health risk. Companies such as 23andMe are using genetics data to model an individual and give them information about their objective genetic parameters. With this, we can make predictions about a person’s future health state, such as genetic predispositions for developing heart disease, cancers, relationship to various foods and activities, and behavioral tendencies [1].

### 3.3.4 Towards Dynamic Personal Models

The modeling techniques mentioned above so far assume that the person is an entity that does not change over time. These models assume that the individual model is static. In reality, this is a false assumption. A human model must continuously change due to circumstance, aging, and events that are impacting the person. There have been further efforts in building adaptive modeling techniques.

Contemporary digital technologies measuring constant change, such as on-engine monitors in airplanes, are used to maintain the well-being of jet engines and also similarly could be applied to humans [260]. Tarrasenko’s “digital twin” consists of a particular engine and a unique computer model that accurately adapts to estimate the state of that particular engine. For example, if an engine flies from Los Angeles to Houston where the weather is dry, compared to an engine that flies a rainy route from Vancouver to Montreal, the model of how the engine state changes is different, even if the engine is from the same manufacturer. When applied to people, the digital twin is a computer simulation of an individual that dynamically reflects the individual’s molecular status, physiology, and lifestyle over time. An alternative concept of digital human twins relies on a large information infrastructure that characterizes each individual for demographics, biological data, physiology, anatomy, and environment, along with treatment and outcomes for medical conditions. Tarrasenko et al. describe “the twin represent(ing) the nearest-neighbor patient, derived from the algorithmic matching of the maximal proportion of data points using a sub-type of AI known as nearest-neighbor analysis.” This example harks back to the collaborative filtering methods mentioned above.

The Howard Chang Personal Dynamic Regulome Lab works on understanding what inputs change the state of an individual. Their concept is to “create the GPS system for navigating the regulomic landscape of human health and disease. Each gene has a system of switches that controls when and where a gene will turn on. The regulome is the complete set of switches for all genes. Regulated gene expression plays key roles in nearly every developmental program and disease state, and dozens of human therapeutics act through this regulation to alter gene expression. We aim for technologies to determine rapidly how the regulomes of individual patients and disease tissues maps into a global landscape, and how each intervention changes their trajectories in real-time. This real-time information feedback between perturbation and outcome is essential to tailor precise medical treatments for individual patients. They face challenges of how these new technologies must be compatible with the small sample sizes of human biopsies and clinical workflows if they are to be widely

adopted” [225]. Essentially, the regulome integrates events with specific changes in biological gene expression. This concept is visualized in Figure 3.12.

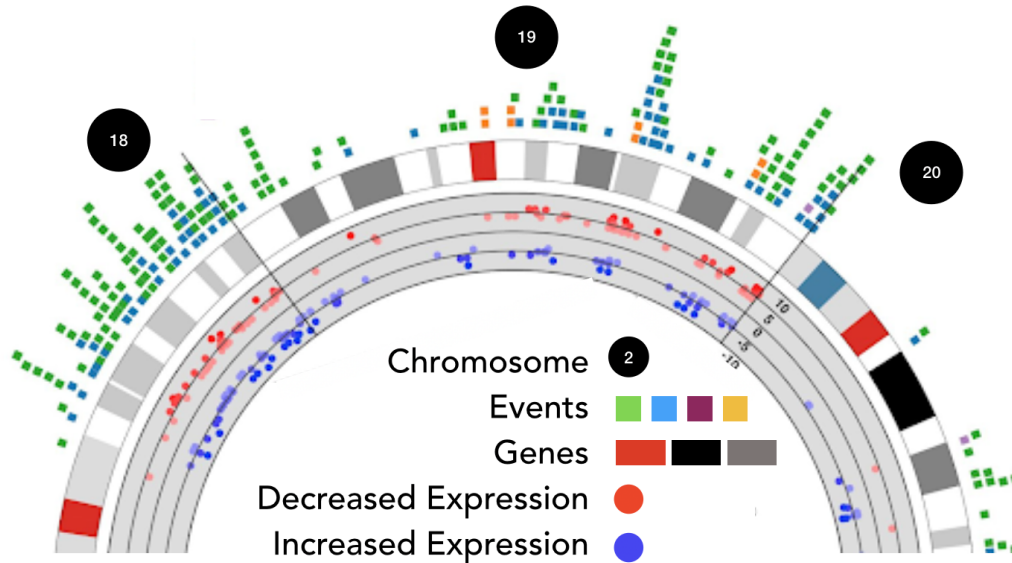


Figure 3.12: **Personal Regulome.** This consists of the complete set of switches for all genes within an individual. Here is a rendition of chromosomes 18-20 of an individual whereby certain events either turn on or turn off the gene expression of specific genes within the individual. This regulome has specificity down to the cell level and is a key pillar in developing dynamic individual models.

Dynamic modeling can be crudely done currently in systems like STELLA, a visual programming language created to run models through a graphical interface. This model is represented by stocks, flows, and variables to represent a single entity moving in the system. Each of the interconnected icons is defined mathematically. STELLA can be used to represent a biological model for human cholesterol dynamics, circadian rhythm control, and more [111]. STELLA can be used for modeling cardiovascular function [233]. Other examples of systems dynamics programs to see entities changing in a system is through Insight Maker, an open-source program to use loop diagrams to create simulation models [168]. Important to note, the “dynamic” word in these approaches refers to the system state is dynamic, while the relationships between variables are generally static.

In order to have a successful dynamic model, the constantly changing health measures are recorded using real-time sensor data. Firstbeat Technologies collects various data sources through wearables and makes sense of the information together and its relevance to an individual's health state. The dynamic model of a person's aerobic fitness level uses the calculation of the maximum volume of oxygen (VO<sub>2</sub>max). The method uses the heart rate and speed relationship and detects the reliable data period for VO<sub>2</sub>max estimate during the exercise. Firstbeat has created a method to calculate the VO<sub>2</sub>max in an uncontrolled real-life exercise. This includes a physiological basis, the linear relationship between oxygen consumption and running speed. The "technology (such as GPS sensors and foot pods in wrist devices or mobile phones) enables reliable measurement of running speed along with heart rate. These parameters can be monitored continuously and automatically during each workout" [84]. They capture the personal information of the individual, such as age. They use a device that calculates heart rate segmentation to estimate the fitness data through the person's heart rate and speed data [84]. These models are a good start, but still remain quite primitive in that they only use a handful of variables to build a dynamic model about the individual.

Once again, there is a common theme of needing to use continuous data streams to build models about people that improve over time connecting the data to the person's model has the most promise in the simple graphical approach. Modern artificial intelligence modeling techniques also have been utilizing graphs as a primary method of connecting data and meaning [235]. For this reason, we move into reviewing the various techniques for representing the individual dynamically through graph modeling below.

## 3.4 Artificial Intelligence and Graphs

Humans have spent a greater part of the last century developing machines that can improve the quality of life. Computing, with machine processing information towards a goal, contributed to a large portion of this progress. Machines that can execute complex thinking tasks such as reasoning, planning, and learning is artificial intelligence (AI) [235]. Modern research efforts continue to investigate how these artificial intelligence principles can apply to health. Health is a product of functional interactions of many objects such as molecules, cells, organisms, and the environment. Based on this premise of connectedness in the world, graphs are one of the most potent ways to represent the real world computationally. We review key relevant principles in AI for HSE.

### 3.4.1 Graph Networks

Established by Leonhard Euler in 1736, graph theory is the study of modeling relationships between objects [102]. Undirected graphs are a tool used to capture symmetrical relationships between objects. Directed graphs contain connections that have a source object and a target object, thus establishing a direction by which relationships flow. Graphs are an excellent computing methodology for capturing the connectedness of information about our health. Graph networks connect how genes are expressed in cancer and healthy tissue to identify topology that distinguishes them apart [239]. Gene identification methods can also determine pathways for CVD, merging omics data with biological knowledge, as shown in Figure 3.14. By identifying the biological networks for a system, we can see transitions from a healthy to a diseased state [67]. Graph networks can construct the human brain networks (brain regions as nodes and pairwise associations as edges) by combining functional and material causes (shown in Figure ) [21]. There may be a clustering of physiologic patterns related to the nodes, as

seen by measuring REM activity as a risk for CVD (shown in Figure ) [124]. Lastly, by using graph structures, we can start observing causal relationships.

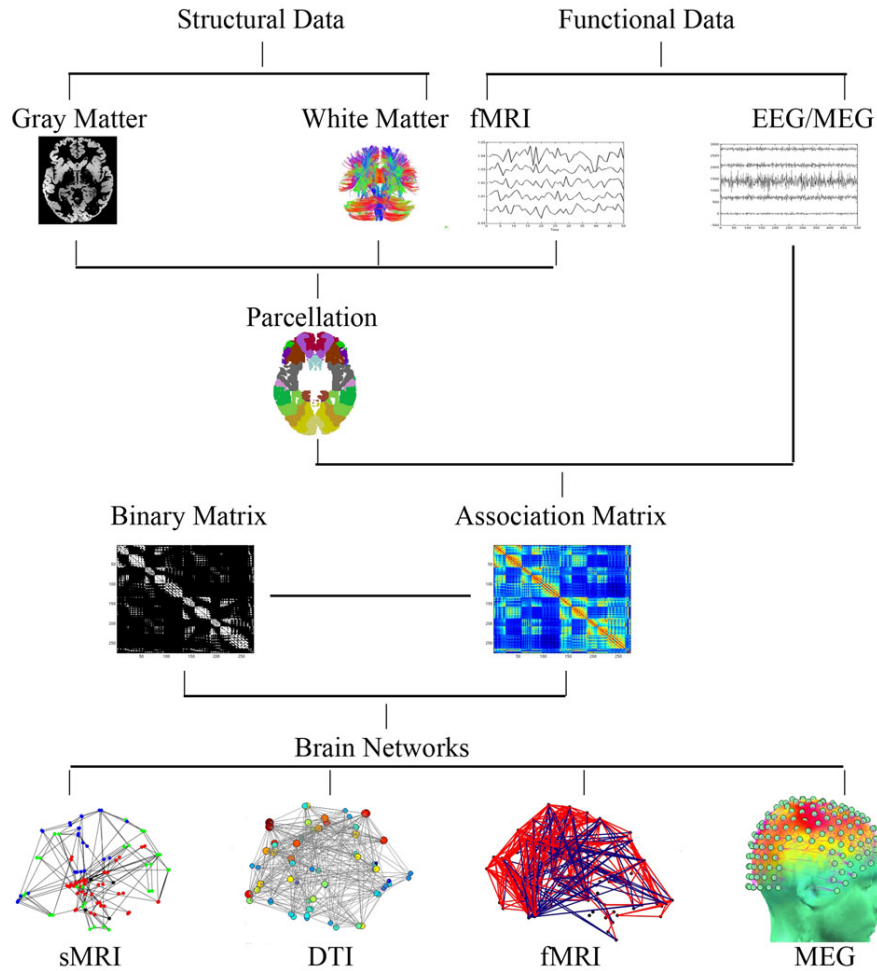


Figure 3.13: **Workflow of Human Brain Network.** Connections can be made using both functional and material causes for establishing the graph network.

### 3.4.2 Causal Reasoning

Causal reasoning is the process of understanding the relationship between a cause and its effect. Causal events usually have a temporal aspect in which the cause is antecedent to the effect. Graph structures are typically employed to represent causal pathways. Computing methods that understand and can discover causal reasoning is of great interest to the AI and



statistical fields. Turing award winner, Judea Pearl, describes causal calculus as “the ability to infer intervention probabilities from conditional probabilities in causal Bayesian networks with unmeasured variables” [219]. A benchmark of this work is the detection and description of confounding variables within data. Pearl’s criterion, which he calls “backdoor,” provides a robust definition of confounding variables, enhancing the identification of variables worthy of measurement [220]. An investigation into causal reasoning in early AI research conducted, uses pathophysiological knowledge in medical diagnoses [218]. The medical community continues to seek a causal understanding. Goodman et al. use the Framingham model as an example to identify causal factors (such as smoking cholesterol, blood pressure, i.e.) through prior knowledge and changes in risk factor [93]. Causal relationships also conduct simulations using demographic information [47].

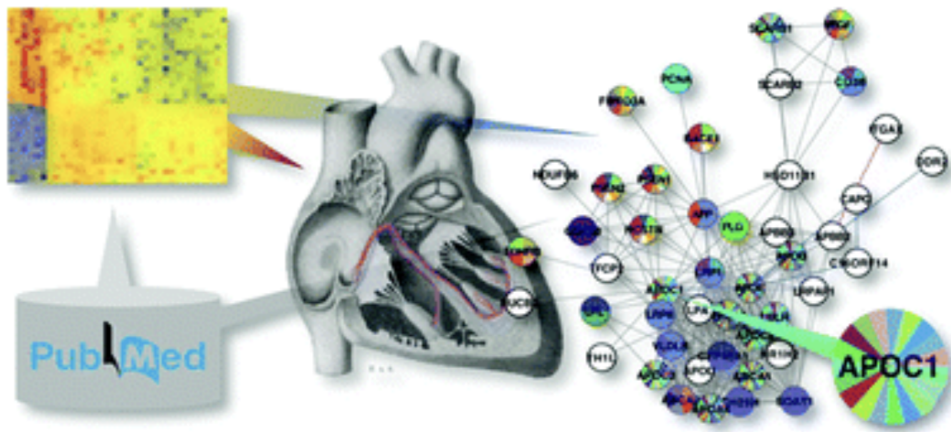


Figure 3.14: **The Heart Network.** The merging of omics data and biological domain knowledge is used to create graph networks, highlighting causal relationships.

### 3.4.3 Knowledge Structures

Similar to a database, a knowledge base stores information about the real world. The first knowledge bases were “expert systems” [44]. Knowledge in health largely consists of

understanding the output of a system after it is perturbed given an input. This knowledge arrives from established causality from scientific discovery. A graph structure can store the causal relationship of various parts of health through directed edges which store the cause as a source node, and the effect as the target node. From a molecular level to a societal level, these relationships are routinely in the scientific literature. In light of this, we review knowledge structures in graph format, also known as knowledge graphs.

Google harbors one of the largest knowledge graphs at the current time, with over 70 billion facts [275]. Knowledge graphs have been described in various forms [74]. Collectively, the aim of knowledge captured in this format is for use in further sharing and computing. Within the domain of health, there have been various efforts to capture this knowledge. Cytoscape is an open-source software platform to capture the relationships between molecular entities for biological research [244]. An example of Cytoscape software is given in Figure 3.15. Non-profit databases such as Pathway Commons store the molecular pathways used in Cytoscape [217]. Heterogeneous networks, or HetNets in short form, have also been described in biology through the work of Himmelstein et al. [117]. By gathering public domain literature into a connected knowledge base, queries are run to determine distances between entities to discover pathways for re-purposing drugs. A data-model of how these HetNets are shown in Figure 3.16. These efforts in accumulating knowledge into data structures can be extended for use in further applications, yet have mostly been limited to specialized domains of research.

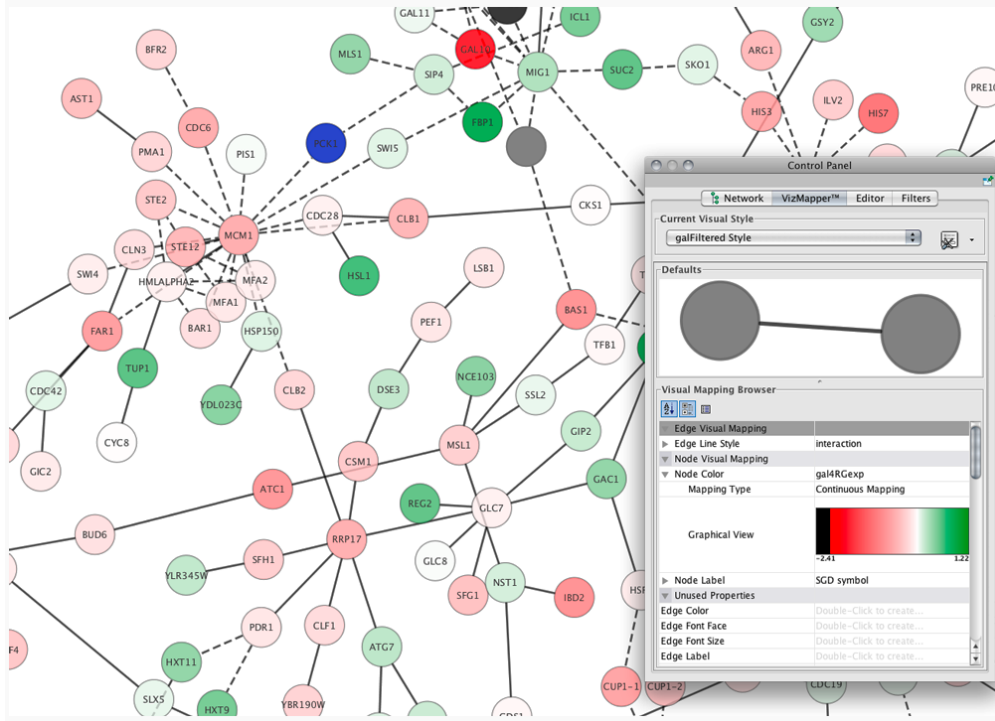


Figure 3.15: **Cytoscape**. Cytoscape is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating with gene expression profiles and other state data. [212].

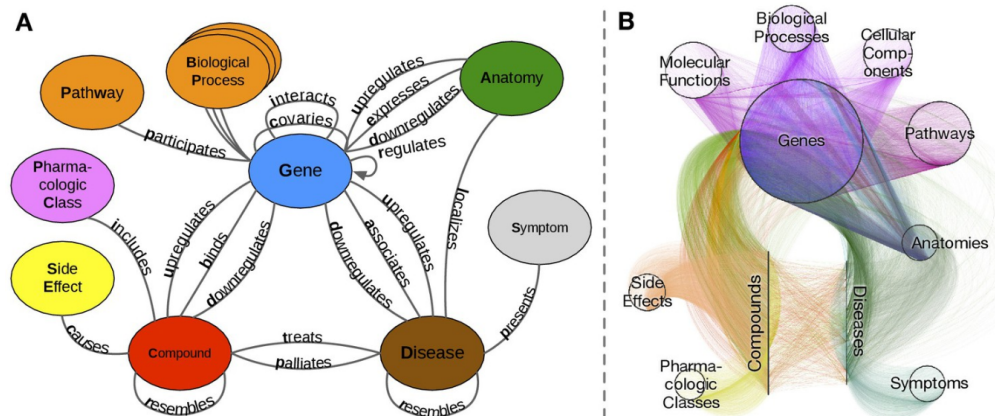
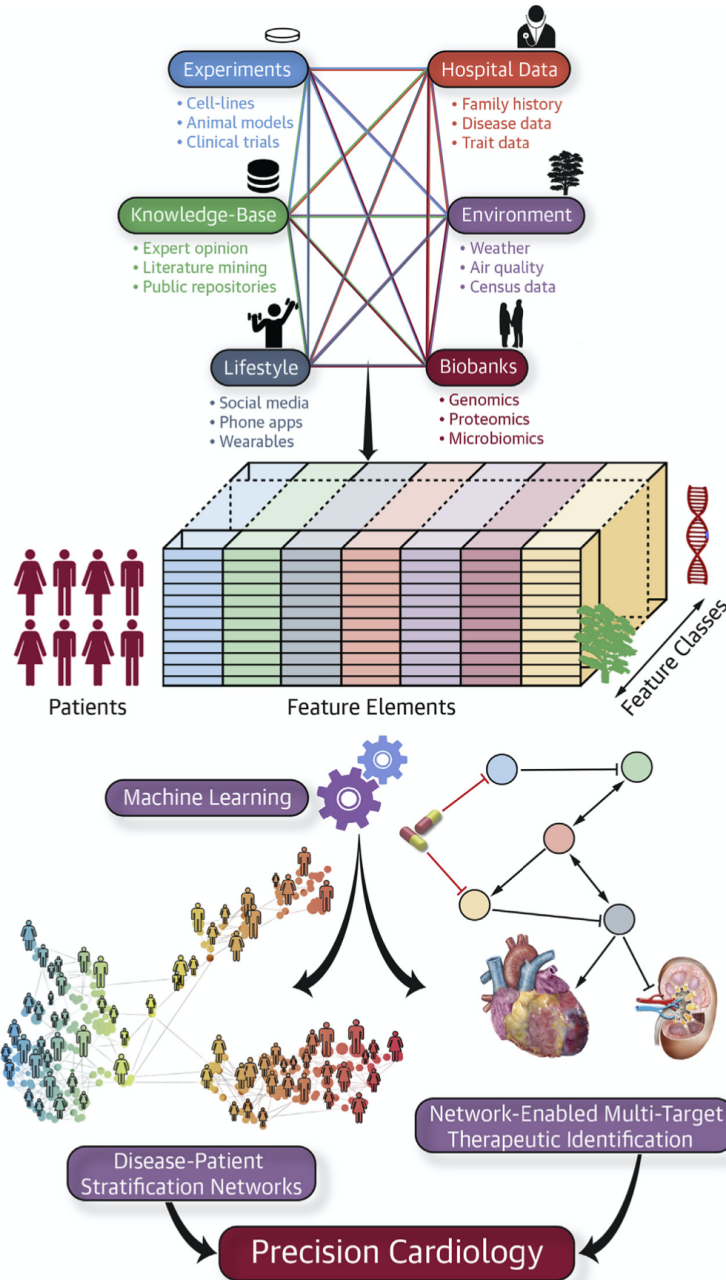


Figure 3.16: **Heterogeneous Networks in Biology**. Hetnets contain edges and nodes that have type specificity as part of their properties. Using these types as distinctions, various visualizations, queries, and algorithms can be applied. This particular HetNet maps relationships between compounds and diseases [117].

In the field of cardiovascular research, there has been a call for the integration of diverse knowledge graphs to work together. This approach has been called multi-scale cardiology

by Johnson et al. Current clinical knowledge in cardiovascular health derives insight broad clinical trials and studies. This knowledge is placed into sweeping standardized guidelines. Johnson et al. describe future research needs in applying AI and knowledge network models to enable a more precise understanding of patients and matching the best therapies for a particular person [138], shown in Figure 3.17. This work highlights the issue of integrating knowledge even within a specific sub-field of research as a formidable challenge.



Johnson, K.W. et al. *J Am Coll Cardiol Basic Trans Science*. 2017;2(3):311-27.

Figure 3.17: **Multiscale Cardiology.** There are multiple layers of consideration for practicing precision cardiology including integration of knowledge, experiments, data, environment, and person’s lifestyle. Computing techniques to extract understanding of how actions can be best taken to help the cardiovascular state are still needed [138].

As we think forward about establishing a state estimation model for health, knowledge graphs can help instantiate a model of an individual at various abstraction scales. Knowledge graphs are essential to overcome the cold start problem of personal HSE.

### 3.4.4 Systems Thinking

Systems thinking approaches problems through an eagle-eye view of interrelated components, contrasting reductionist approaches that try to isolate the view of study to a highly controlled environment within which we understand perturbations. A system may express emergent behavior when the parts accumulate, giving homage to the famous term “greater than the sum of its parts.” As described by Fang et al., “a fundamental tenet of systems biology is that cellular and organismal constituents are interconnected, so that their structure and dynamics must be examined in intact cells and organisms rather than as isolated parts” [83]. Inputs to one part of the system can propagate to other parts of the system. Systems approaches can predict how these effects will propagate.

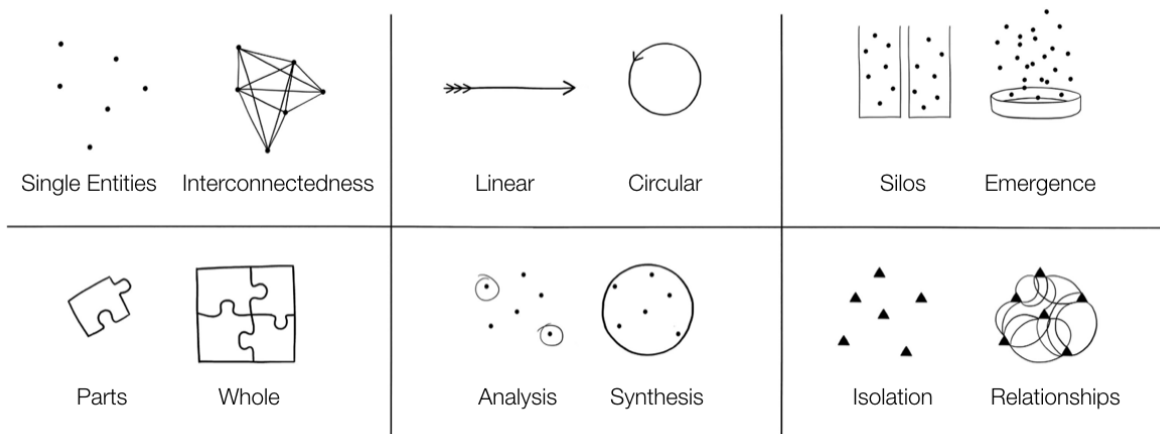


Figure 3.18: **Systems Thinking.** By considering an entire ecosystem of components, system thinking aims to understand the behavior of complex entities. This approach includes understanding how sub-entities are connected and related, how effects can circularly impact the system, what emergent properties arise from the system as a whole through synthesizing the system into a single architecture. *Adapted from Leyla Acaroglu.*

STELLA and InsightMaker are examples of modeling systems with stocks and flows [168, 68]. The user places interconnected icons and defines functions mathematically of how a particular value is flowing into and out of stocks. The most classic example of this method is a bathtub with an in-flow from the faucet and an out-flow through the drain. This approach has been used in modeling cardiovascular function, as shown in Figure 3.19 [233]. Systems models that are intelligent that change their structure remain a challenge in this field.

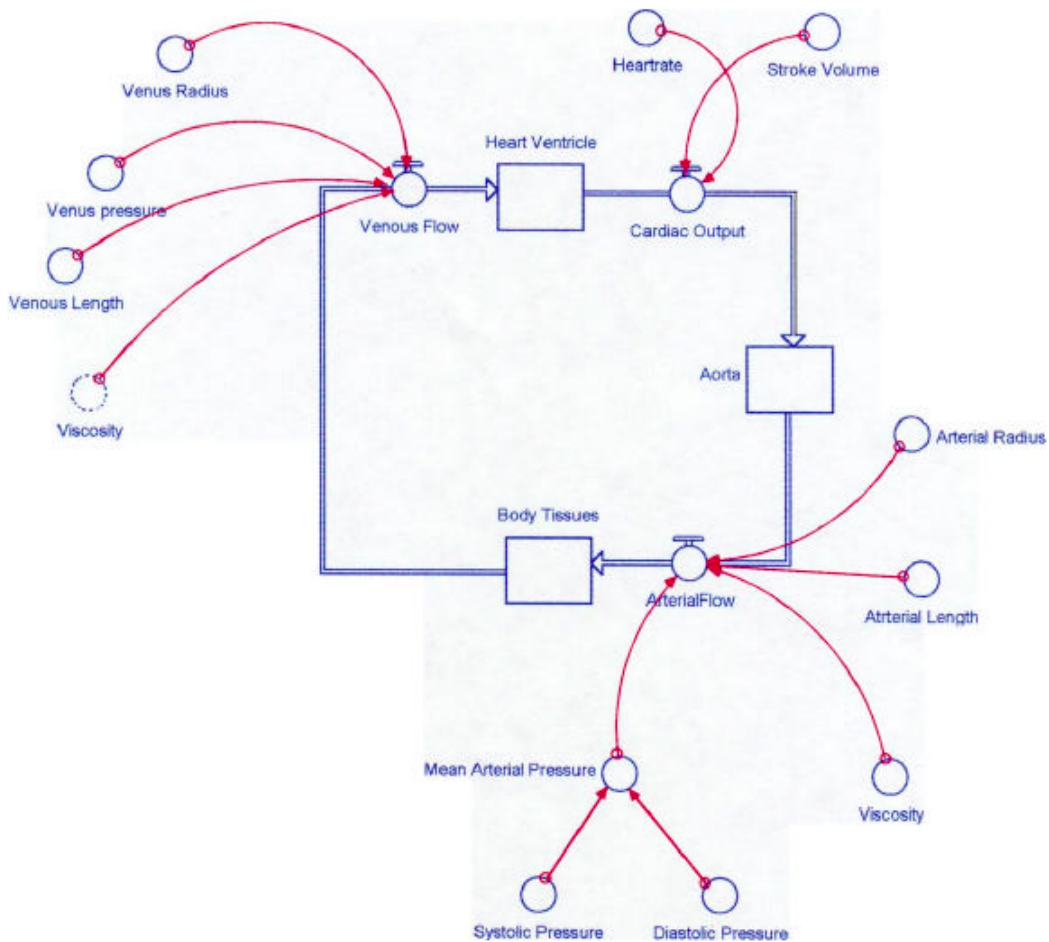


Figure 3.19: **Cardiovascular Function.** Blood flow is shown here moving through stocks and flows through the STELLA interface. This model can only focus on a single dynamic variable within these containers [233].

A robust HSE requires a systems approach due to the complexity of biology. At the same time, we must incorporate reductionist science discovery to have the best performance. To address this computationally, we look towards microservices.

### **3.4.5 Microservice Architecture**

Microservice architecture is a software development technique whereby an application is made from a compilation of loosely coupled services. These services provide a specified interaction with other services. Reductionist science is a microservice reporting on the behavior of an isolated system. An extensive collection of these systems that have knowledge-driven connections between them can start to build a systems approach to using these smaller nuggets of knowledge. The modular nature of microservices also means that a system can easily swap out, add, or delete a service if there is an updated or out-dated version without breaking the whole system structure. Hence microservices are readily understood as a graph network of interacting Application Programming Interfaces (API).

This approach is used in primarily identifying device or software performance health, and security. As examples, microservices are used in modeling IoT devices for health [116], monitoring the health of the software itself [263], and implementing security measures [77]. No one known to date has used the microservice approach to represent the health of an individual directly.

### **3.4.6 Data-Driven Learning**

Uncovering complex patterns specific to an individual requires harmony between top-down approaches through knowledge, as mentioned above, alongside data-driven bottom-up approaches. These data-driven approaches can discover patterns that human analysis may fail



to see. We can use data in learning how to adapt models to best fit reality constantly. A large part of these patterns in health and life is temporal. This approach is the particular aim of event-mining research.

*Event mining* encompasses techniques for automatically and efficiently extracting valuable knowledge from the historical event and log data [157]. *Event Mining: Algorithms and Applications* presents state-of-the-art event mining approaches and applications with a focus on computing system management [157]. Event mining can be challenging as it requires familiarity with domain literature for meaningful pattern discovery.

In general, temporal data is considered as a sequence of a primary data type, most commonly numerical or categorical values and sometimes multivariate information. Temporal data is presented in a numerical time series of values or events (e.g., EEG, stock price), click streams). Some primary goals of temporal data mining include pattern analysis, search, prediction, and clustering. Translating knowledge maps for events into a temporal visual representation [283]. Different variants of sequential patterns are described [5, 45, 155]. Multidimensional patterns are particularly interesting for health since there are usually many parameters to compare at once [312, 110, 15]. There have been extractions of knowledge in literature for event mining in the context of systems biology [9]. By using the knowledge above as a basis for constraints, the search space for mining these patterns can be reduced in computational complexity [221, 182, 222, 298]. Machine learning approach based on Bayesian networks is trained on medical health records to predict the probability of having a cardiovascular event. Unfortunately, their model fails to use high temporal resolution data [18]. Temporal patterns are just one avenue by which to use data-driven approaches. A breakthrough in machine learning and AI systems in the last decade has been through the advancement of neural networks.

*Neural Networks* represent a broad class of algorithms modeled after the human neuronal connections for pattern recognition. A network starts with input as stimuli and propagates

a signal to neighboring nodes by an activation function, such as sigmoid or tanh. Each pair of nodes have particular edge weights. Below are these activation functions and weights.

$$\text{Activation}(W(\text{weights}) * X + b(\text{bias}))$$

Training the network is based upon defining a cost function that is optimized. This concept is shown visually in Figure 3.20. As of 2020, neural networks have provided a valuable computing method to recognize patterns when given large amounts of data. This data-hungry need for neural networks is a double-edged sword. For systems that continuously produce large amounts of homogeneous data, it is advantageous. Unfortunately, large amounts of data about an are rarely collected, and thus there has not yet been much progress in using this method for personal HSE. Regardless, there has been some research into using neural networks for HSE.

We explore two avenues within neural networks for HSE: 1) temporal data streams and 2) deep learning. Temporal data streams such as signal based learning for EEG signals [206]. We also see temporal-based data streams in health condition management and predictions using neural networks [303, 104]. There are several examples of using deep learning for better classifying diseases. Image recognition in a clinical setting is prevalent. We see this in radiology, dermatology, pathology, as these fields are dependent on images [227, 63, 78]. Deep imaging continues to grow in in biomedicine, such as for predicting genetic variations and altercations [280]. Image Recognition applied to health state has a broad scope of research conducted to date and will be beyond the scope of this thesis to cover.

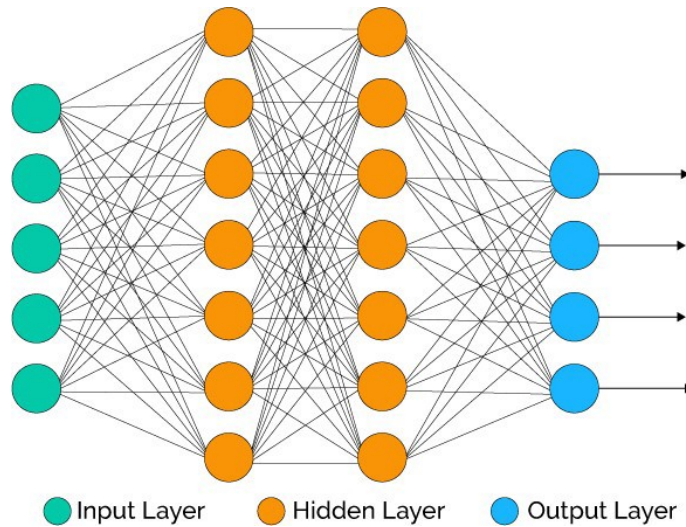


Figure 3.20: **Neural Networks.** Given an input layer of data, the neural network learns patterns and features to match an output layer.

### 3.4.7 Robust Artificial Intelligence

Recent AI progress in domains such as vision, language, and decision-making fit the strengths of neural networks. Many characteristics desired from intelligent systems are still sought after and remain out of reach. In particular, generalizing beyond the story of observable data remains a formidable challenge for modern AI, requiring approaches that structure meaningful representations and integrate data coming from the real world. We will explore approaches that benefit from the complementary strengths of top-down and bottom-up thinking below.

*Probably Approximately Correct* approaches use a combination of estimation and probability as a learning approach. Turing award winner, Leslie Valiant, describes how probably approximately correct algorithms can explain how effective models are learned without direct, explicit instructions. Satisfactory solutions are generated in the absence of any theory about the problem. This approach rests on choosing the appropriate information gathering mechanism and a reasonable learning protocol [270, 269].

*Knowledge Infusion* aims to endow computers with the ability to apply commonsense knowledge with human-level performance. Here the computer represents pieces of knowledge about the world with uncertainty. Knowledge infusion rules are learned from the world to reduce this uncertainty [271].

*Combinatorial generalization* is the ability to understand and produce a new understanding of familiar components through new inferences, predictions, and behaviors from known building blocks. This signature of human intelligence to extend data from life experiences with knowledge is “infinite use of finite means” [140, 50]. Battaglia et al. have explored implementing a combinatorial generalization approach by biasing learning towards structured representations and computations that operate on graphs. They claim, “Just as biology uses nature and nurture cooperatively, we reject the false choice between hand-engineering and end-to-end learning, and instead advocate for an approach which benefits from their complementary strengths” [22]. They explored how to capitalize on the structured representations through nested layers of graphs called graph network blocks (GNB). Combinatorics can help answer what combination of blocks is optimal in modeling the system [102].

Other researchers have attempted to build frameworks for combinatorial generalization. A hybrid model approach of mechanically producing real-world data to feed into a neural network has been attempted by combining neural networks with a physical audio engine [284]. This approach is mostly data-driven and doesn’t take advantage of knowledge, even though the authors claim to have achieved combinatorial generalization. One explored approach to using both data and knowledge has been to use combinatorial building blocks into neural networks [277]. Representing symbols through “vectors approach to representing symbols” (VARS) allows training standard neural architectures to encode symbolic knowledge explicitly at their output layers [272]. In these cases, the knowledge put into an overarching data-driven approach is the primary driver. At the time of publication, there are no known HSE approaches that use graph network blocks or combinatorial generalization techniques.

### 3.5 Unmet Challenges and Scope of Dissertation

Especially in chronic diseases, the current medical practice continues to use sparse hospital-based biological metrics (i.e., blood tests, expensive imaging) to understand the health status of an individual. The apparent research direction ahead points to utilizing a combination of the approaches above to shift towards a continuous and abundant HSE paradigm. To achieve this requires the harmony of using personal data, estimation techniques, and modeling. Enabling data assimilation will require an intelligible understanding of how sensor data relates to health status. We must use existing medical and biological scientific domain knowledge to guide the data assimilation and conversion matrices from signal to state, as shown in Equation 2.1 and Equation 2.2 with causal intelligibility. Knowing *why* an individual has a particular health status will be paramount to explanation, recommendation, and treatment. Purely data-driven methods can be components in an extensive system but will have difficulty in leveraging existing knowledge. Purely knowledge-based systems do not take advantage of data giving insight into the dynamic state of the individual. For this reason, complementary methods needed to take advantage of both approaches. Combinatorial generalization approaches allow for the pooling of understanding at different levels of abstraction, the encoding of domain knowledge, and learning from data. In so far, at the time of writing, these approaches seem the most promising for leveraging computing power to aid in the HSE problem.

We will discuss the unmet challenges in this work by focusing on bridging together several components of the literature mentioned uniquely. Foremost, continuous sensing through the self, sensors, and external agents is an absolute must for in-depth high temporal resolution HSE. A graph-based approach aims to fuse the connectedness of causal reasoning with domain knowledge. By nesting these graphs into microservice blocks, we gain the ability to zoom between abstraction layers fluidly. This ease in movement allows for a flexible, holistic, and reductionist approach to work symbiotically. These blocks can contain their unique

models and estimation techniques that suit their level of abstraction optimally. We will describe these efforts in the context of bioenergetics and cardiovascular health at an individual level.

# Chapter 4

## Health State Estimation Framework

Knowledge is power.

---

*Francis Bacon*

In the era of information technology, we harness the power to know the state of key aspects of our life. In an instant, we can know our physical location, financial situation, or the quality of our batteries. Based on this high-quality state estimation, we have a host of rich computing services that serve our needs, such as mapping directions, trading assets, and energy optimization. To date, we cannot see the details of our evolving health state in this way: on-demand and at every moment. The host of problems this issue poses are described in Chapters 1, 2, and 3. As defined in Chapter 2, health must go beyond just biological components by including elements of function. In order to know the state of our health in each moment, we need to combine all possible sources of data and knowledge into a seamlessly connected framework in the context of both biology and real-life needs. These two factors are the core of how we build a useful HSE paradigm. This chapter describes the principal concept of this dissertation by which to build a state estimation model for each person connecting these two factors.

## 4.1 Health State Estimation General Overview

The value of good health to an individual is largely based upon the wishes of how they wish to live. In order to practice a well-lived life, certain abilities determine the possibility of these wishes coming true. Take the example of an individual who is keen to play the guitar, ride a bicycle, travel to hiking destinations, or solve interesting computing problems. Each of these endeavors requires abilities for the individual to embark on the respective goal. For example, an individual who wishes to play the guitar and enjoy music must have the ability to use their fingers with dexterity, have the endurance to hit the strings repeatedly, and a heightened sense of hearing to engage satisfactorily in this activity. All these abilities depend upon harmoniously operating biology of neuro-muscular function, cellular metabolism, and auditory sensory cells. These biological mechanisms are in constant flux. So in order to determine how the state is continuously changing, we must understand what inputs cause change to the system state as given in Equation 2.1 and Equation 2.2.

These inputs are the events of our life in every moment aggregated in sequence. Each event interacts with our biology and causes changes in the system. The basis of this change rests in the central dogma of biology, whereby these events change our gene expression, metabolism, and other molecular processes, which then cause a change in the biological state [59]. Our health state is rooted in the flow of events shaping our biology, which in turn determine our abilities to do the things we want to do. This flow is shown in Figure 4.1. This HSE and modeling distinctly capture these flowing elements. The proposed framework thus contains four key abstraction blocks that serve these unique needs. The first is the semantic laminae, which are the layers of information by which humans can easily interact with natural language and visualizations. In the above example, this is at the abstraction level of wanting to play the guitar. The second is the utility state block, which captures all the fundamental functionality a human has. In the above example, this refers to the abilities to capture audio and the mechanics of playing the strings. The third block is the biological block, composed



of all the natural components that give rise to the utility functions, exemplified through the muscular, neuronal, and sensory biological systems that allow us to play and enjoy music from a guitar. The fourth block contains all the events that are inputs into the human system. In the case of a guitar player, they could be practice or jam sessions with other musicians. The four blocks are described below. The remainder of this chapter outlines the intricacies and interactions of these blocks.

These four blocks interact in a specific flow. Events in a person’s life change their biology, and these biological changes are reflected in the change of our utilities. The set of utilities relevant for a particular domain provides the dimensions by which to position the user state within the health state space for that domain. The flow of these four blocks is demonstrated in Figure 4.1. We go over the atomic construction material of this system next, the GNB.

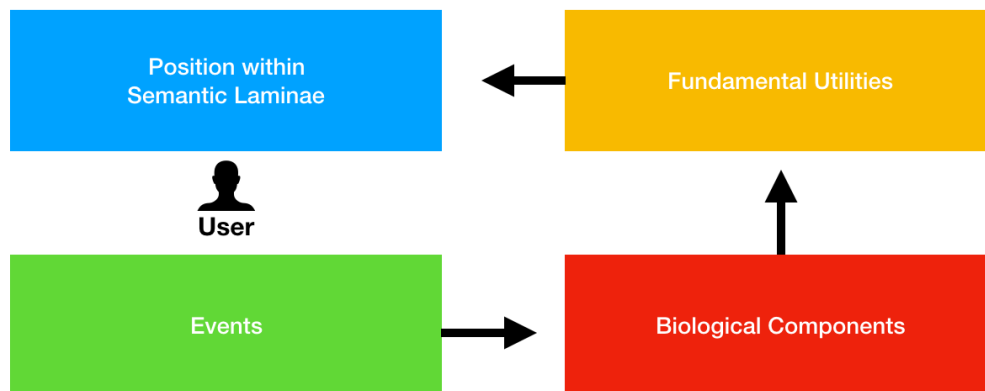


Figure 4.1: **Health State Estimation Flow.** Events drive biological changes in the user’s body. These biological components determine utility state dimension values. The combination of these utility dimensions provide the health state, akin to the GPS position. The user’s position is visualized to view the current state within a certain semantic context defined by the user intent.

## 4.2 Graph Network Block Structure

To begin describing the framework details, we start with the most basic component, a GNB. This approach takes inspiration from the work of Battaglia et al. to use relational induction

within a system to elucidate a generalized understanding of a system [22]. Because all the parts of life are related together, a graph concept suits our needs of connectedness the best. The GNB structure is defined by a directed multi-graph, summarized into a global set of attributes. A visual description is given in Figure 4.2. This GNB concept represents the utility and biological blocks in this estimation framework.

**Directed** : Edges travel from a “source” node to a “target” node.

**Attribute** : properties of each node or edge are stored as a set within the respective edge or node.

**Global attributes** : This is a set of attributes about the entire graph structure, which may be computed from the properties of the nodes and edges.

**Multi-graph** : Between two nodes, there can be edges in both directions and multiple edges in any given direction. A node may also have a self edge.

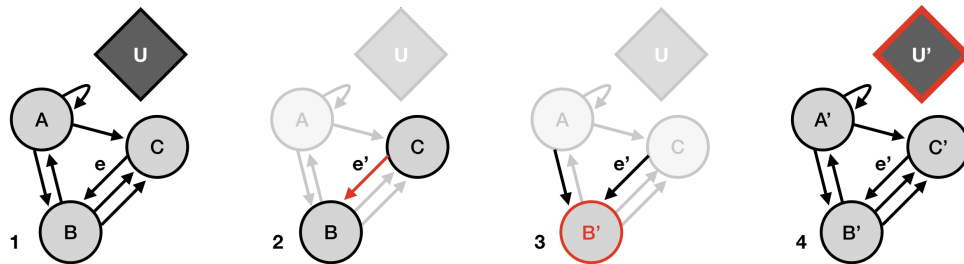


Figure 4.2: **Graph Network Block Unit and Update Process.** Each GNB describes a directed multi-graph, including multiple edges in the same direction or opposite directions, as shown between node B and C. Edges can also be to a self node as shown for Node A. The properties characterizing the entire multi-graph are summarized as global attributes within U. Step 1 shows the initialization of a GNB. In the following steps, red highlights the updates. Step 2 shows the edge update. Step 3 shows the node update. Step 4 shows the global update, a visualization of algorithm 1.

We represent the relationship between two parameters as a directed edge from the cause to its effect. Every node is a GNB with its global update. The update cycle for each block is as follows:

1. Update incoming edges: We call the update method associated with each edge coming in to block and defined within the block. For most edges, this method reads the global parameters of the source node and passes those to the target node. If needed, we can also perform more sophisticated transformations on the incoming data. An edge can be updated when there is a new update in an associated source node or an update in the relationship itself.
2. Update the nodes in the block: Every node in the block updates by applying the node's global update function on the incoming edge parameters.
3. Global update: Once all the edges and nodes in the block updates, we can aggregate the updated values of nodes and edges to update the global parameters of the block.

#### **4.2.1 Nested Microservices**

Each of these GNBs can represent a set of microservices working together to provide a higher level microservice. The output of the higher level service is within the global attributes of a GNB. The purpose of using this microservice approach is to ensure the framework is able to allow for modular exchange with any biological (i.e., organ transplant, chimera) or non-biological system (i.e., metal bones). It also allows for systems and focused reductionism to work together by zooming to the appropriate block. This structure is shown in Figure 4.3. This design allows for incorporation of the nested service layers that biological systems provide to the human body. This hierarchy of services is shown in Figure 2.4.

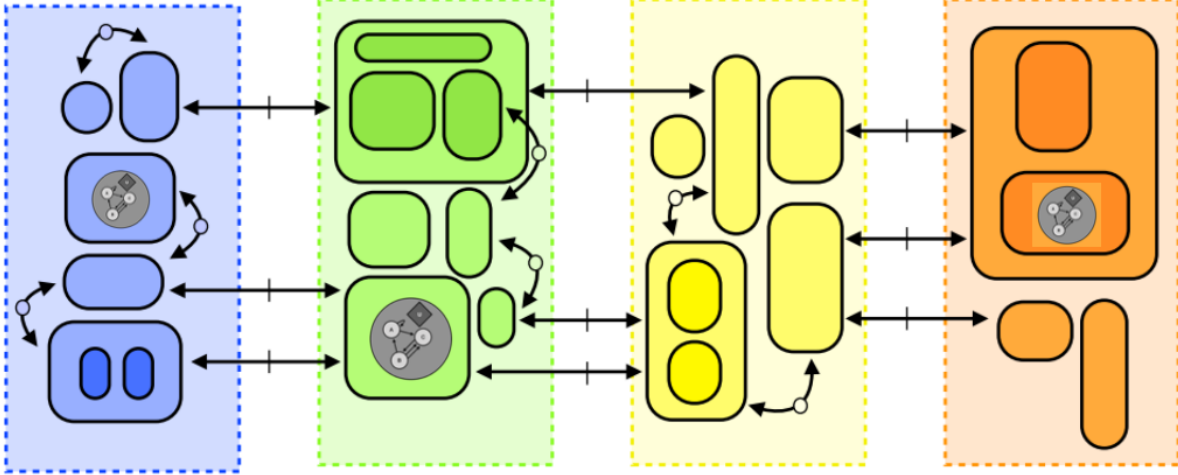


Figure 4.3: **Nested Microservices.** The GNB structure global attributes give rise to a microservice. These microservices are structured in a nested manner that can interact with each other in complex networks. This allows for the relevant level of abstraction to communicate in any system that wishes use the GNB. *Figure adapted from Meyer et al. [179].*

### 4.2.2 Causal Knowledge Infusion

In the case of the directed edges within a GNB, each edge represents a relationship between two real-world entities. These entities and relationships are built initially on domain knowledge. Thus, each edge is a causal relationship that extends for causal calculus or backdoor confounder analysis (beyond the scope of this dissertation). By using the nested microservice approach, we can insert knowledge at the correct abstraction level and can scale between depth and breadth through the blocks. The knowledge sources for each edge are also stored within the edge for reference. An example of the knowledge transformation process is shown in Equation 4.1 and Figure 4.4 as an example for cardiovascular physiology. The equation dictating cardiac output  $CO$  is below, where  $SV$  is the stroke volume of blood each heart-

beat ejects and  $HR$  is heart rate.  $CO$  at rest is determined by body weight. The knowledge source is from *Exercise Physiology: Human Bioenergetics and Its Applications* [33].

$$CO = SV * HR \tag{4.1}$$

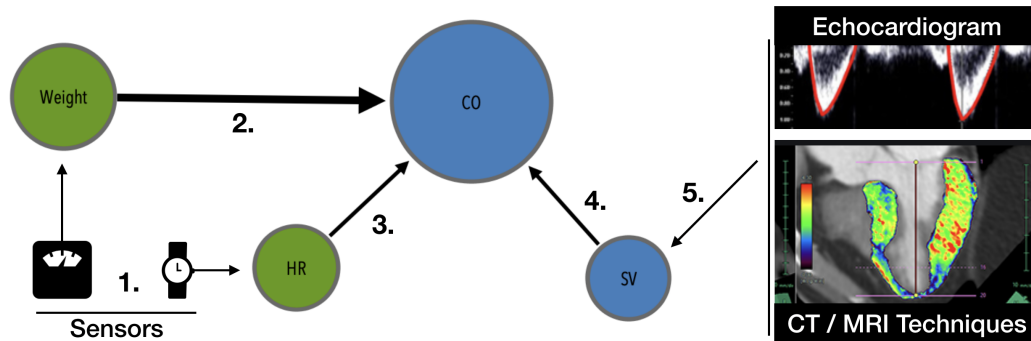


Figure 4.4: **Knowledge to Graph.** Here, we translate cardiac output equations to the graph structure. Cardiac output refers to the volume of blood that the heart can pump per minute. Once we have the relationship in graph format in conjunction with data streams from either cheap continuous sensors like a weight scale or watch, or a more costly sensor like an ultrasound, we can update the relationship to more accurately reflect the model of the individual.

### 4.2.3 Data-Driven Model Updates

The relationships encoded in the edges can be modified in the future when we identify two nodes that have incoming data streams of both the source and the target node. This approach allows for both domain knowledge and data-driven structures to co-exist and help each other. In Figure 4.4, this is shown as either low-cost or high-cost sensor data validating the value of specific nodes. If we can measure the heart rate and stroke volume, then the value of cardiac output can be known with high accuracy. We modify the base relationship of weight to cardiac output to fit the data.

## 4.2.4 Combinatorial Generalization

In any situation where we update a value of a node or edge through either a learning process or through new domain knowledge, we can recompute the global attributes of all edges and nodes within the total graph network system. These updates result in new values for all global attributes, including the utility states, resulting in a combined new generalized state of the system used in the next time iteration. Hence, this updates the value along each dimension of the user health state.

---

**Algorithm 1** Graph Network Block Update

---

**function:** Update ( $E, N, \mathbf{U}$ )  
**Input:** All blocks ( $E, N, \mathbf{U}$ ), Updated Knowledge, Attributes, Values  
**Output:** Updated Values of ( $E, N, \mathbf{U}$ )'  
**Begin**  
1: for each block  
2:     for each edge ( $E$ )  
3:         Update all edge relationship attributes with domain knowledge.  
4:         Update all edge relationship attributes with learned knowledge.  
4:         Return ( $E$ )'  
5:     endfor  
6:     for each node ( $N$ )  
7:         Updated node attributes with new values from data.  
8:         Propagate updates through outgoing edges from updated nodes.  
4:         Return ( $N$ )'  
9:     endfor  
10:     for each set of global attributes within the block ( $U$ )  
11:         Using ( $N$ )' and ( $E$ )', compute ( $U$ )'  
12:     endfor  
13: endfor  
14: Return ( $E', N', \mathbf{U}$ )'  
**End**

---

Algorithm 1 is the method by which we compute the estimated global attributes of each block microservice within the system for both biological and utility blocks at large. Each update cycle refreshes the utility dimensions to reflect the current state.

### 4.3 Semantic Laminae: Interactive Layers

Lamina (plural: Laminae) refers to a thin sheet. These layered sheets represent themed data in a particular state space, such as in the case of bicycle riding or guitar playing. The laminae integrate perceptual elements so that a user can interact at the abstraction level that is relevant for their needs. Parallel to Geographic Information Systems (GIS), the map layer selected can be a street map, satellite imagery, or weather information depending on if the user is looking for directions, planning a camping spot, or choosing to go to a pleasant vacation destination. The underlying dimensions that support these map layers are all tied to latitude, longitude, and altitude. These are the three dimensions of space that GIS systems use. For a person to exist in Irvine, they must ensure that these three dimensions are within the appropriate range. Similarly, people generally think about their life in the semantic space of natural language. Engaging in any particular goal in life requires having the person exist in a specific range of health state dimensions. These interaction layers need a foundation concept of dimension. In the case of health, we call the dimensions, utilities.

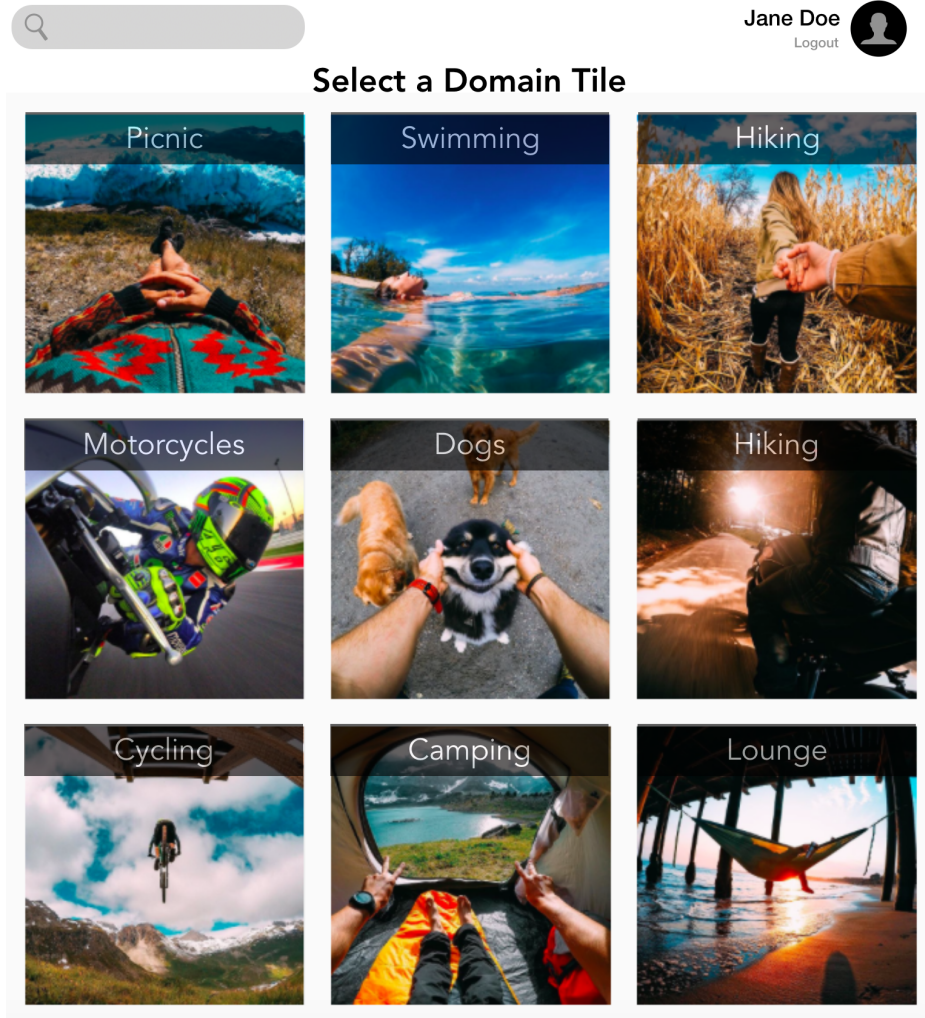


Figure 4.5: **User Interface.** A user can browse through various domains of interest, or search for a specific domain of interest in order to populate the a lamina.

User intent can be captured through various means. A sample browser interaction to capture intent is given in Figure 4.5. In this case we use natural language to query the user for a general topic of interest or regions-of-interest (ROI). If the user chooses a ROI, the associated topic is retrieved. This topic then is used to retrieve the relevant coordinates or dimensions. Mapping the coordinates of a “healthy pregnancy” ROI or “marathon running” topic in the state space requires matching relevant laminae and dimensions that are of interest to the user intent. Each lamina has a name, a set of dimensions, and a set of ROIs. ROIs each contain a label, an enclosed boundary, and a set of attributes. The operation of the



laminae within the developed system occurs through four components, which are explained in Figure 4.6. The first is a user interface where the goal or domain of interest is entered. The second component uses the user input to retrieve all laminae that relevant to the user. The user can select a lamina to view from this list. When selected, the visualization loads the utility dimensions and plots the regions of interest on the user interface. The user health state location may or may not be shown on this visualization. The state space shown on the screen can be selected to be the general health state space of all possibilities within the domain, or confined to just the user’s own health state space. This whole system relies upon the set of utility dimensions, which we discuss next.

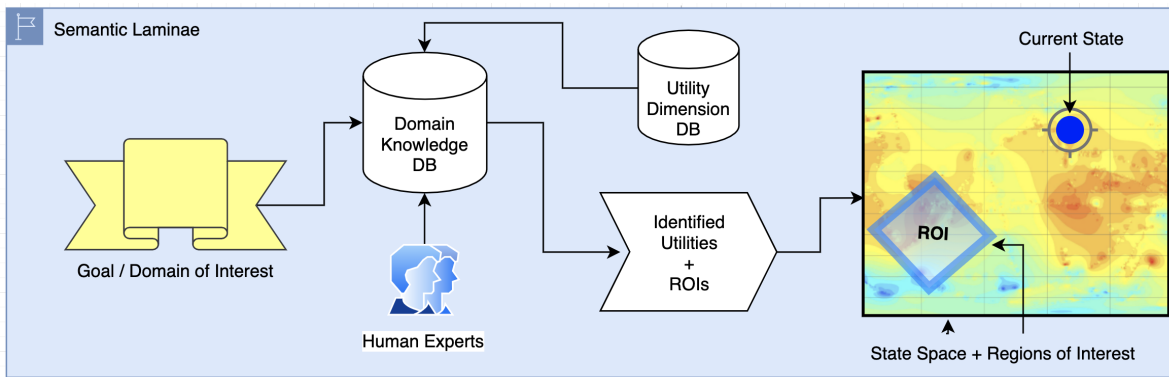


Figure 4.6: **Semantic Laminae**. For the system to initialize, the user must specify what goals or domain of interest they have, which is referenced against a domain knowledge database (DB) curated by experts of various fields. These experts outline the relevant utility dimensions and have map layers that contain ROIs. These map layers may also contain the connections between regions and the layers of information on how to move between states in the region. These connections will be expanded upon further in Chapter 6.

## 4.4 Utilities: The Dimensions of Health

Utilities describe the fundamental dimensions of health state, akin to the latitude, longitude, and altitude of GIS systems. Utilities are the fundamental abilities of a living human. In this work, we define five parent nodes for the utility taxonomy based upon consensus literature in the life sciences. These major categories include movement, sensing, thinking, managing

homeostasis, and reproduction. These primary groups have many nested sub-dimensions within them. The importance of utilities rests in abstraction away from the material cause of existence and instead focuses on the final cause of use, as described in Figure 2.5. The utilities generally arise from biological underpinnings but are not dependent on biological function. A person may have a biological joint or a metal joint. The end utility is the same: to move the body in a specific manner.

As mentioned above, the dimensions of health state space are defined through the utilities they provide. Utilities serve as a parameter with a particular value or categorical assignment. Structured taxonomy of human utility state space in this work is given in a simplified version to demonstrate the concept of utility space. These five categories within this block are shown in Figure 4.7. If “movement” is the parent node, then there are many child nodes that are part of movement. One of the child nodes in Figure 4.7 is the flexibility of the skeletal system, which encapsulates the entire summary of joint function in the body. A list of joint articulation utilities is in Table 4.1 to show how there can be categorical strings, Booleans, floats, or integers as attributes for each utility. Because the majority of a person is generally composed of biology, we need to connect how utilities arise from biological systems.

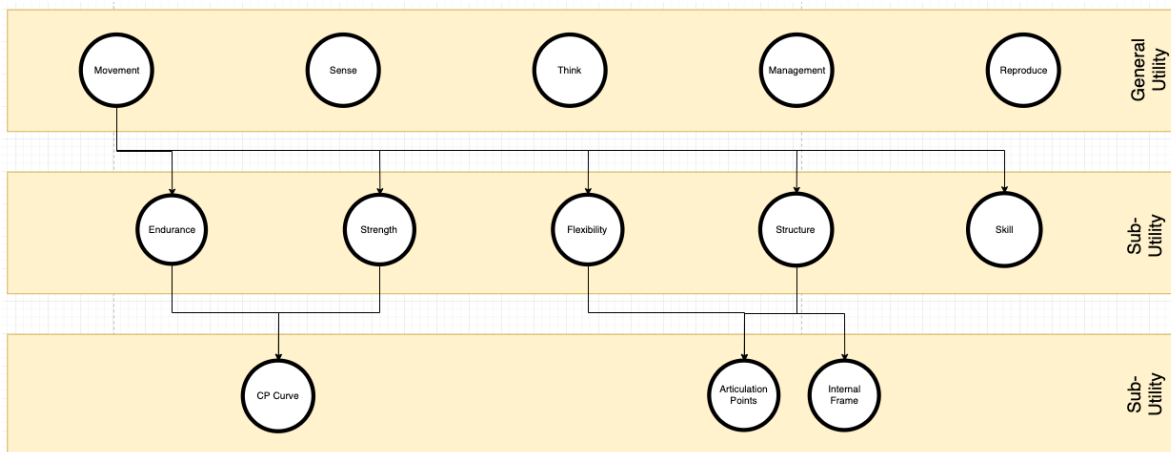


Figure 4.7: **Basic Utility Taxonomy.** These nodes show the nested microservice architecture of the utility dimensions. The movement node is expanded to reveal the next layer of microservices. This layer can be further expanded to more detailed child nodes. This is a simple overview of the utility GNB at large.

Joints	Joint Classification	Range of Motion
proximal radioulnar joint	Pivot joint	190
distal radioulnar joint	Pivot joint	210
median atlanto-axial joint	Pivot joint	50
elbow	Hinge joint	150
jaw	Hinge joint	30
proximal interphalangeal joints	Hinge joint	90
distal interphalangeal joints	Hinge joint	100
foot	Hinge joint	100
knee	Hinge joint	150
talocrural joint	Hinge joint	50
subtalar joint	Hinge joint	60
distal tibiofibular joint	Hinge joint	45
carpometacarpal joint	Saddle joint	53
sternoclavicular joint	Saddle joint	35
incudomalleolar joint	Saddle joint	20
acromioclavicular	Plane joint	30
metacarpophalangeal joints	Condyloid joint	40
metatarsophalangeal joints	Condyloid joint	45
hip joint	Ball and socket joint	300
shoulder joint	Ball and socket joint	320

Table 4.1: **Joint Utility Table.** This table is a sample list of human joint utilities. For each utility, there is a set of output attributes that determine the utility value, in this case “Range of Motion.”

## 4.5 Biological Blocks: Natural Micro-Services

This block contains the material cause of what humans are composed of [81]. The biological components all serve certain utilities. The isolated state of biology has been traditionally the study of bioscience and is the target of health state study in most medical fields. These components nest in various levels of scientific and spatial abstraction, which can be molecular to an organ system. The microservice architecture for the GNBs allows for representing these levels of abstraction. The nesting is represented in GNB format in Figure 4.8.

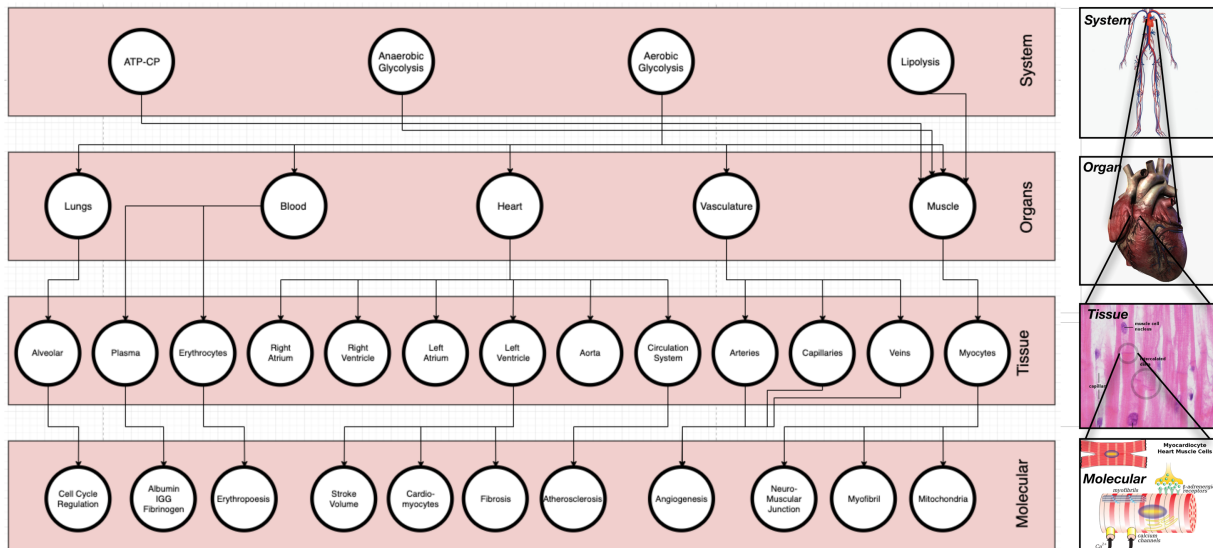


Figure 4.8: **Nested Biological Microservices.** Human biology is represented as GNB nodes at various layers of abstraction within the framework. This figure translates the cardiovascular system knowledge layers in Figure 2.4 into the GNB structure. The four chosen abstraction layers for this work are the system level, organ level, tissue level, and molecular level.

Personalized information that is in this structure includes anatomical, physiological, cellular, and signaling interactions. Demographic information about an individual such as age and gender may also be used to personalize the biological blocks. Genetic information is stored as relationship or node attributes that modify a generic model. Biological systems are highly

complex and always in flux. A changing system inherently is temporally dependent on its inputs. Temporal data about an individual is captured in the details of their life events.

## 4.6 Events: Inputs to Life

What changes our health state, and when does it happen? Every event at each moment of the day affects our health. Investigating the significant factors that impact health, we see that life events are the central premise behind changes in health state. As shown in Figure 4.9, these factors are all part of daily life events for an individual.

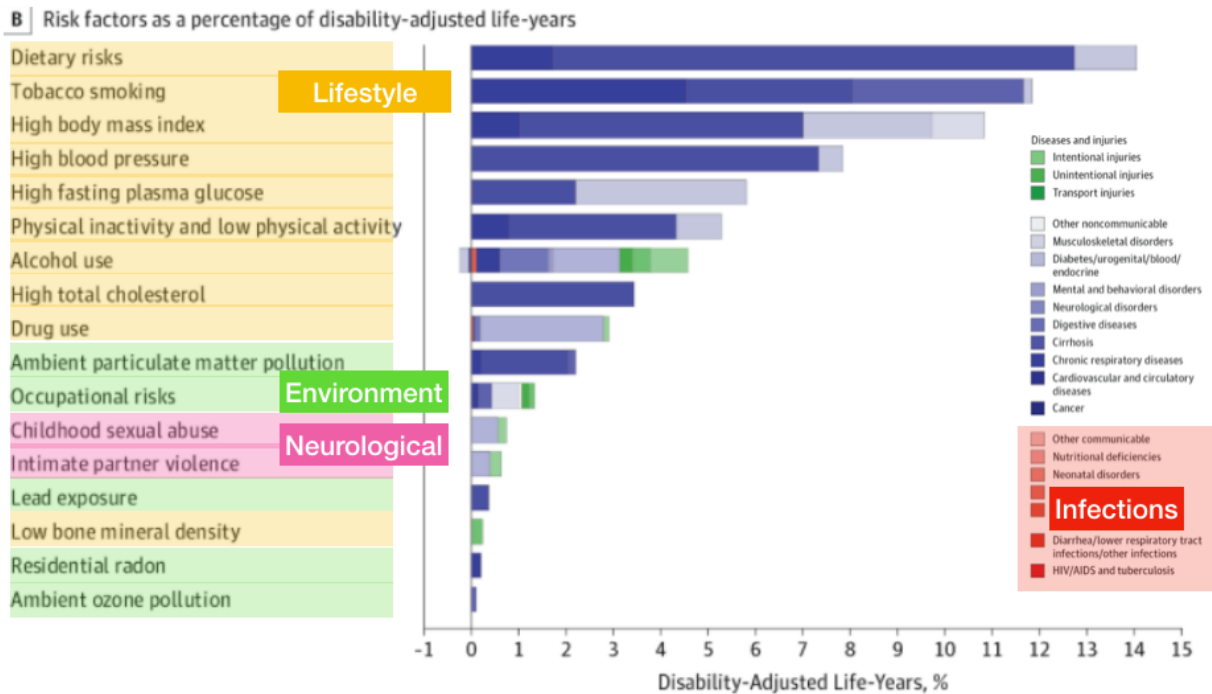


Figure 4.9: **Major Factors that Change Health.** This figure displays the burden of health factors in the USA for the 20 years of 1990 to 2010 [185]. The top three items that change the health state of an individual in the United States are lifestyle events, followed by environmental and neurological events, validating the necessity of tracking events for estimating health state.

A person's life history is traditionally documented with a biography. A biography consists of a series of events that flow over time. These events of life happen at varying time-scales. Take, for example, nanosecond molecular interactions of cell receptor binding compared to a four-year college degree. The difference between these time scales is 16 orders of magnitude. The temporal scope of life events is shown visually in Figure 4.10. Event architecture allows for a flexible design where various event types are recorded. Within any particular time window, a set of events will contain inputs that the change in biology, a phenomenon constant from conception until death.

The HSE framework classifies events into three main categories. The first category, the exposome events, are all events happening outside the body of the person, including all environmental exposures of light, sound, weather, pollution, and any other physical attribute outside the skin. This environment can also include social events or interactions with external people or agents. The second category of events, the life events, includes all events the person's body engages in from a molecular level to the whole body. At a coarse level, these events include any activity, food intake, sleep, or any general life events as described by the Personicle system or Daniel Kahneman's day reconstruction method [205, 141]. At a granular level, this could include events of organ systems such as the heart (such as an arrhythmia event, volume overload event, pressure overload event), pancreas, or gut. Even more granular are events such as reactions of molecules such as ultraviolet light damaging DNA. The third category of events is called neurological events, including all events that are perceived by the mind. These events are generally subjective in nature and include events such as joy, stresses, or dreams. In summary, the three classifications of stored events are everything outside the skin, inside the skin, inside the mind.

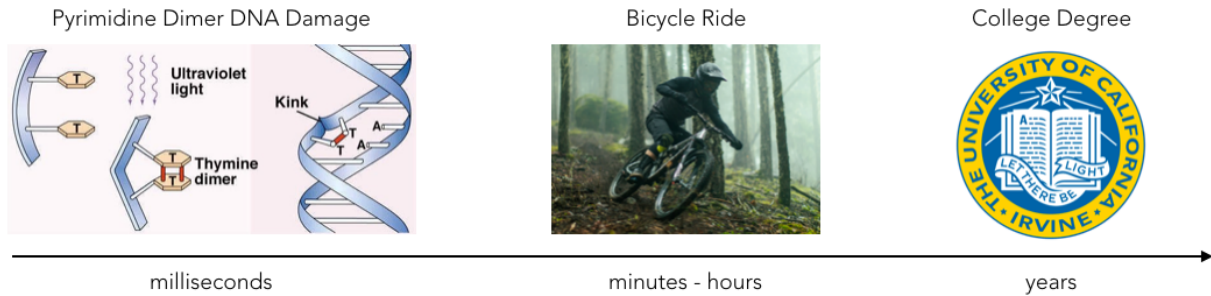


Figure 4.10: **Personal Events: Time Scale of Intervals.** The temporal scale of events that happen to a person have a large range that can span over 15 orders of magnitude.

Events happening to a person are captured and then stored in a DB system with three containers. The first container is an event stream, where interval-based events can be stored and retrieved. The second container is the stream database, where continuous data streams can be stored and retrieved. The third container is a specialized version of the stream database, where the location data stream can be stored and retrieved. A table characterizing this storage schema is in Table 4.2. Operationally, the event module connects to incoming data streams through parsers and adapters. These data are stored in a Raw file DB and redundantly backed up as the most original form of the user’s data. After cleaning the data of erroneous data through predetermined filters, the data is translated to either the event format or the stream format and copied into the Event DB or Stream DB appropriately. The location stream in the Stream DB then fetches environmental data at the particular location and time stamps the user has data to gather weather, pollution, demographics and any other requested geospatial data. Both the Event DB and Stream DB are used in the Interface Event Retrieval process, which we will go over in the following section. A visualization of this framework is given in Figure 4.11.

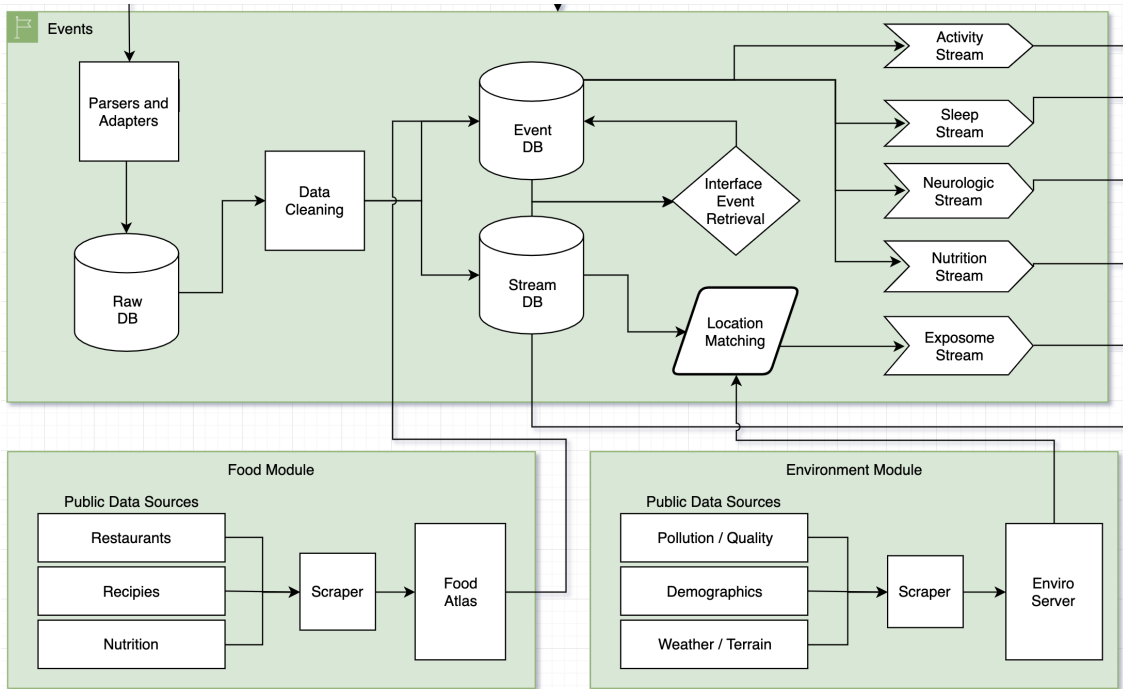


Figure 4.11: Event Processing Framework.



Stream	Field name	Data type
Event	Event type	String
	Event name	String
	Start time	Timestamp
	End time	Timestamp
	Parameters	Key-Value pairs
	Data streams	Set
Real valued Data Stream	Timestamp	Timestamp
	Value	Numeric
	Unit	String
	Source	String
Location Data Stream	Timestamp	Timestamp
	Value	Location
	Unit	String
	Source	String

Table 4.2: **Event Schema.** Event schema uses an interval approach to define the event and associated parameters. Data streams are split into two schemas, one for general data streams, and one specifically for a location that includes a geospatial value of latitude, longitude, and altitude stored as “LOCATION.”

We conclude the introduction to the four major components of the HSE framework. To set up the system, we jump into initializing the system in the next section.

## 4.7 Initializing the System

In order to begin the state estimation process for the individual, we need to populate the four main components concerning the person. The state space of all possible dimensions of health is vast, and to narrow the state space to an appropriate level, we query the user for their particular interest. Given this intent, we can give a list of laminae that match the intent. The user can then select to visualize a particular lamina. Once this lamina is chosen, the relevant dimensions must be populated to generate the state space. This process requires retrieving the utility dimensions and the supporting biological blocks, as visualized in Figure 4.12.

1. **Obtain User Intent.** By providing a certain intent, the user can request to view their health state on a lamina that is their interest. Similarly, if a computing agent wants to understand a particular aspect of the health state, it may also have an intent.
2. **Retrieve Associated Semantic Laminae.** Given this intent, relevant layers are retrieved from a database with domain knowledge of the particular query intent.
3. **Populate Utility State Dimensions.** These layers are placed upon their respective dimensions of utility.
4. **Populate Biological Blocks.** Given this set of utilities, the system then populates all relevant biological components that determine the utility state.

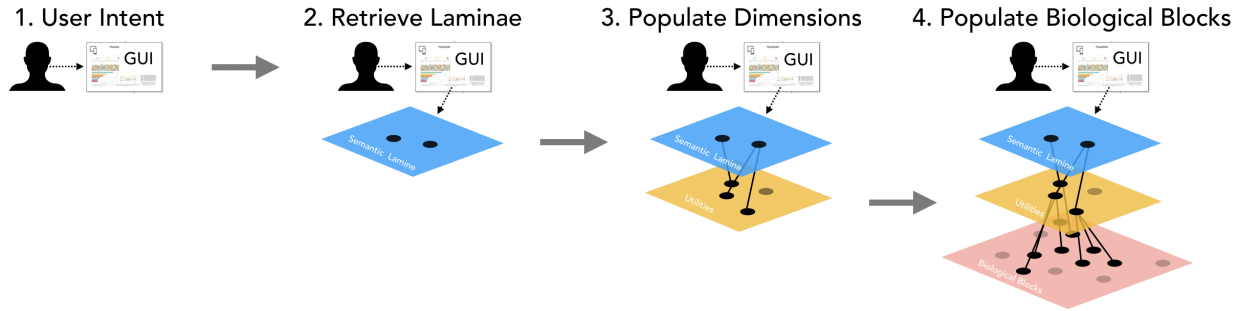


Figure 4.12: **Initialization Steps.** These are the four steps in initializing the graph structure for the individual. **1.** First, we query the user for their intent. **2.** Using this query, we retrieve the appropriate laminae they can visualize. **3.** These laminae require a certain set of utility dimensions to be displayed. These are populated next. In order to know the user position on these dimensions, the appropriate biological blocks need assessment. **4.** The fourth step populates these appropriate blocks to monitor.

## 4.8 State Update Method

There is constant flux within an individual’s personal health state space (PHSS). We must be able to assign an accurate location within the PHSS for state estimation. Understanding this location in real-time is best powered by multi-modal data sources that update the health estimation. For example, monitoring a cardiovascular health state is useful to both endurance athletes and heart disease patients. Estimation techniques have been of great interest in designing many applications, but health applications will require increasingly deep biological knowledge layers to define and estimate health states. This process uses the knowledge generated graph structure model from the system initialization of an individual to deliver updated states. To compute the state of health using this flow, we need to continually compute the individual’s position along each dimension of utility from any new source of data or knowledge. Below is outlined the step-by-step overview of how this occurs. Figure 4.13 shows a visual flow of how these steps carry out.

1. **Initialize System and Connect to Personicle.** As given above, the set of utilities and relevant biological components are populated. The person's event and data history are then loaded from a personicle.
2. **Time Based Updates** Any nodes with self-edges update within the biological blocks. This will be the only update within the whole cycle if there is no data in the personicle. The only piece of information used here is the time since the last update.
3. **Event Based Updates.** If there is an event that has occurred within the personicle, the interface event retrieval system gathers any events that connect to the populated biological states. The attributes of these events determine the change in a particular biological block. Interface event retrieval is in more detail below.
4. **Observed Biology** If there are direct observations in the personicle of any biological components (nodes or edges), they are updated.
5. **Observed Utilities** If there are direct observations in the personicle of any utility dimensions (nodes or edges), they are updated.
6. **Update Utility dimensions.** All estimated values are propagated forward along the directed edges to update the utility nodes required to show the position of the user in the health state space.
7. **Visualization of Current Health State Position.** The updated values for the utility states are then shown as an updated position on the selected visual lamina to the user.

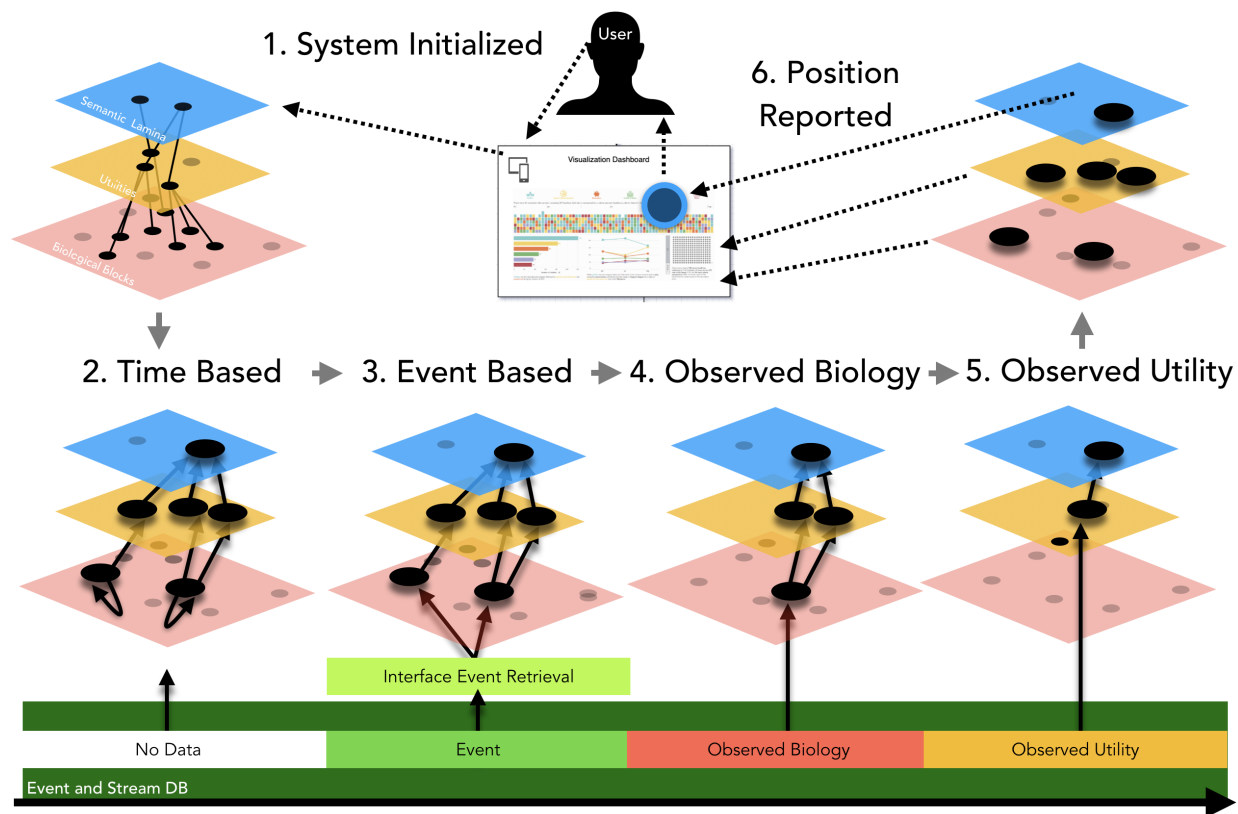


Figure 4.13: **State Update Methods.** **1.** The system initiates from the user intent. **2.** If there is no data from the event or stream DB reported, self-edges update nodes which then propagate along the directed edges to give new utility values. **3.** If interface events are detected from the event stream, they update the relevant biological states, which then propagate. **4.** If there are observed biological attributes directly from sensors, they update the respective biological nodes. **5.** If there are observed utility attributes directly from sensors, they update the respective utility nodes. **6.** The health state position is reported on the lamina of interest, and/or can be viewed as raw utility dimensions, and/or can be viewed as biological attributes.

#### 4.8.1 Interface Event Retrieval

Interface events are particular sub-events with specified attributes that are responsible for a particular change in health state, interfacing between the actions in our lives and the state of our health. After combining general events and data streams from different sources, we specify the interface events as logical statements. Each interface event has direct causal edges on particular biological nodes. These events and relationships are well studied in biomedical

literature and can be defined as a combination of data and event streams using an event operator language. When we detect an interface event, the relationship with the biological variables contributes to the update of the health state. This work will retrieve interface events from five main event categories in chapter 5. These categories include nutrition, exercise, neurological, environment, and sleep events. Logical statement examples are given below describing left ventricular volume overload cardiac events (*VolOverload*), and pressure overload left ventricle cardiac events (*PressOverload*).

Representations of these events are given below:

$$VolOverload := (HR > 140) \vee (Cycling \wedge detect - climb(Altitude))$$

$$PressOverload := detect - spike(HR) \vee (Power > 400W)$$

The state update mechanism uses the same node and edge structure as new data arrives. In order to change the graph structure itself, we need to update the GNB model itself.

## 4.9 Individual Dynamic Model Updates

To improve the performance of the GNB model, we use both data and domain knowledge approaches to edit the GNB structure itself. We update the node attributes, edge relationships between nodes, adding new nodes, adding new edges, or removing edges and nodes. There are four key methods by which these updates can occur, one based in knowledge-driven approaches, and the other three based on data-driven approaches.

Knowledge-based approaches reference a knowledge DB in order to establish the graph structure. When a scientist discovers new knowledge or a new genetic association, the system can access these updates to update the individual GNB. An overview is given below in a list format. Figure 4.14 describes these four approaches visually.

1. **Knowledge Update.** If there are new static knowledge that is associated with the GNB or individual (such as their genetic information), they apply to the relationships within the graph.
2. **Event Mining.** Event mining establishes relationships on a temporal scale between events. This work does not directly address event mining. It provides a framework that allows for this type of approach in future research.
3. **Parallel Observations.** When data is observed regarding two nodes that are connected, the relationship can be updated using the data relationships. If the predictive quality is suitable, the respective edges are updated with this relationship.
4. **Cross-Modal Estimation.** If there are multiple relationships with a certain node, they can be used in combination to learn their relationship to the target node. If the predictive quality is suitable, the node updates with this relationship to its sources.

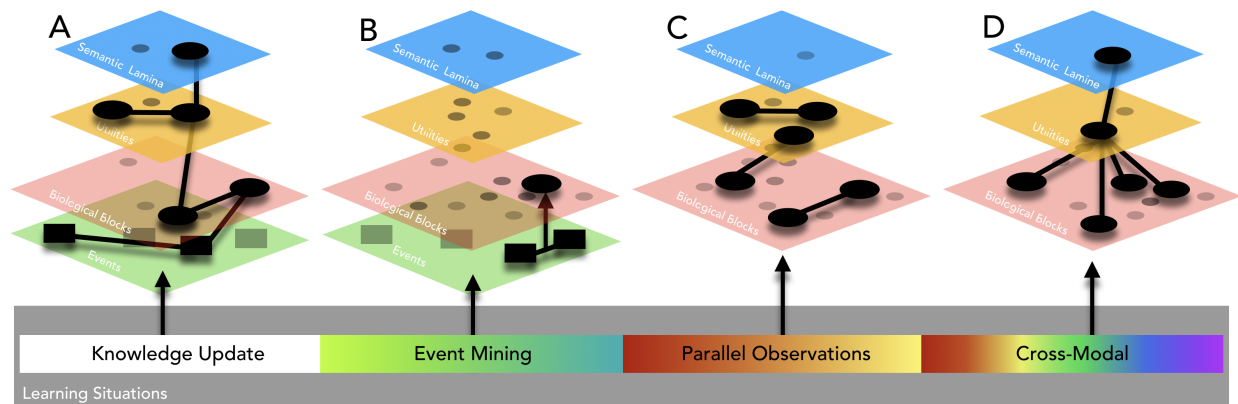


Figure 4.14: **Learning Situations.** Relationships between nodes update in both inter-layer edges or intra-layer edges. **Situation A.** When new knowledge is given to the system, the appropriate edges or nodes are updated. **Situation B.** Event-mining can discover relationships between temporal events within the user personicle. These events can reflect an update to a biological node that models behavior or any other biological component. **Situation C.** When two nodes that are connected are both measured in parallel, the relationship between the two nodes can be characterized to match the incoming data. **Situation D.** When many source nodes are related to a target node, we can use all the nodes to learn how they impact the target node.

The above two sections cover the process of how we estimate the health state and continuously modify the graph structure model of an individual to reflect what we learn from data. In order to verify if these approaches satisfies the needs of HSE, we compare this framework overview against the design requirements.

## 4.10 Revisiting the Design Requirements

Based on the discussions so far, we have defined the following design principles essential to the HSE framework in chapter 2. Below describes how we meet these particular demands with the above framework.



1. **Participation with User.** We begin addressing the design requirements by directly engaging with the user's intent in the semantic laminae. This is the root of instantiating the system and ensuring that it works towards the user's goals.
2. **Encompass Holistic Definition of Health.** Because a holistic definition of health goes beyond biology, we abstract health needs into utility states.
3. **Capture Connectedness.** In order to capture how aspects of health are constantly evolving as a system, we choose to use a graph-based approach.
4. **Flexible and Modular System.** By nesting the graphs into blocks, we can take advantage of microservice architecture to represent and model the components of interest.
5. **Zoom Fluidly Between Abstraction Scales.** These nested blocks allow the user or system to interact at an abstraction level that is appropriate for the situation. Scales can range from molecular to person level health.
6. **Causal and Explainable.** To have a causal and explainable system, we choose to have directed edges in the graph that represent causal relationships. Understanding the root of how something is changing can be traced in reverse on this causal pathway.
7. **Domain Knowledge Integration.** To determine what directed edges to add to the system, we reference domain knowledge to establish the relationships as a cold start instantiated personal model.
8. **Diverse Data Fusion and Transformation.** Using the high-level abstraction of events, biology, and utility, we can transform various types of data into these three layers.
9. **Perpetually Updating.** When new data arrives into the system, we can propagate updates to other nodes via the directed edges.

10. **Personalized Modeling.** Physiologic function, psychological function, risk, and any other attributes that are related to genetics, life history, age, or other demographics data can modify the relationships between nodes to personalize the graph network.
11. **Precision through Understanding and Learning.** When data about two related nodes arrive in parallel, we can use learning approaches to characterize the relationship better. The general domain knowledge relationship can then be updated to reflect a precise model for the individual.
12. **Predictive and Preventive Capability.** By using this constantly adapting network, we can simulate inputs to the graph and compute future outputs. If a future output is undesirable, we can prevent it by testing inputs that lead to a desirable future state.

In summary, the structure of this framework satisfies the design requirements we set out to address. In the next chapter, we dive into the details of how we can operationally implement such a framework for a real person.

# Chapter 5

## Instantiating Personal Health State Estimation

A healthy vision of the future is not possible without an accurate knowledge of the past.

---

*Daisaku Ikeda*

Chapter 5 describes using the framework described in chapter 4 within a real-world implementation of personal HSE. The chapter begins with an overview of the methods and tools used to initialize the HSE GNB at an individual scale. Initialization is for both biological breadth and depth with a focus on CRF. We then cover the types of data collected, and how they are processed towards the state estimation given various data situations. Insights from the state estimation are discussed. We conclude this chapter by showing how we dynamically edit the model with new knowledge and data-streams to fit the individual better. A visual experimental set up is shown in Figure 5.1.

## 5.1 Experimental Overview

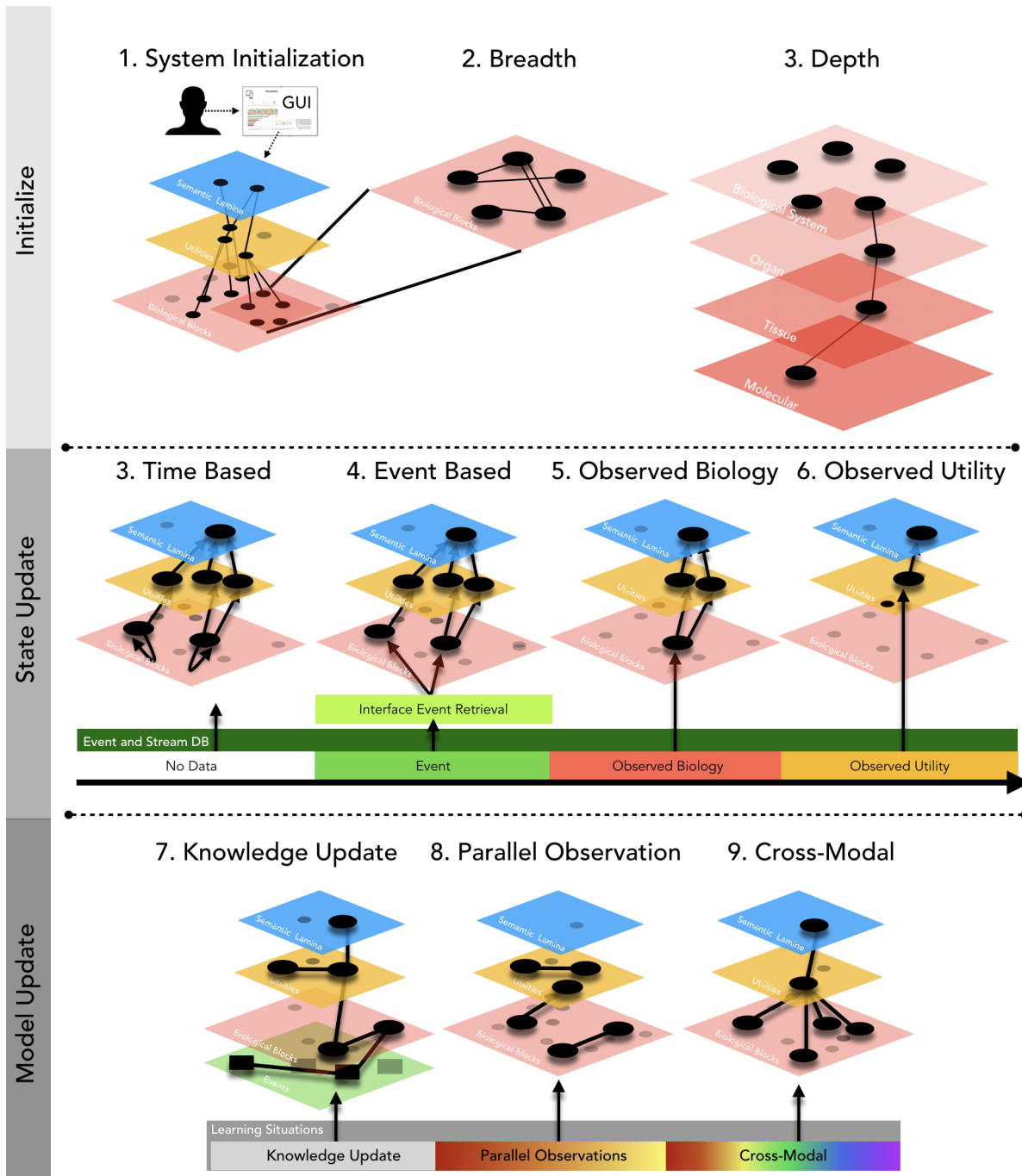


Figure 5.1: **Chapter 5 Experimental Overview.** The first part of Chapter 5 is initializing the system for the test subject in both breadth and depth. The second phase demonstrates the state estimation in action. The final set of experiments demonstrate three approaches to update the personal GNB model.

## 5.2 Methods

Applying the concept described in Chapter 4 within a real experiment set-up provided a few challenges. First, there is no dataset in existence at the level of detail required for this research, posing the most significant barrier to implementing this research. Although a great task to build such a comprehensive data collection pipeline, we decided to create the dataset organically to specification around CRF. We chose to focus on a single subject, with a personalized approach. We also needed data over a long temporal span to show how the health state is changing. Data collection is also necessary to be non-intrusive to cause minimal observer effect. We also want these experiments replicated within the scientific and clinical settings. Hence we chose mostly commercially available hardware and software. These decisions resulted in exemplifying the state estimation framework in chapter 4 within real-world settings and allow for using complex personal data. The next four subsections cover these experimental methods in detail.

### 5.2.1 Longitudinal

As we want the state estimation to provide continuous estimates over time, the most appropriate experimental approach would be a longitudinal one on a specific subject, rather than a cross-sectional study of many individuals. Longitudinal studies consist of repeated observations and measurements of the same variables, such as people, over periods (could be decades at a time). The data collected is a combination of quantitative and qualitative exposures and outcomes data of the controlled variables. This type of study is useful in determining relationships between risk factors and disease development [37].

A cross-functional observational study, on the other hand, analyzes data from a population with multiple variables at a specific point in time, often used in cohort studies [153]. These

studies can be useful to understand the prevalence of disease in a clinic-based sample [243]. However, with no information with the influence of time, the difficulty is to derive causal relationships [37].

### 5.2.2 N=1

Because we are trying to build personal estimation models, the incoming data must only be from the individual. Population-level data is helpful as a part of the knowledge-based edges, but cannot be used to personalize the graph network for the individual. For this reason, we choose to have an N=1 approach. N=1 approaches are also gaining traction in clinical settings. An N of 1 trial is a clinical trial in which one patient is the entire scope of the case study. These studies include experimental and control interventions in the design. The goal of this type of trial is to determine the optimal intervention for a patient, given data-driven criteria [158]. While N=1 studies are not highly common in the medical community, Lillie proposes further use of this method given the “contemporary focus on individualized medicine.” There has been further research conducted supporting the feasibility of N=1 clinical trials [107, 176, 204]. Table 5.1 lists several simple experimental designs to conduct these N=1 trials. When updating the state, we use an A-B method using domain knowledge to establish a causal relationship. When we update the model using data-driven approaches, we use the A-B-A-B experimental method to capture many instances of A and B and then derive the relationship.

Design	Causality	Use
A-B	Quasi experiment	Often the only possible method
A-A1-A	Experiment	Placebo test design where A1 is a placebo
A-B-A	Experiment	Withdrawal Study where effect of B can be established
A-B-A-B	Experiment	Further Validation of Above Study Design

Table 5.1: **Quasi Experiments.** Quasi experiments cannot definitively establish causality, but provide evidence towards a related mechanism (correlation or causal). Experiments establish proof through demonstration [46].

### 5.2.3 Multi-Modal Fusion

We explored the vast multitude of data acquisition approaches in Chapter 3. To have the most holistic and high-resolution health state estimate, we need data of all types about the individual. This data range includes low-cost sensors that have a high temporal resolution or low user burden and high-cost sensors that may only have sparse readings. This approach takes inspiration from sensor fusion approaches that aim to use all possible information to get better situational awareness, precision, and accuracy. Sensor fusion combines sensory data from different sources of uncertainty but provides greater insight when combined. There may be “uncertainty reduction” as the aggregation of the data can be more dependable than each piece alone [106, 75]. Distributed sensors are represented in a network with dynamically changing topology between each data source [307]. Numerical representations of sensor fusion, such as the “certainty grid,” create probabilistic geographic mapping given the assimilation of sensor data [181]. Our approach uses a centralized sensor fusion approach, where all sensors are uploaded together into a single system for analysis. When appropriate, we use weighted averages of various sensors depending on their data quality to derive the fusion.

### 5.2.4 Living-Labs Implementation in the Real World

The final choice involved in choosing methods for the following experiments was to use devices and software that is readily available to the public. Other than one of the exposome sensors, the entire suite of products used for the experiments is available in the consumer goods market without any special restrictions. We chose this approach to reduce the burden of developing proprietary components, demonstrate how the fusion of various sources is key for novel health state insight, and scale and implement these types of experiments easily for future researchers or for those who wish to implement these experiments for themselves.

Furthermore, the available sensors are minimally intrusive to collect the best continuous data streams in a natural setting. Nobel Laureate Richard Thaler’s work inspired the thought process in behavioral economics regarding nudges that unconsciously incentive our behaviors [261]. By designing the data collection to be mostly unconscious, we can help ensure that observer effects are at a minimum, and the data represents real life most accurately. Observer effects are also known as the Hawthorne effect, a type of response in which people modify their behavior when they are observed. This response can change the integrity of the research and relationships between variables [238]. Aims for quantifying the Hawthorne effect have been conducted [4, 291] and specifically in clinical trials [174, 62]. By having measurements as a background part of normal daily life, these observer effects are minimized.

The first step in starting these experiments is to collect data about the subject. We will review the tools for data compilation in the next section.

## **5.3 Data**

To fuel the HSE framework, we need to collect various types of data. The data needs are dictated by the user’s intent and any biological or utility parameters related to this intent. We use a combination of hardware and software alongside public data resources and open source-based tools to collect data about the user within our scope of CRF.

### **5.3.1 Collection Tools**

There are three main central hardware components: a computer, mobile phone, and wearable described in Figure 5.2, which tether together all the sensors and data collection through local or cloud network connections. Because the user participates in many activities, the suite of activity sensors includes rowing machines, cycle ergometers, heart rate, and cadence



accelerometers, as shown in Figure 5.3. The data streams resulting as an output from these devices are explained further in Figure 5.9 and Figure 5.10. We capture biological measurements about the user through a host of smart-connected devices, as shown in Figure 5.4. The ability of these devices to track and record metadata alongside the primary sensor without any extra burden to the user is critical towards encouraging data collection and reducing friction to use. This reduction in friction is part of the real-world implementation concept as people need to have a seamless experience to be nudged towards engaging with these tools. Some measurements are still challenging to capture through smart devices due to this need for user engagement.

In this experiment, all these measurements are related to body surface measurements are captured with traditional analog tools, as shown in Figure 5.5. Integration with costly clinical tools is also an important factor in using data for HSE. These types of tools can aid in ensuring the accuracy of lower-cost sensors and validate their use in the implementation. They can also serve as a ground truth reference from which low-cost sensors can track the change in the state. The four clinical tools that are used in this work to demonstrate this integration (Figure 5.6). Three sets of hardware are used in sensing the environment of the individual. The first includes synchronizing the GPS location from the mobile phone and wearable with the nearest public sensors and data. The second involved a host of readily available IoT devices that monitor the home setting (Figure 5.7). The third includes a low-cost constructed gas sensor kit linked to a Raspberry Pi (Figure 5.8). An overview of these hardware components and the outputs they produce is in Table 5.2.



Figure 5.2: **Central Hardware Devices.** These are the three main central devices to which other sensors link via local wireless connections of either Bluetooth, WiFi, or ANT+ protocols. The devices themselves include their sensors. The notebook computer contains a microphone, front camera, and HCI capture software. The mobile phone includes a light sensor, microphone, accelerometer, gyroscope, barometer, front, and rear camera. The wearable watch includes an optical sensor, accelerometer, gyroscope, barometer, and GPS.



Figure 5.3: **Activity Sensor Hardware.**



Figure 5.4: **Biological Measurement Hardware.**

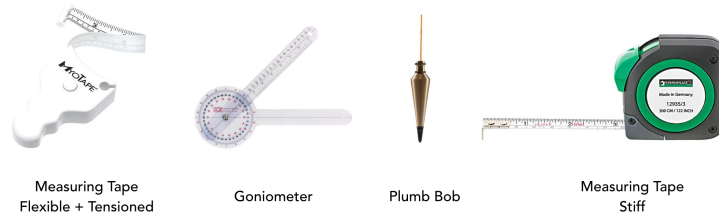


Figure 5.5: **Manual Measurement Tools.** These tools were used to capture basic physical measurements.



Figure 5.6: **Clinical Hardware.** These machines are used as examples of reference values to compare with the accuracy of lower cost sensors.



Figure 5.7: **IoT Devices.** These sensors are used to monitor the local exposome of the individual including light, sound, temperature, humidity, and smoke.

## Local Air Exposome Sensor Kit

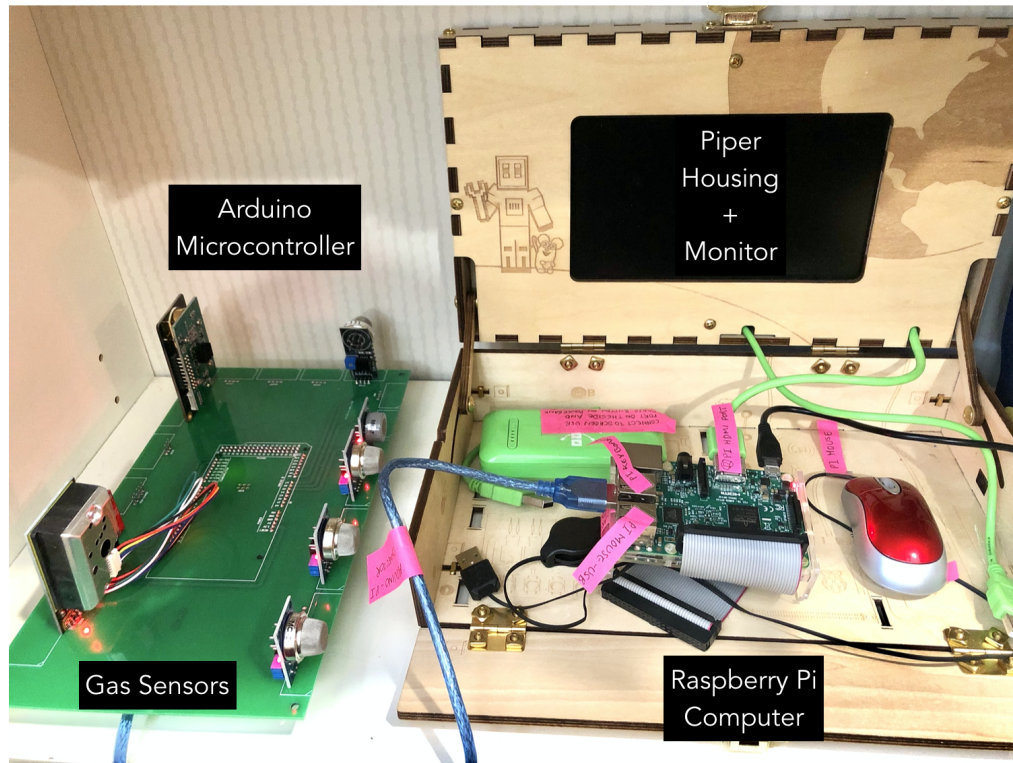


Figure 5.8: **Custom Gas Sensor IoT Device.** This sensor IoT system was built with multiple gas sensors connected to an Arduino micro-controller that links to a Raspberry Pi computer. The computer sends the signal via a web Rest-API that allows for a minute by minute monitoring of the gases in the local environment.

Software that was used as a sensor to track the user includes two varieties. The first variety of software sensors directly query the user for input. For the mental state of the individual, the open-source tracking mobile application (app) Nomie was used. The Sleep Cycle app was used to gather information about the mental state before falling asleep and assess the mood and perceived quality of sleep upon waking up. The second variety of software used as a sensor operates through background tracking that does not interact with the user. Google Timeline captures location history and the personicle of the individual. LifeCycle app is a secondary personicle capture. Garmin MoveIQ is a third source that captures background activity personicle. The RescueTime mobile app and desktop applications track screen time. Time Out software tracks desktop computer visual strain by engaging with the user for

Hardware	Manufacturer	Sensors	Datastreams
Cadence	Garmin	Accelerometer	Cadence
Crankset	Stages	Accelerometer	Torque
		Gyroscope	Force
		Strain Gauge	Power
		Thermometer	Cadence
Chest Strap	Garmin	Electrical conductance	Heart Rate
Rowing Ergometer	Concept2	Strain Gauge	Power
		Magnet	Cadence
Cycle Ergometer	Elite	Accelerometer	Torque
		Gyroscope	Force
		Strain Gauge	Power
		Thermometer	Cadence
Smart Watch (Fenix 5)	Garmin	Optical	Heart Rate
		Accelerometer	Cadence
		Gyroscope	Movement
		Barometer	Altitude
		Thermometer	Temperature
		GPS	Location
Measuring	MyoTape	Distance	Body Measurements
Goniometer	EMI	Angle	Joint Measurements
Plumb Bob	Generic	Gravity	Body Alignment
Spirometer	AirMD	Air Pitot	FEV1 / FVC
Pulse Oximeter	Hypoxico	Optical	O2 Saturation
Breathalyzer	BAC Track	EtOH	Blood Alcohol
Blood Glucose Monitor	One Touch	Glucose	Glucose
Weight Scale	Weight Gurus	Strain Gauge	Weight
		Electrical Impedance	Body Fat Percentage
Thermometer	Kinsa	Thermometer	Body Temperature
Gas Analyzer	Lumen	CO2 / O2	Respiratory Exchange
Blood Pressure Machine	Balance	Pressure	Blood Pressure
Thermostat	Nest	Thermometer	Local Temperature
Smoke Detector	Nest	CO2 / CO	Gas / Smoke
Thermometer	SensorPush	Thermometer	Local Temperature
		Humidity	Humidity
Echo Plus	Amazon	Microphone	Audio
Smart Plug	SmartLife	Electrical Usage	Electrical Usage

Table 5.2: **Hardware Sensors.**

screen breaks. The software used to track the individual is summarized in Table 5.3. Public datasets are the third source of “tracking” we use about an individual. Based on the GPS location and timestamp, we match the location to the current nearest sensors for various

environmental streams. These environmental streams are in Table 5.3. We will discuss the details of any relevant sensing in the sections below. The Appendix contains further details on hardware, software, and public data sources used in this work.

Category	Interaction	Method	Source	Data-Streams
Neurological	Active	Mobile App	Nomie	Mood Pain Symptoms
Neurological	Active	Mobile App	Sleep Cycle	Mood Rest Quality
Personicle	Background	Desktop+Mobile	Google Timeline	Location Activities
Personicle	Background	Desktop+Mobile	LifeCycle	Location Activities
Personicle	Background	Wearable	Garmin Move IQ	Activity Type Intensity
Screenome	Background	Desktop+Mobile	RescueTime	Focus Level Screen Timeline
Screenome	Active	Desktop	Time Out	Eye Strain

Table 5.3: **Software Sensors.**Software sensor systems used to track the user’s neurological state and lifestyle events.

### 5.3.2 Events

The scope of implementation around CRF determined the tools we choose for data collection regarding events that can power state estimation. The data collection methods focus on capturing the details of five categories regarding the subject’s life events. These events were chosen because they have a well-understood impact on health state in large bodies of literature. They also compose a large portion of the daily life of individuals. We will describe these five categories in the next sections.

## Activity

All instances where the user is conducting any movement, physical activity, and exercise are captured in detail. Movement and physical activity are defined as energy expended to move, occurring through any physical action taken. Exercise, in particular, is defined as an intentional activity designed around exerting effort to improve physical fitness. These terms have been well described in the literature [38]. Both activities at large and exercise events are logged in the user's personicle. The activities at large are recorded as events with time intervals of 2 minutes. Because exercise events provide a higher insight into the operational status of both the utility and biological states, they are an excellent event source to capture in detail for state estimation. Therefore exercise events are captured at 1-second intervals. Figure 5.9 and Figure 5.10 show sample exercise event data-streams. Table 5.4 lists all the sensor streams collected during the activity.

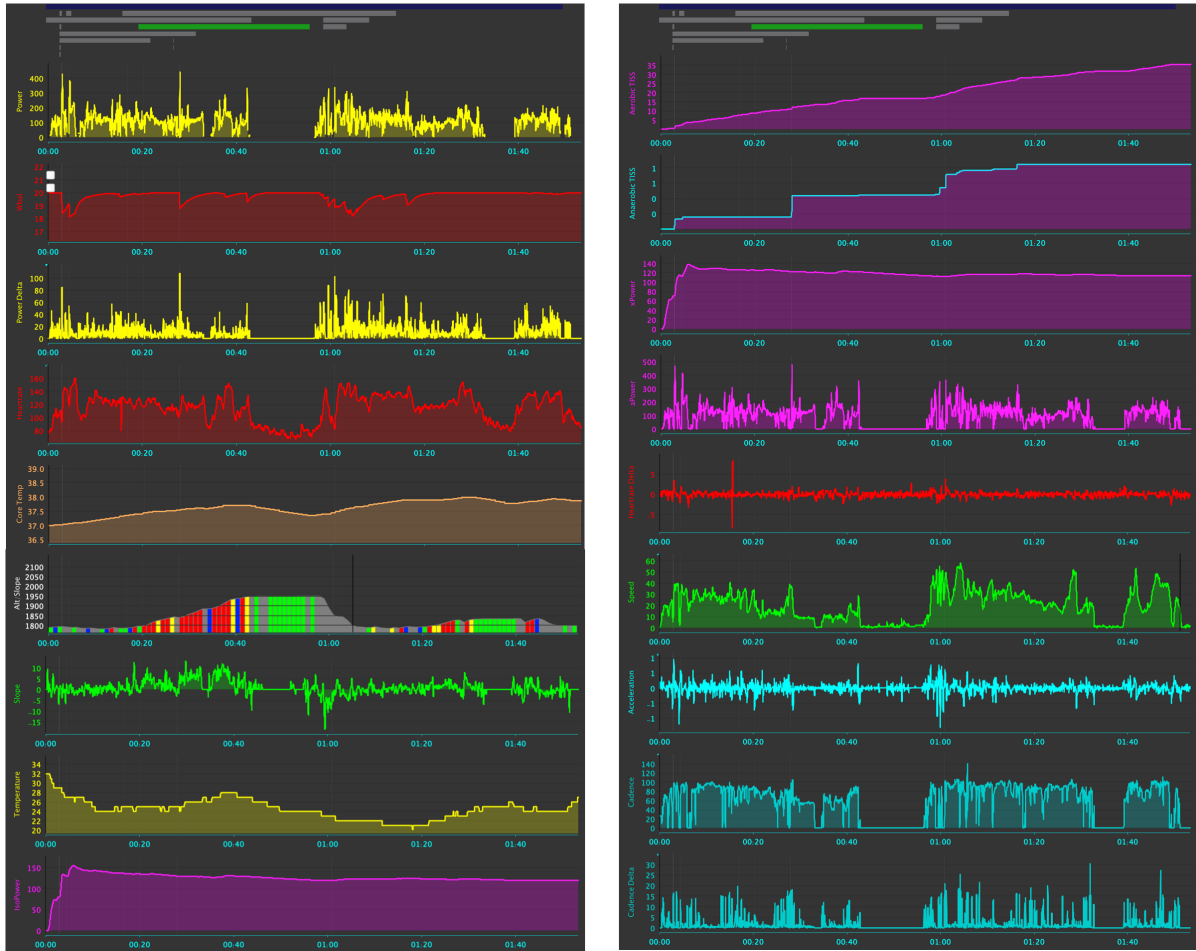


Figure 5.9: **Exercise Event Understanding.** Single exercise event streams.



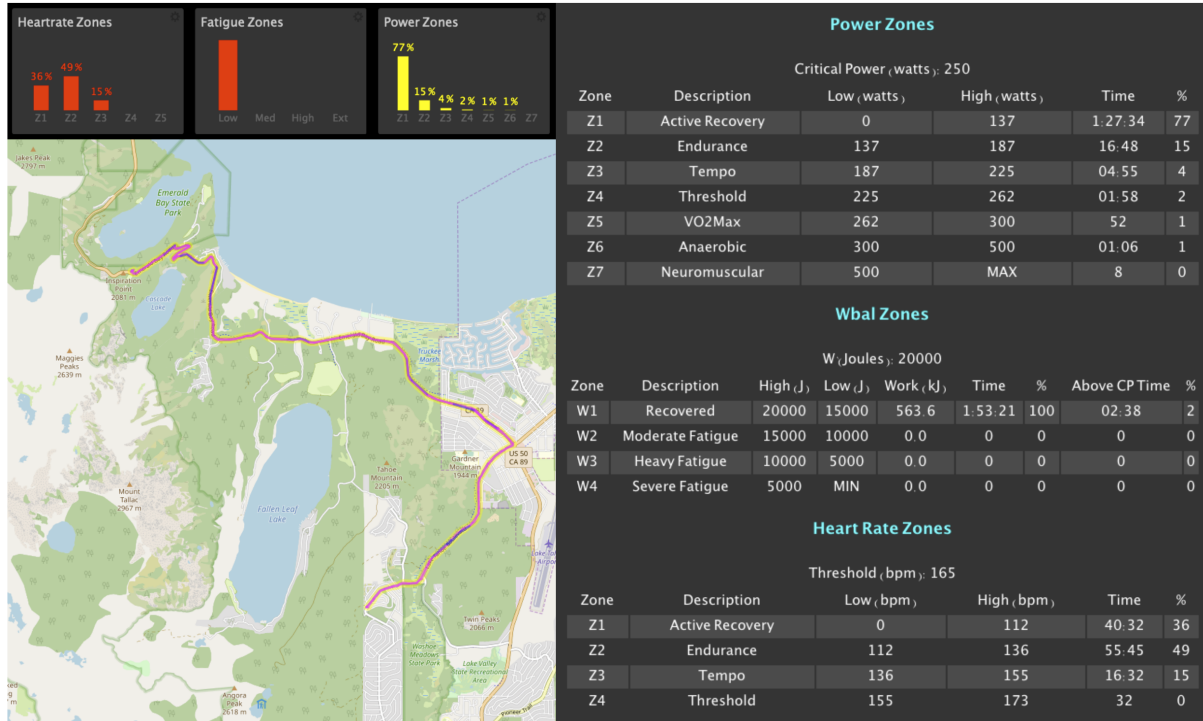


Figure 5.10: **Exercise Event Understanding.** Single exercise event summary with location data, heart rate zone charts, and power data zone charts.

## Nutrition

Epidemiological data suggest that food is the most influential event on human health, according to the Journal of the American Medical Association [185]. Nutrition includes all energy or nourishment obtained from food intake. All nutrition events include any food consumption that occurs. Nutrition events are tracked at each instance of food or beverage ingestion through images. Images are run through the Calorie Mama platform for object detection. Objects are then verified for labels and weights by the user. The app references the United States Department of Agriculture DB, extracting nutritional content of each item and then can aggregate the nutrition intake per meal. The nutrition data is sent to HealthKit on the user's mobile phone as a central data aggregator. Metadata of timestamps and location are also captured.

Category	Source	Data-Stream	Units
Utility	Stages	Power	Watts
Biology	Stages	Energy Balance	Wbal
Utility	Stages	Power Delta	Watts/s
Biology	Garmin	Heart Rate	bpm
Biology	Garmin	Temperature	Celsius
Environment	Garmin	Altitude	meters
Environment	Garmin	Slope	percent
Environment	Garmin	Temperature	Celsius
Utility	Stages	IsoPower	Watts
Biology	Stages + Garmin	Aerobic TSS	TSS
Biology	Stages + Garmin	Anaerobic TSS	TSS
Utility	Stages	xPower	Watts
Utility	Stages	aPower	Watts
Biology	Garmin	Heart Rate Delta	bpm
Utility	Garmin	Speed	kph
Utility	Garmin	Acceleration	m/s <sup>2</sup>
Biology	Garmin	Cadence	rpm
Biology	Garmin	Cadence Delta	rpm/s

Table 5.4: **Sensor Streams.** Sensor streams captured at 1 second intervals during exercise activity.

## Neurological

Understanding the mental perception of each individual’s state is a subjective component of a health state. Neurological events include stress, mood, and even one’s screen time. *Stress*, in the medical context, is a physical, mental, or emotional factor that causes psychological or physical strain to an individual. Stress also includes an individual’s response to a change. One’s *mood* is a temporary state of feeling that the individual experiences, reflecting their current state of mind. Each digital interaction also gives insight into neurological events. The *screenome* consists of all interactions with one’s screen. The screenome reflects the use of digital media and screen time to give insight into how individuals are choosing to spend their time [228]. Although viewed as subjective, there are clear physical impacts on human physiology, as described in Chapter 2. Therefore, the mental state of the individual is in both subjective and objective formats. Stress and mood are captured subjectively through a

direct query to the user, as given in Table 5.3. Stress is captured objectively with heart rate variability (HRV) through the Garmin Fenix 5 smartwatch, as HRV is a strong indicator of stress. Lastly, the screenome is collected through background tracking software RescueTime and Time Out, also shown in Table 5.3.

## **Sleep**

This portion of data collection includes time spent in the prone position. Prone position activities refer to the body in a horizontal position. The individual may be awake or asleep while tracking this activity. Sleep is a condition in which the nervous system is relatively inactive, eyes are closed, and postural muscles relaxed. During sleep, consciousness is typically suspended. Sleep data is collected through objective and subjective sensors. The objective sensors include ambient noise, accelerometer-based movement, heart rate, HRV, temperature, humidity, and timestamps. Subjective measures include mood before sleep and perceived quality of sleep upon waking up.

## **Environment**

The external environment directly impacts various components of health, as described in Chapter 2. The external environment includes all factors outside the body that affect the individual; these can be local factors in one's home like lighting or broader environmental exposures such as the air. We use both public GIS data and devices that are precisely placed within the immediate vicinity of the user to collect environmental information. GIS data is collected through public sensors near the user's time and location. Public data about the user covers a large variety of data types, as shown in Figure 5.11. Hyperlocal systems gather data about the user's in-room environment. These streams come from the IoT devices and

the Raspberry Pi. Table 5.5 lists the sensors we use to build the exposome stream for the individual.

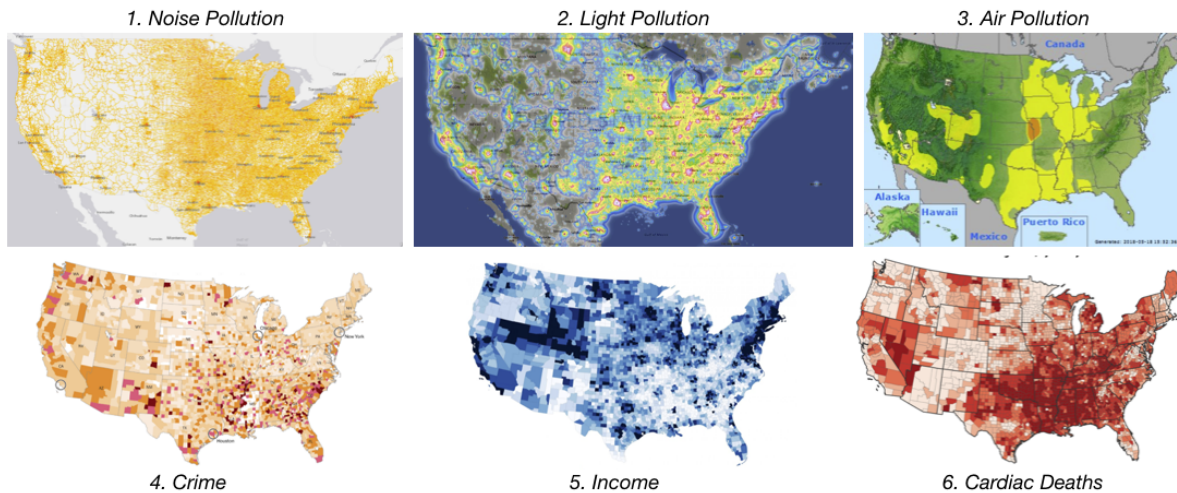


Figure 5.11: **Public Geospatial Dataset Integration.** Many aspects of the environment, including ones that induce stress, can be elucidated from domain literature and publicly available GIS datasets. Related to pollution, we gather air, audio, and light as factors that impact the health state. Related to demographic information, we gather crime, income, and cardiac deaths to estimate socioeconomic stress [193].

We use this detailed collection of data for the event and stream DB system to power the HSE.

### 5.3.3 Biology and Utility Direct Measurements

Data capture about the biology of an individual is done concurrently with many of the events, as shown above. Data from smart devices and manual tools are shown in Figure 5.4 and Figure 5.5. The data capture of utility in the case of this experiment is the power output capability of the individual, also captured concurrently with activities. One additional method of biological data capture is with images of the user. Jaw shape and facial proportions are associated with testosterone levels. Here we use social media images of the individual to estimate the facial proportions and shape and derive a relative score of testosterone

Type	Source	Data-Stream
GIS Matched	EPA	PM2.5
	EPA	PM10
	EPA	NO
	EPA	NO2
	EPA	NO3
	EPA	Ozone
	EPA	UV
	Open Street Maps	Altitude
	In-the-Sky	Light
Hyper Local	Sensor Push	Temperature
		Humidity
	Decibel-X	Sound Level (dB)
	Sleep Cycle	Sound Level (dB)
	Nest Protect	Smoke
	Nest Thermostat	Temperature
		Humidity
	Raspberry Pi	Temperature
		Pressure
		Altitude
		Humidity
		Carbon Monoxide (CO)
		Smoke
		PM10
		PM2.5
Large Particulate (LPG)		
Volatile Compounds (tVOC)		
Ethanol / Fuel		
Carbon Dioxide (CO2)		
Ozone (O3)		
Dust		

Table 5.5: **Exposome Sensors.** There are two forms of exposome data collection within this experiment. The first is GIS matched data, which uses the user’s location to retrieve the nearest public sensors as an estimate of their local environment. The second set of data uses hyper-local in-room sensors to capture data about the immediate vicinity of the user’s environment.

for the individual [193]. Testosterone levels affect the cardiovascular health of individuals and the ability to recover from fatigue [20, 247, 14, 35]. From a user’s full social media data, images were filtered that had a high-quality resolution and direct frontal view of the subject. Open Computer Vision resources were used to pinpoint the face height-to-width

ratio. An example of a photo analysis is shown in Figure 5.12. The theme of this biological measurement technique from a computing perspective is that we can garner a plethora of information regarding the subject’s health if we use modern image recognition techniques. We also investigated these images for skin health through pigmentation tracking and wrinkle detection. This investigation is shown here as part of the image-based health tracking theme, but it will not be used in further CRF estimations. The appendix contains resources used in data collection, analysis, and visualization in Table B.2.

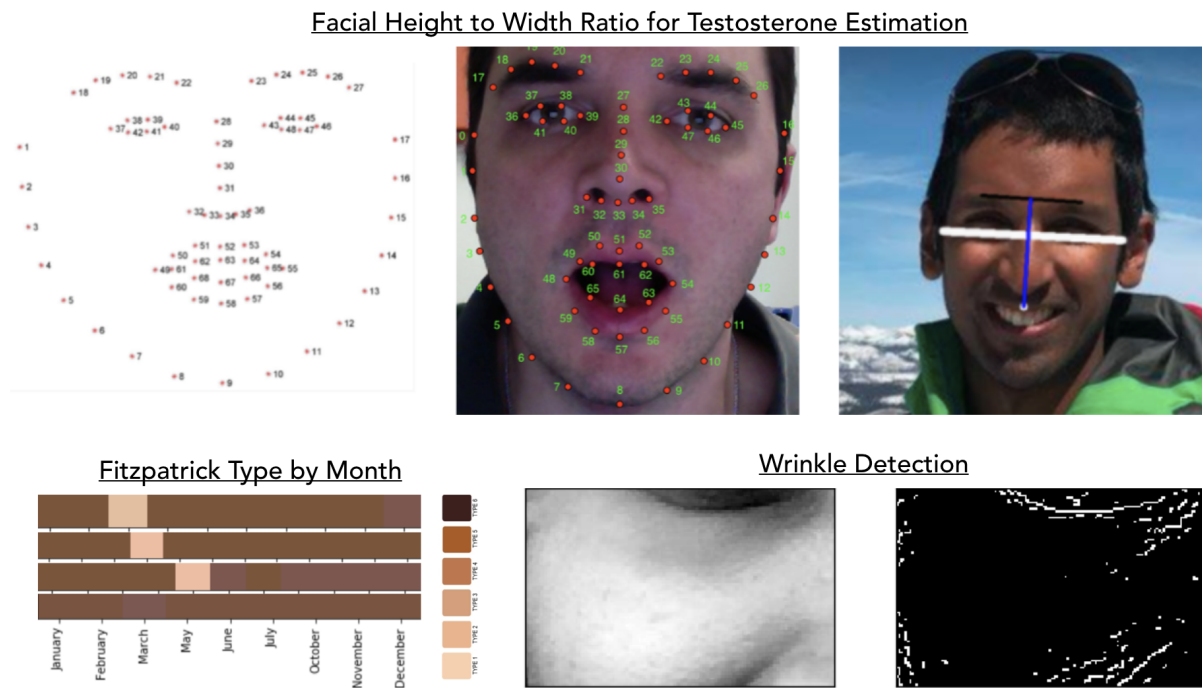
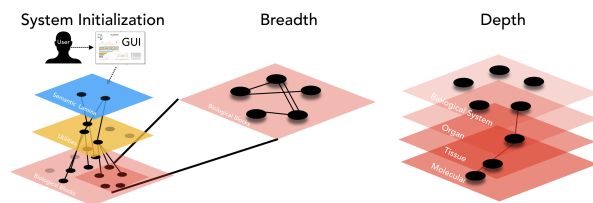


Figure 5.12: **Computer Vision for Biological Measurements.** Images can give insight into a person’s health state. In this figure, we use facial images for three purposes. First, we detect landmarks on the person’s face to identify their height-to-width ratio, lending insight into their baseline testosterone production. Second, we use color detection to see how the user’s skin color is changing based on each month. Skin color is classified using the Fitzpatrick type used routinely by dermatology physicians [232, 236]. We see on average that spring time, the user has lighter skin than the rest of the year. Third, we use edge detectors to characterize the wrinkling of the skin to understand collagen health.

## 5.4 HSE System Initialization



The precursors to the state estimation system require having the above data streams and background work. Next, we begin the process of implementing the system for the test subject. The first sequence of steps is initializing the system, as explained in Chapter 4. We run through this process to set up the GNB structure, created once the user specifies a domain or intent. Following this, we can populate the relevant layers of utility dimensions and biological layers. This section walks through all these steps for the test subject.

### 5.4.1 Translating User Intent



To start the HSE process for the subject, we begin by inquiring the user for their intent or domain of interest. The test subject's chosen interest in the experiment is in the realm of cycling. We display a lamina of cycling types to the user for visualization and interaction, as shown in Figure 5.13. The user can see the relationships of different types of cycling on a particular health state space. The proximity of the various types of cycling shows the similarity of the health state required to participate. For example, BMX, track, and downhill styles of riding are all quite comparable on the upper left-hand side of Figure 5.13. Similarly, cross country, road, and gravel are also close together. Enduro riding sits in between these two clusters. The subject selects interest in the disciplines of enduro and gravel riding and wishes to see the health state in the context of this region-of-interest in the lamina. The next step for the system is to generate the utility dimensions for this lamina.

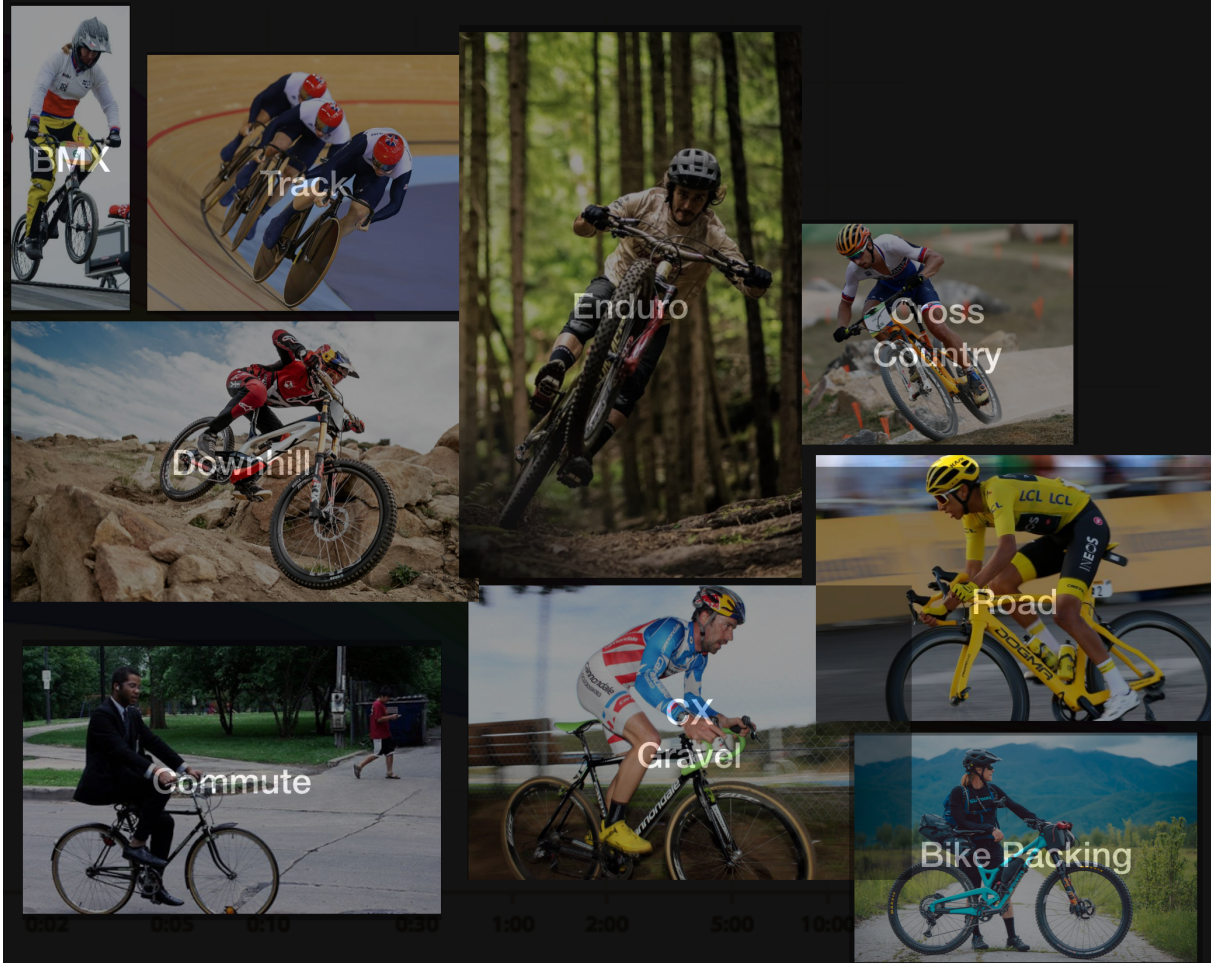
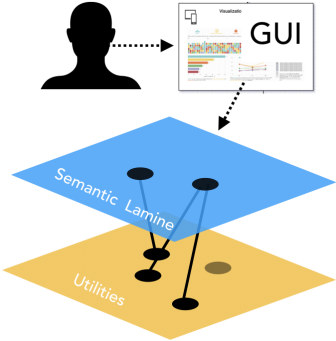


Figure 5.13: **Cycling World Lamina.** This interface displays many different types of cycling of interest to the user. The user can select a specific type, zoom in and out, or view the whole lamina at once.

### 5.4.2 Retrieval of Relevant Utility State Space



Given the intent of cycling by the user, the knowledge DB retrieves the dimensions of endurance, specifically the Critical Power (CP) curve, to display this specific lamina. The CP is the amount of power that can be applied maximally in a given period. It serves as the tolerance or biological threshold for the whole body in exercise (such



as in cycling) for an individual [139, 224]. For example, a 5-minute CP is the maximum amount of power an individual can produce for 5 minutes. An untrained person can generally produce about 2 Watts per kilogram of body weight. For a 70kg person, this would be 140 Watts of power. A world-class athlete can produce around 7 Watts/kg, sustaining 490 Watts of power for a 70kg person. CP is usually shown as maximum power for each one-second interval of time from 1 second to several hours. Every living individual has a CP utility dimension value for producing power for 1 second, 2 seconds, 3 seconds...60 seconds (1 minute)...3600 seconds (1 hour). For measurement of these dimensions up to 5 hours, there are 18,000 utility state dimensions (5 hours x 60 minutes x 60 seconds). There are many other laminae related to cycling using utility dimensions of skill, sensing, or thinking abilities. We do not cover these in the dissertation for the sake of brevity. This selection of CP as part of the dimensions within the GNB of movement is shown in the utility state space in Figure 5.14.

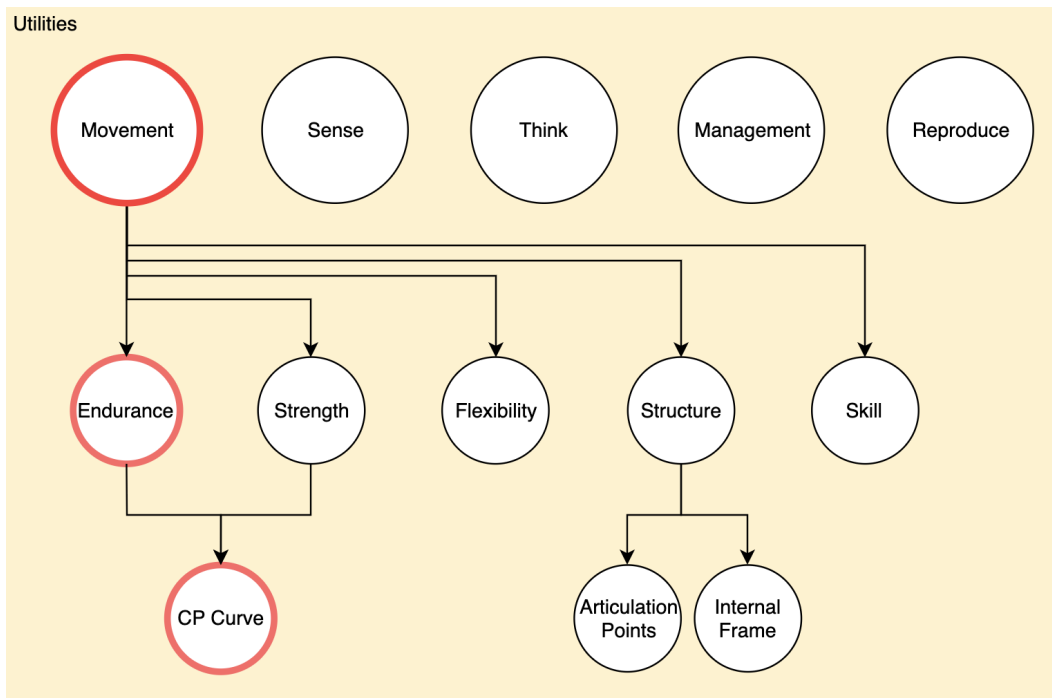


Figure 5.14: **Utilities of Human Health.** In this work, we have defined five basic clusters of human utility. Given the selection of the user, we focus on the utility blocks that are connected to power production. In this work, we focus on the endurance capacity of critical power production. These nodes are highlighted in red.

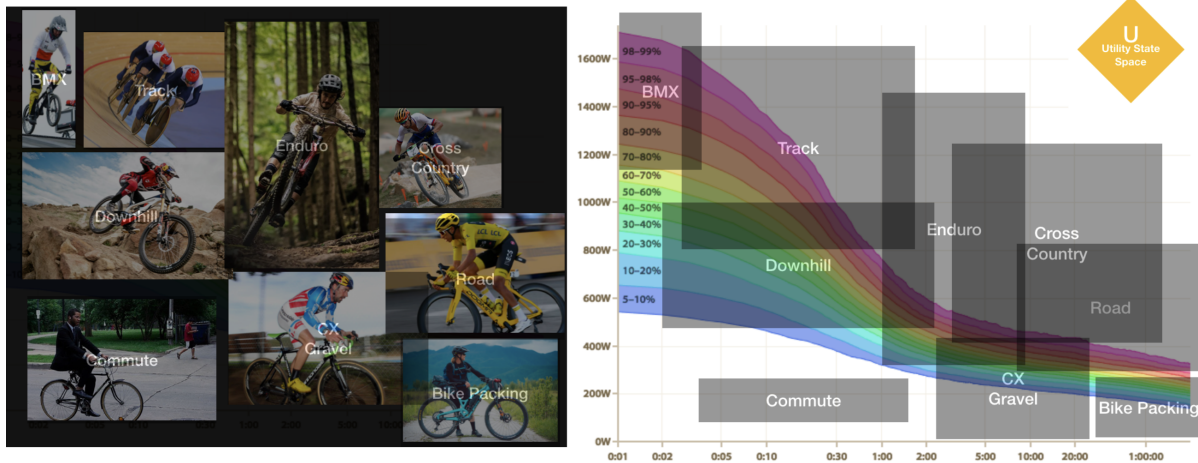
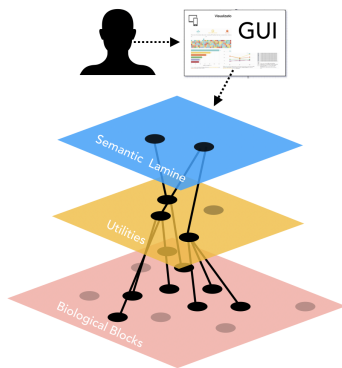


Figure 5.15: **Translating Regions-of-Interest to Utility State Space.** On the left, we see all different types of cycling that a user may be interested in participating in a cycling lamina. On the right, these types of cycling are overlaid on the utility dimensions of power production versus duration. This state space of power given duration is the CP curve.

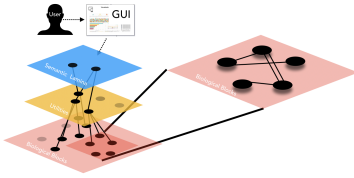
### 5.4.3 Populating Biological Blocks



Once we have retrieved the relevant utility dimensions, such as CP, we need to retrieve next what nodes in the biological block determine the value of these dimensions. There are two challenges to implement this. First, there is a breadth of biology that is needed to explain a particular utility dimension. For example, if the utility is walking ability for movement, it is insufficient only to understand the isolated biology of the toes or knees. The

whole biomechanics and bioenergetics of walking must be populated into the system for a proper understanding of walking. A superficial understanding is a good start, but understanding the depth of details about the system will give a better picture of the factors that are changing the utility state. The next two sections cover these two challenges of breadth and depth in the biological GNBs.

## Breadth of Biology



To capture the parent nodes that are directly related to the utility dimensions of CP, we reference the pertinent domain knowledge of human bioenergetics [33]. Bioenergetics focuses on how cells transform energy in cells, tissues, and organisms, which may be done by storing, transferring or consuming energy [197]. Depending on the duration of the energy need, there are four significant systems the human body uses to move. These four systems of interest are in Figure 5.16 overlaid on top of the utility dimensions. The dimensional boundaries of enduro and gravel riding (the user's selected region-of-interest), the utility dimensions to focus on are from the 100 second CP value to the 1,200 second CP value (20 minutes). The biological energy system connected to these utility dimensions is aerobic glycolysis, as shown in Figure 5.16. Aerobic glycolysis ability is synchronous with CRF and measured in units of  $VO_{2max}$  (mL O<sub>2</sub> / minute / kg).

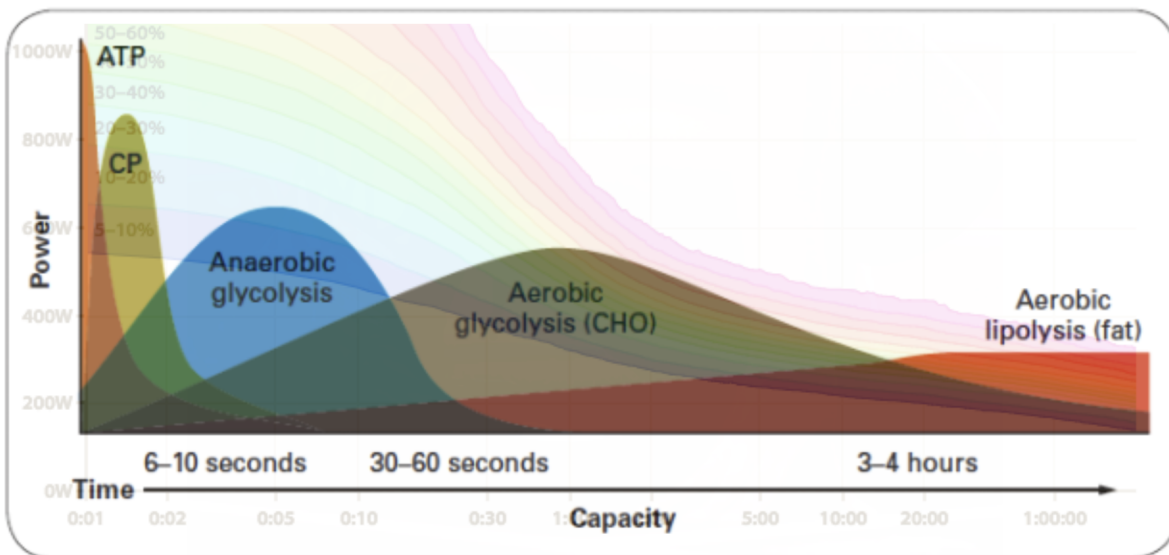


Figure 5.16: **Populating Biological Blocks from Utility Blocks.** The CP utility state space (background) exists through 4 main biological systems (foreground). The first is the ATP-CP system (orange and yellow), the second is the anaerobic glycolysis (blue), the third is aerobic glycolysis (grey), and the fourth is lipolysis (red).

Given this target system, we populate the range of involved biological nodes at the gross organ level. The organ systems involved in this process and the edges between them are given in Figure 5.17. Each organ has a set of major global attributes describing its microservice output to the system. The microservice edge relationships are captured in the equations in Figure 5.17. The global attributes of these organ nodes are in Table 5.6. Next, we dive into deeper layers of biology by expanding these organ level nodes.

Organ Node	Global Attribute
Lungs	O2 diffusion, CO2 diffusion
Blood	O2 carrying capacity, heat capacity (through fluid volume)
Heart	Cardiac Output
Vasculature	O2 diffusion, CO2 diffusion, density
Muscles	O2 extraction, rate of ATP synthesis, myofibril force

Table 5.6: **Organ Global Attributes.** Global attributes of major organ system GNBs

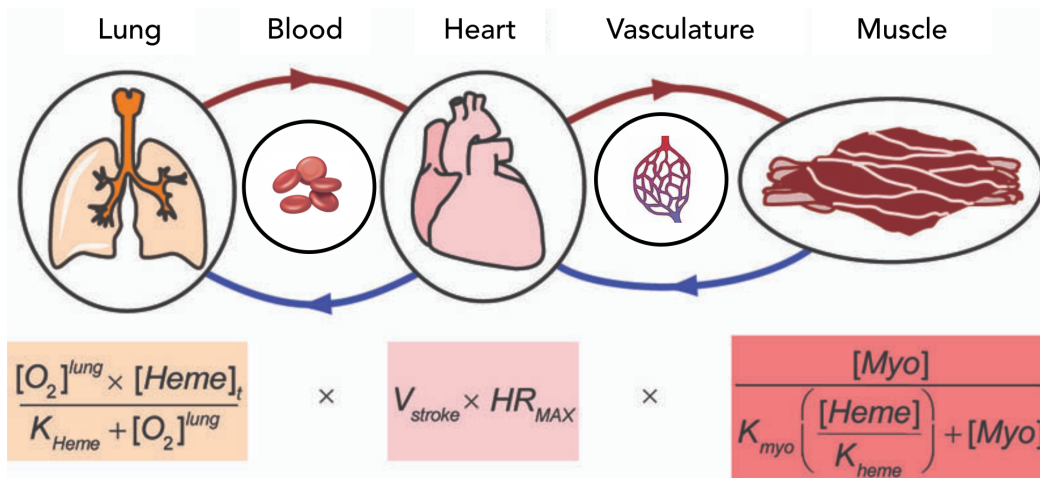
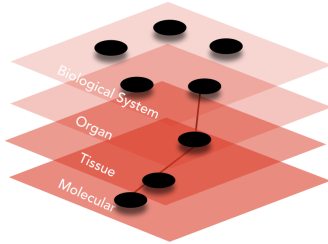


Figure 5.17: **Aerobic Glycolysis Domain Knowledge.** This figure shows the flow of oxygen (O<sub>2</sub>) from the air to the muscles used in glucose oxidation, also illustrating how the nodes are connected in the GNB. The processes are well-characterized through the scientific literature, and in some instances, can be described mathematically. These processes are captured in the edges of the GNB.

## Depth of Biology



To face the challenge of biological depth, we build the GNB with deeper layers of biology. We will address four layers of depth in the experimental model. The first and second layers are the system layer and organ layer, as described in the broad sense above. We then dive into the tissue level and the molecular cell biology level in further depth. For

the cardiac (heart) system, we see there are several key tissues that comprise the heart. The left ventricle is the tissue chamber that is responsible for cardiac output, which is the key attribute of interest for the utility dimensions at hand. This nesting of microservices in the GNB for biological depth is shown in Figure 5.18.

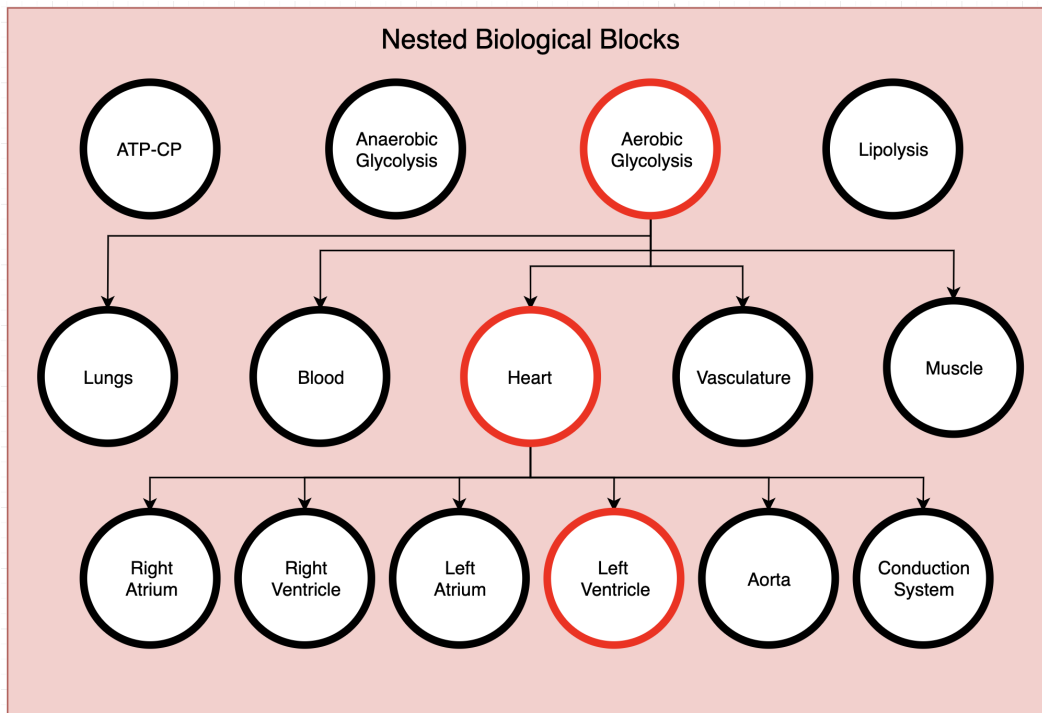


Figure 5.18: **Nested Graph Network Blocks within Aerobic Glycolysis.** There are many biological components nested within aerobic glycolysis. We focus on the details of gene expression inside the left ventricle in this experiment set. These nodes are highlighted in red.

Within the left ventricle, there is a highly complex network of molecules that signal changes in the tissue. These changes include growth or degradation of tissue and cellular proteins through changing gene expression. Events are the primary input that alters gene expression. The understanding of these networks is usually stored in natural language within the body of traditional literature. Recently, graph DB systems that store domain knowledge in cellular signaling have also started flourishing. Cytoscape is the most popular of this graph-based knowledge and research systems [212]. Harnessing the knowledge in these systems, along with traditional literature, helps to populate the molecular networks for the tissue of interest. Figure 5.19 describes the process by which mitochondrial biogenesis networks for this experimental GNB is populated from public domain knowledge. Domain knowledge of how events impact this network are also curated from traditional literature to build this knowledge graph. The process of using public graph networks of biology and traditional literature is also shown for cardiac tissue in Figure 5.20. The biological network for cardiac tissue is sourced from Cytoscape, and the event relationships to these nodes are sourced from traditional literature.

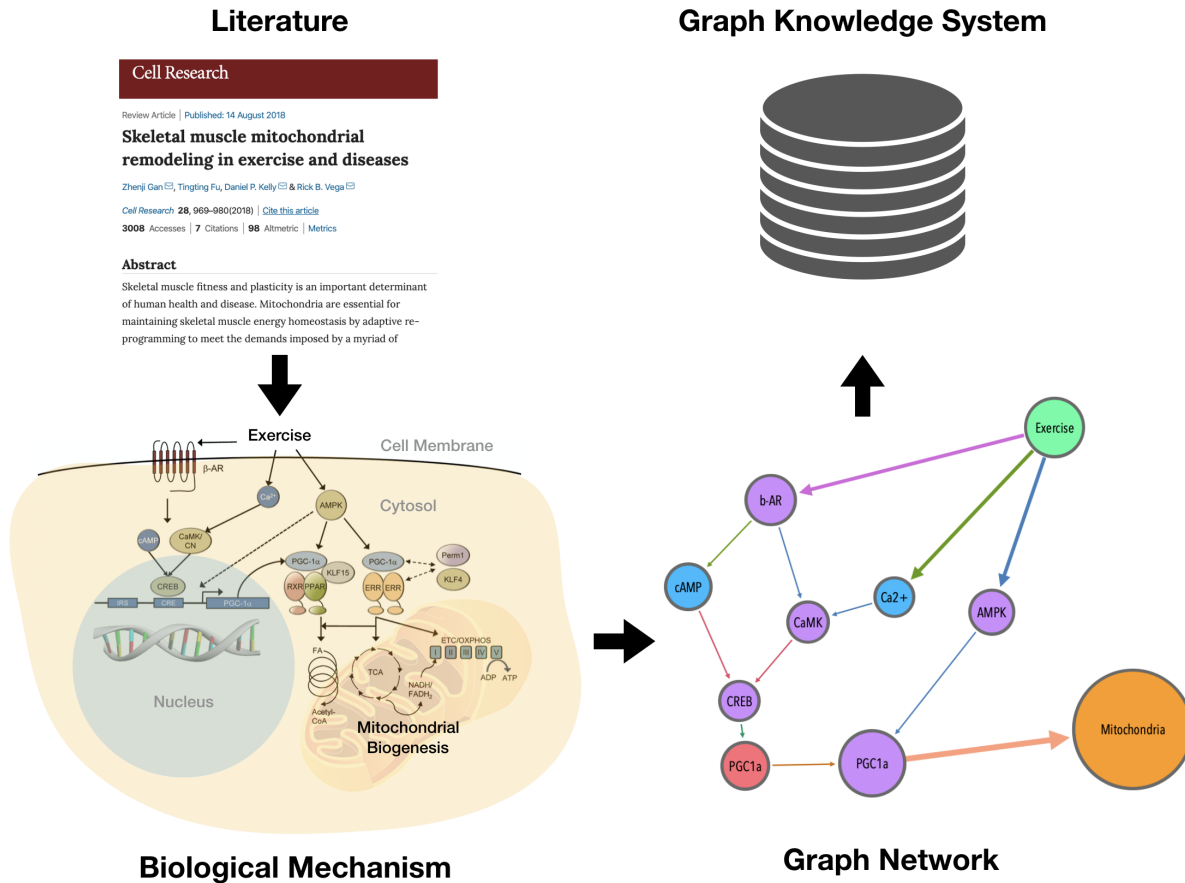


Figure 5.19: **Translating Domain Knowledge to Graphs.** Mitochondrial biogenesis pathways are taken from literature to populate the molecular GNB layers of the HSE system for our test subject. This GNB is stored as a node within the muscle tissue nodes.

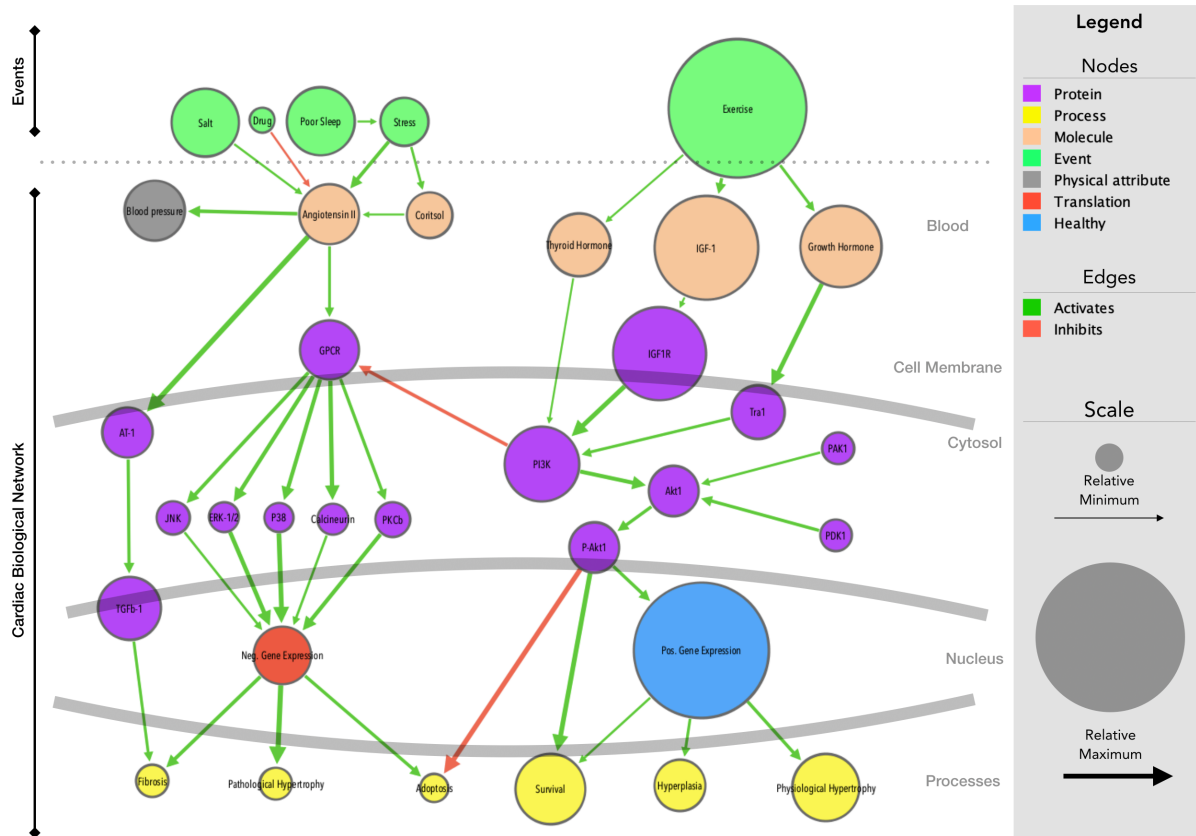


Figure 5.20: **Cardiovascular Graph Network Block.** This network shows how events impact the cardiovascular state down to the gene expression level. The graph network of biology is taken from compiling open-source data from graph knowledge database Cytoscape and other literature. Consolidation of this into a unified graph was done by the author. Green arrows denote an activating effect on the target node. Red arrows denote an inhibitory effect of the source node to the target node.

These knowledge graphs at the molecular level populate the deepest level of the personal GNB structure in this experiment. Now that we have the HSE GNB system built at all the desired levels of abstraction, we can plug incoming data streams to see how they are changing the state.



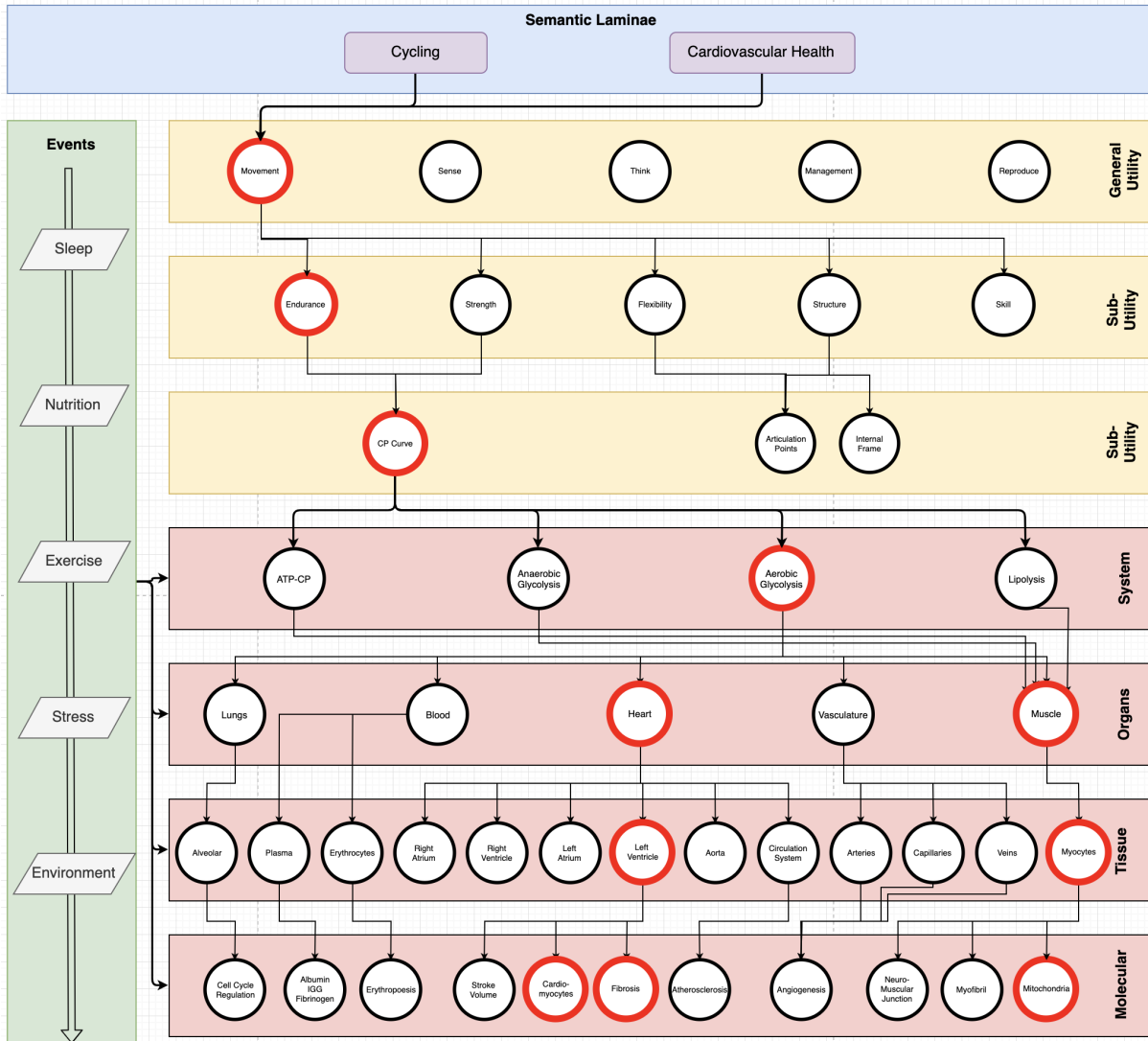
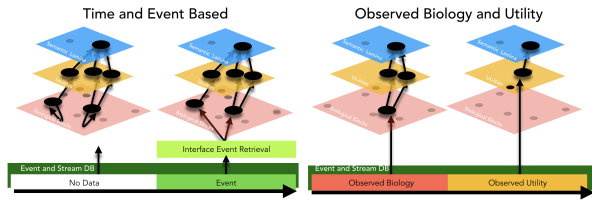


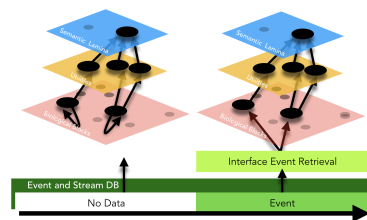
Figure 5.21: **Graph Network Block Structure Summary.** The total overview of the 4 main system components instantiated for the test user. Red circles denote areas of focus for the experiments carried out in the next chapter.

## 5.5 Health State Update



After setting up the GNB structure for the individual's intent, the next step is to use the network to show how the state is changing. Time intervals at which to update the state were chosen to be one day. There are two overarching approaches to update the state. The first approach uses event data streams about the individual as inputs. Events can include time intervals where there is no other data stream other than time itself. The second approach is measuring a biological or utility attribute directly through a sensor, and we use that information to arrive at an updated estimation of the health state. This section addresses these two approaches in separate sub-sections below.

### 5.5.1 Time and Event Based HSE Updates



The first update approach uses temporal streams as events that perturb the GNB. Temporal streams are understood as events in this work. Not all events impact the health state regarding the utility dimensions of interest. To detect events that are relevant in changing the health state we are focused on, we need to know what types of features to look for and characterize the impact of that event on the health state. This process is called *interface event retrieval*.

## Interface Event Retrieval

“In computing, an interface is a shared boundary across which two or more separate components of a computer system exchange information” [121]. In the context of our work in HSE, the two sides that exchange information are the events in our life with the state of our health. Interface events bridge this gap by describing how to transform a particular event attribute into an input value for nodes within the HSE GNB structure. For events used in this experiment, we use validated aggregation metrics that connect the events to how they impact the health state of CRF. For activities, we use three algorithms of Chronic Training Load (CTL) [80], Gravity Ordered Velocity Stress Score (GOVSS) [281], and High Power Activity (HPA). In a nutshell, CTL captures the long-term (42-day rolling average) of cardiac volume overload output [80]. GOVSS captures the work done against gravity and inertia to produce motion [281]. HPA captures high-intensity activity that stimulates muscle fiber growth and motor unit training adaptations. HPA interface events are any 5 second periods with over 10 watts per kg of power. The extraction of these features from activity data is shown for the two years of test subject data in Figure 5.22. Using the time-stamps from the event stream, we can also inspect the circadian pattern of when particular interface events occur during the day. We show this for both activity events and environmental events in Figure 5.23.

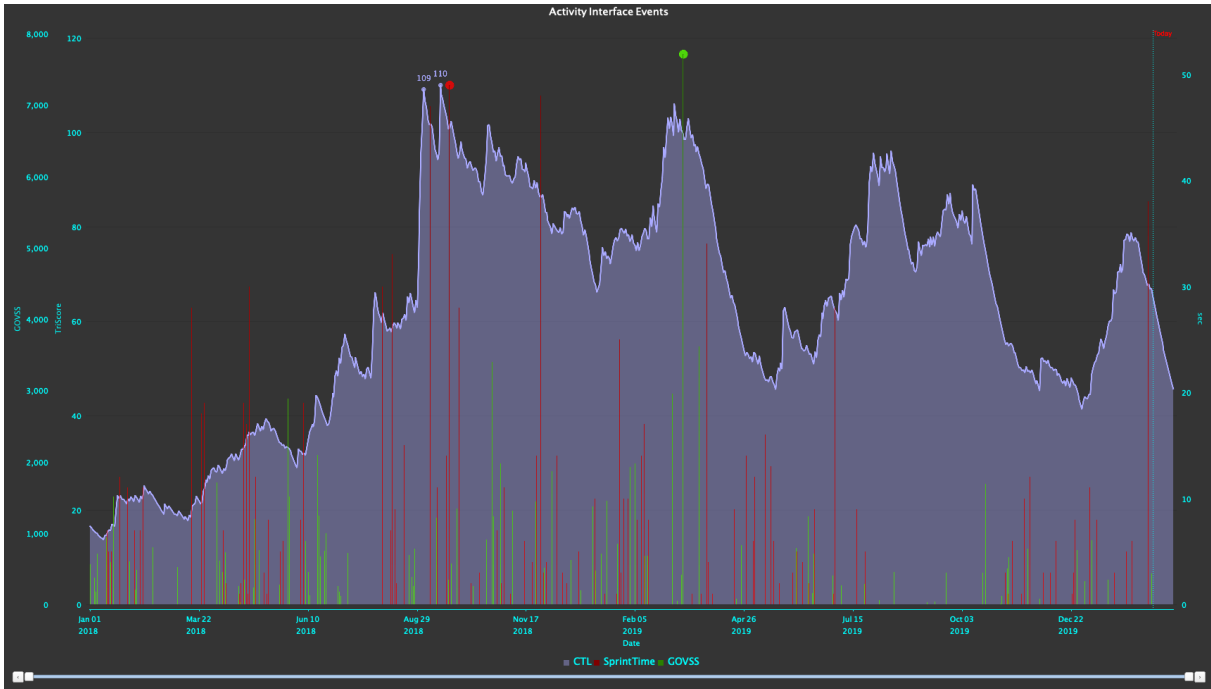


Figure 5.22: **Exercise Interface Events.** For the years of 2018-2019, CTL is given in purple, GOVSS is given in green, HPA Load by day is given in red. These specific interface events are used to determine impact to the CRF health state.

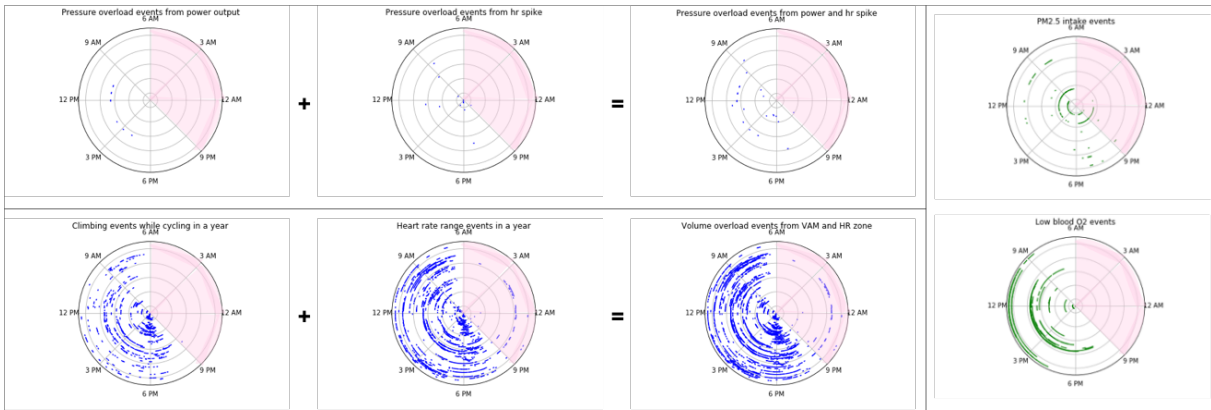


Figure 5.23: **Circadian Rhythm of Interface Events.** This figure shows different interface events retrieved from the test subject data. Different days of the year are represented as concentric circles, and various sectors of the circle represent different times of the day. Events in a day are represented as colored arcs on that circle. Volume and pressure overload events that occur in the left ventricle are retrieved as inputs towards changing the health state. This figure also shows the environmental interface events. High PM2.5 intake events are recognized over one week. These are the instances where per minute PM2.5 intake was greater than  $0.7 \mu\text{g}$ . Low blood oxygen events are recognized over one year, where blood oxygen saturation goes below 95%. The highlighted sectors roughly represent the time between sunrise and sunset. Thus we can see how different events overlap with circadian patterns.

In this experiment, we summarize the impact score of exercise (Figure 5.22), sleep as a total quality metric, food through relevant nutrients, stress through a combined score of objective and subjective data, and environmental pollution as quantity per day. The algorithms for these computations are given in the appendix. The CTL, sleep, and stress scores per day are presented in Figure 5.24, preparing the system to now propagate the effects of these events into the GNB.

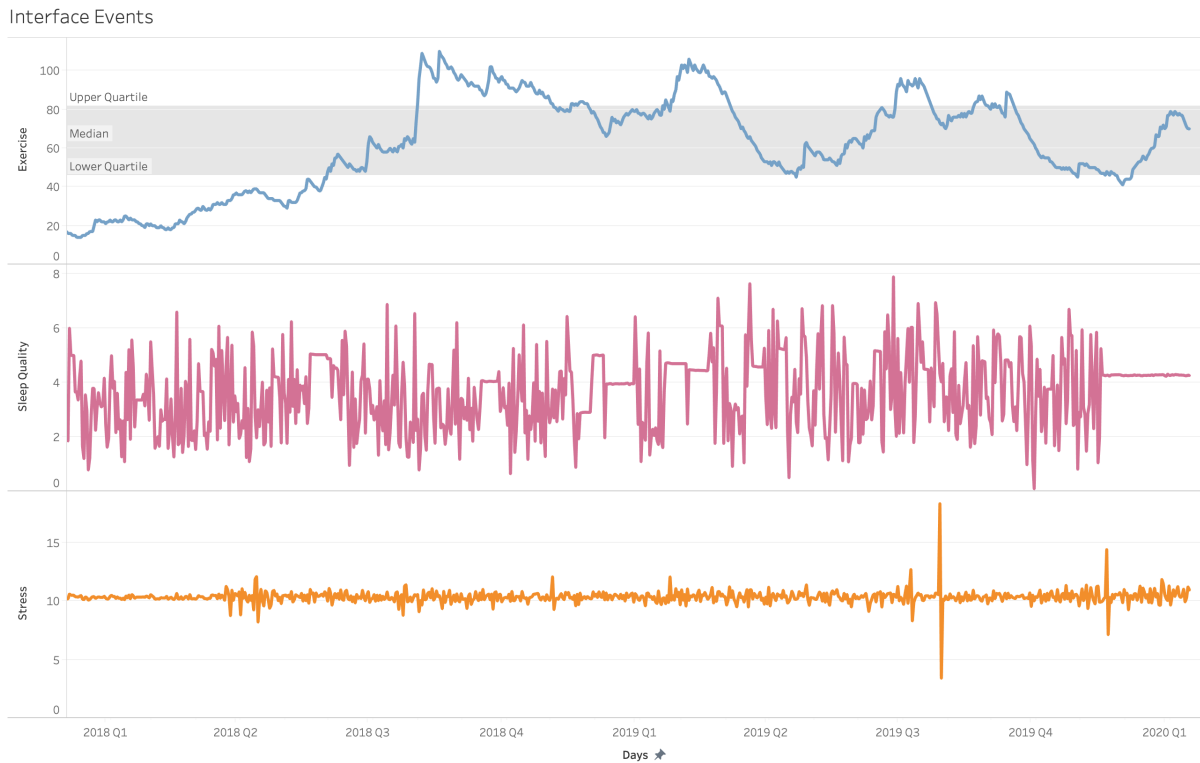
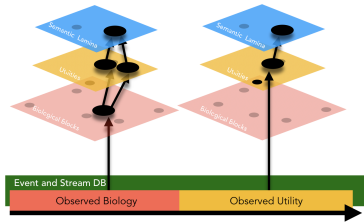


Figure 5.24: **Interface Event Summary.** Total Exercise impact, sleep quality impact and stress impact by day for years 2018-2019 in the test subject. Exercise events are shown in quartiles showing periods of higher impact in the upper quartile, and minimal impact in the lower quartile. High variance within sleep and stress occurs daily.

## 5.5.2 Observation Based HSE Updates



When we have instances of directly measuring the utility or biological values, we use these values in priority over the estimated values provided from event-based inputs. For the biological layers of the GNB, these values are provided at the organ level, or system level through the sensor collection. Examples of these values include body weight and heart rate. The node that is measured may have parent nodes that can be updated. In this experiment, we use resting heart rate and weight to derive stroke volume of the left ventricle, and thus cardiac output. For the utility states, we directly measure CP and weight during instances that indicate maximal power output. These measurements can be retrieved through looking at power intervals where heart rate is within ten beats of age-defined maximal heart rate.

## 5.5.3 Propagating Updates

Using the data streams of these events and direct observations, we feed new input values into the connected nodes as described in Algorithm 1. An example time-lapse of this update cycle for the CRF case is given in Figure 5.25. We use this update system to compute the health state of the individual on a daily cycle interval for the two years of data collection (2018-2019) on the test subject's GNB structure. The results in this section show updated values for all the layers in the GNB. Through the story of queries by the user, domain expert, or interfacing computing agent may have, we describe the health state update results in various levels of depth.

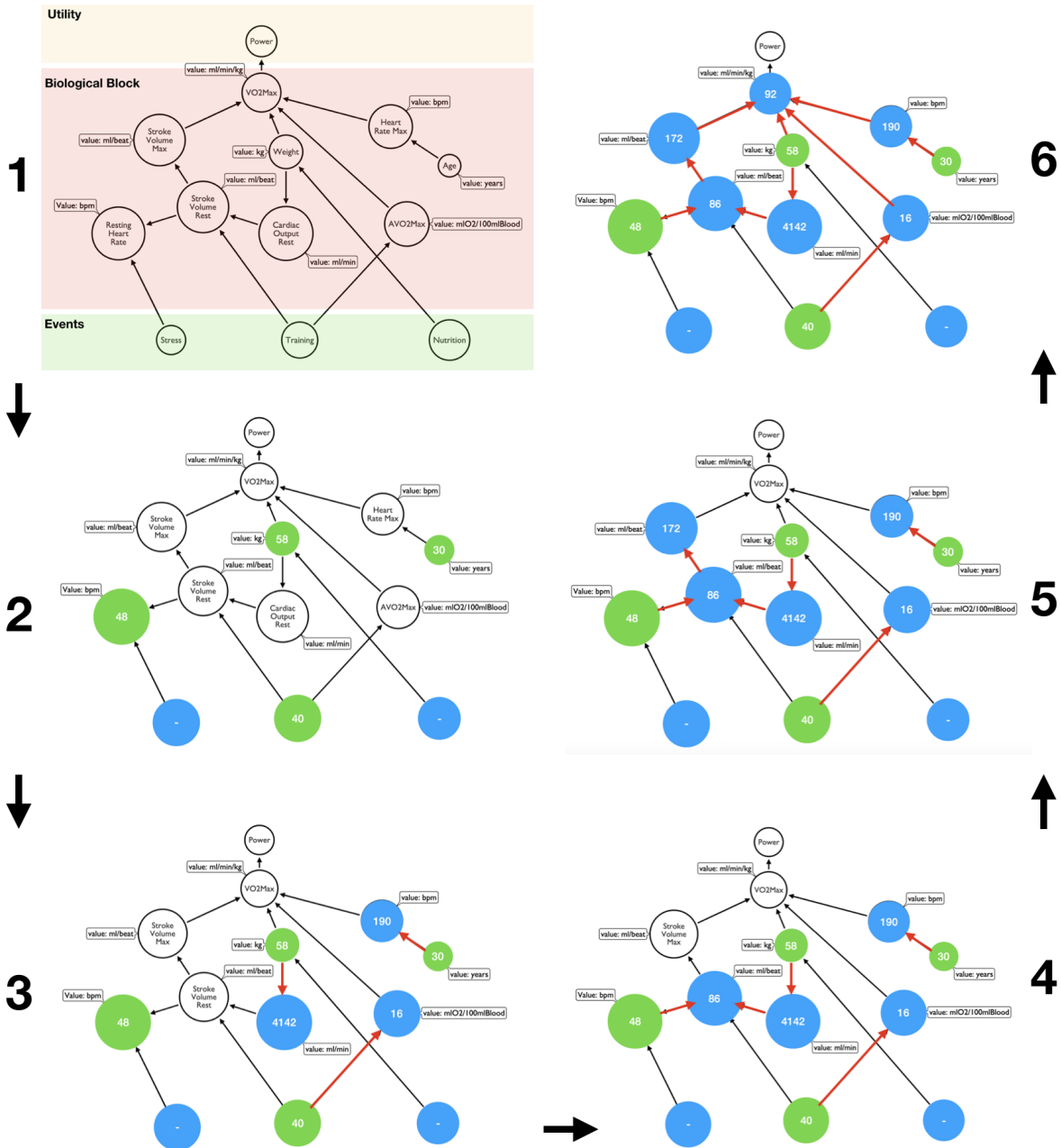


Figure 5.25: **GNB Update Cycle.** This illustration shows the time-lapse of steps within the update cycle for a GNB system. Step 1 colors the layers of the HSE GNB structure from the initialization. Step 2 shows how the value updates are given for inputs (green) consisting of an exercise CTL of 40, age of 30, weight of 58kg, and resting heart rate of 48. The estimated nodes are blue.

## Q1 User: How ready am I for the race?

The original intent of the user was to be ready for enduro and gravel cycling. Based on the user weight, the power to produce at minimum for these events is 2.5 watts per kg for sustained periods of over 1000 seconds, equating to 152 Watts of power for this particular CP dimension. If we can estimate the utility dimension of this particular value over time, we can see when the user is capable of participating in the races they desire to do. Figure 5.26 shows this particular utility dimension, with the green top section showing if the user is in the health state to fulfill their desire. We see the user has a total of 5 periods in the last two years where the user was within the appropriate utility state. Figure 5.27 shows the values of the entire 18,000 utility dimensions of CP (x-axis) at four different time points throughout the last two years. Figure 5.26 shows just a summary daily value of the HSE for one of these dimensions over time.

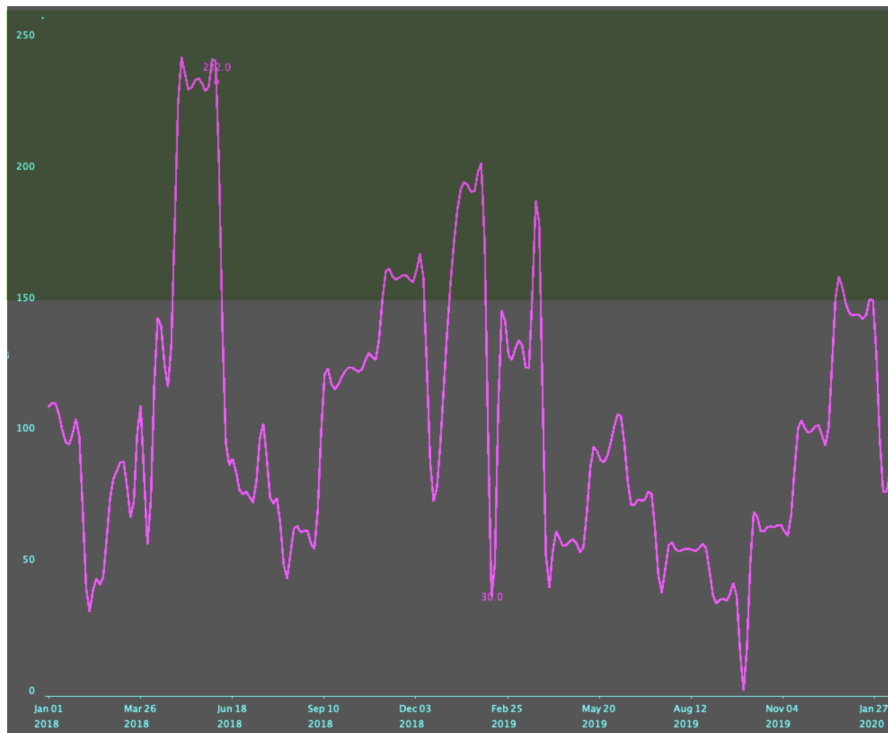


Figure 5.26: **Single-Dimension Health State.** Here is a single dimension of one hour CP value for the subject over the course of time.



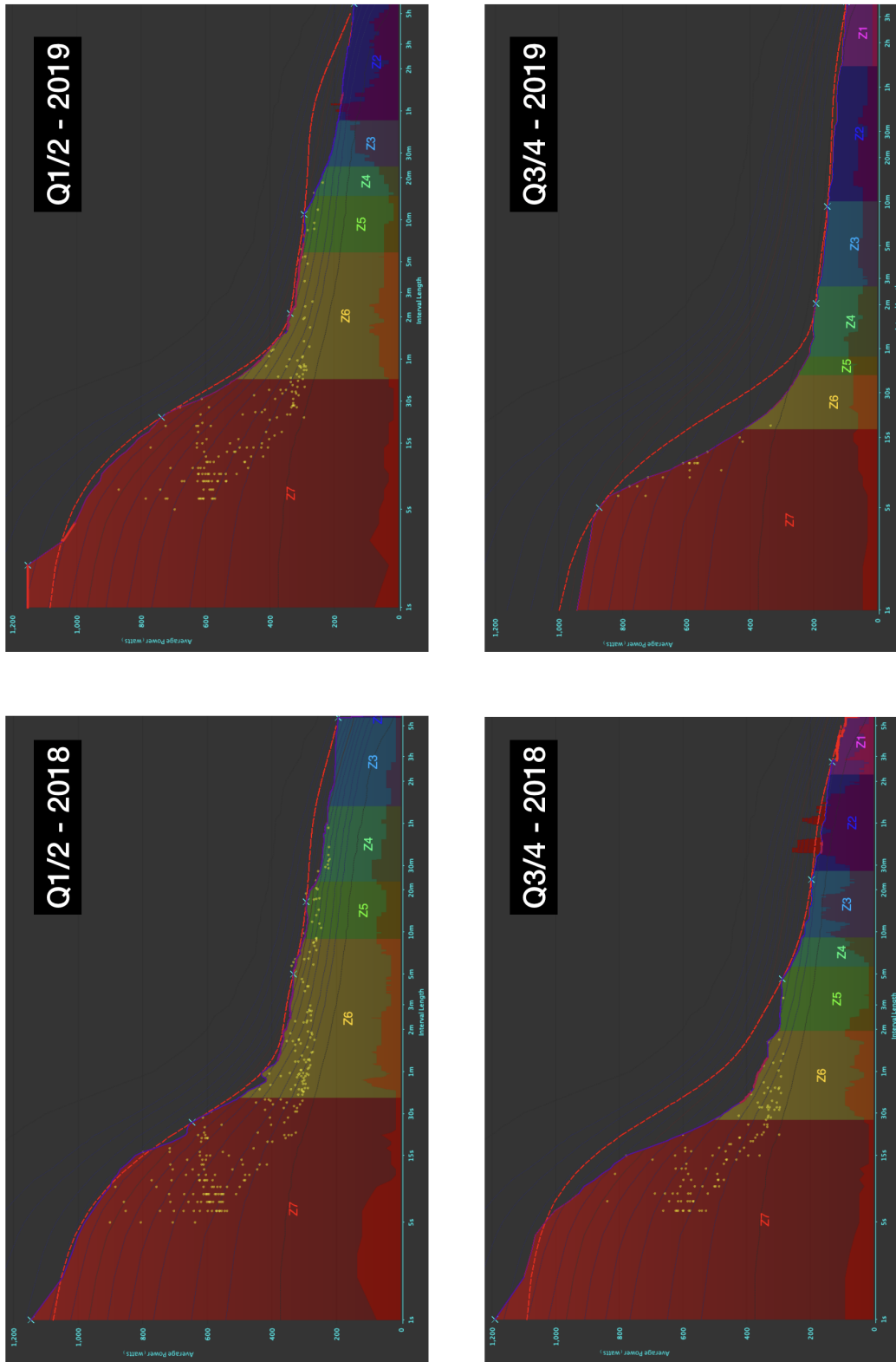


Figure 5.27: **Multi-Dimension Health State.** Here we see all 18,000 utilities of CP at six month intervals.

## **Q2 User: How does the air I breathe during a workout impact my health?**

Our test subject wants to know how the environment interacts with their health state. GIS visualizations can give readily accessible methods by which to answer this query. Figure 5.28 shows how pollution parameters like CO, O<sub>3</sub>, PM<sub>2.5</sub>, and PM<sub>10</sub> are monitored for understanding where respiratory and cardiovascular health is changing during outdoor exercise events. This visualization uses the location tracked activities synchronized to the public data sets generated personal exposome. For improved accuracy in providing this GIS estimation, we use heart rate data to understand breathing volume mapped with the local air quality sensors via constant GPS tracking [191]. Breathing rate is derived from heart rate streams generated during exercise [87]. Tidal volume is the volume of air inhaled during a normal breath. It is computed from gender, weight, and height of an individual, which are sourced from the respective GNB nodes. Estimation of the total volume of air intake would be the product of breathing rate and tidal volume. The total volume of air multiplied by the air particulate content gives total exposure to the particulate of interest. In summary, this query uses both the environmental data and the biological streams in conjunction to provide the result.

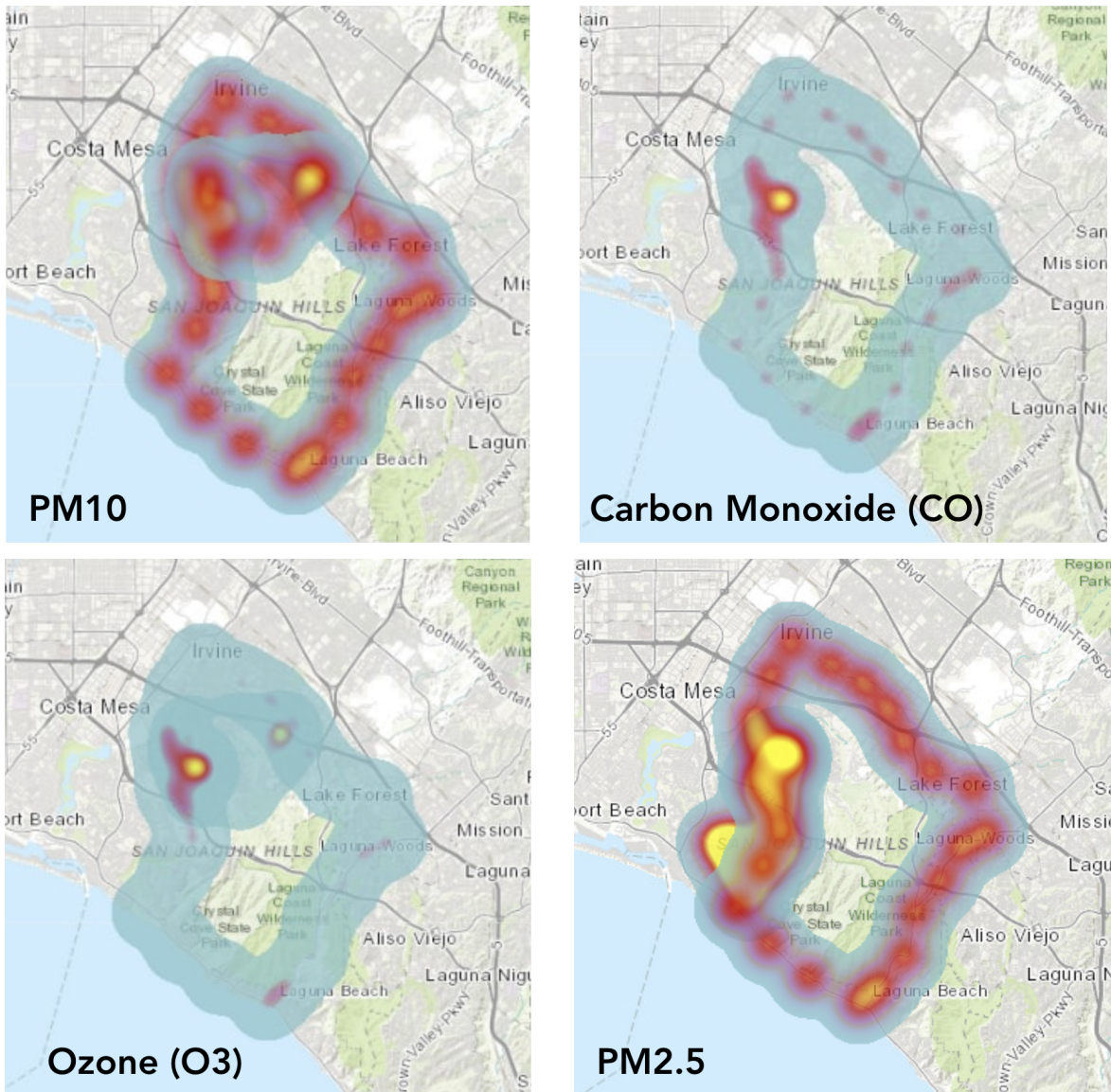


Figure 5.28: **Location Impact on Health State.** Here we can see how four different pollutants impact a single bike ride activity. Blue is low exposure, red is moderate exposure, and yellow signifies high exposure to the particular pollutant during the unique time-points and location during the bike ride. The EPA defines these levels. The levels can be used to inform the user, health experts, civil engineering, and recommendation systems of health state impact to aid in planning.

### **Q3 Doctor: How is my patient's heart health?**

Our test subject visits the doctor and wants to understand if they are healthy enough to continue riding bicycles and how their health is changing. For clinical use, organ and tissue level analysis is most common. A cardiologist specialist wants to know how the left ventricular wall thickness is changing based on various life events as shown in Figure 5.29. The physician can use the GNB to compute the summary of the heart node global attributes. This state is shown in Figure 5.30. The subject can see there are periods where heart health is improving, and heart health is hampered. The top row of the figure shows if healthy genes are expressed in the heart, and the bottom row shows the unhealthy genes. The second row shows how much positive physiological hypertrophy is occurring. This is when the heart's left ventricle expands in its interior volume, thus allowing for an increased stroke volume. The fourth row shows the pathological hypertrophy, where fibrosis and left ventricular wall thickness may be increasing, causing the left ventricle interior volume to decrease and leads to unhealthy states such as heart failure. The middle row shows the balance of healthy and pathological processes to see if the global attributes of the heart are improving or degrading. Based on these observations, the clinician can discuss best actions towards improving heart health.

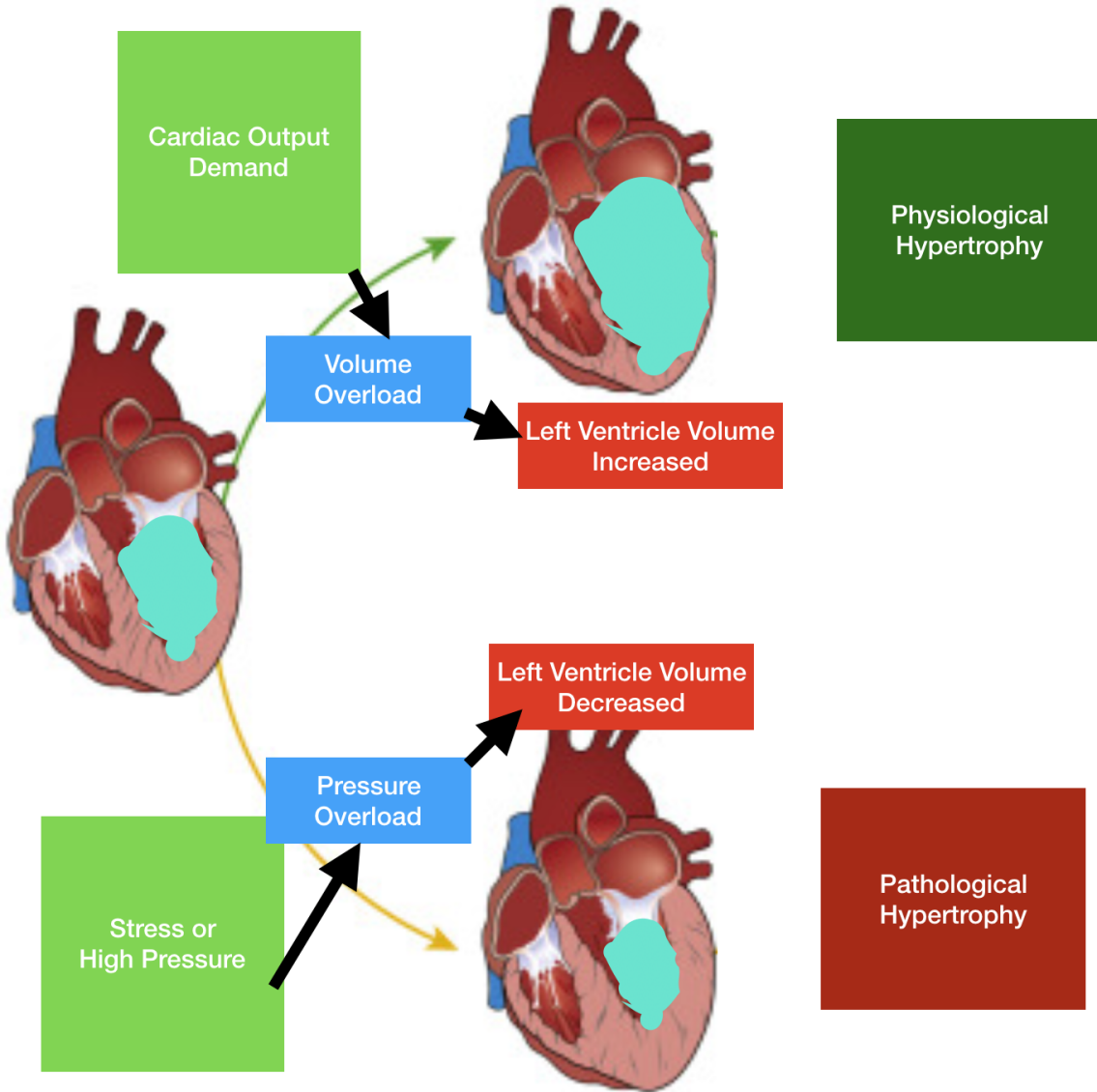


Figure 5.29: **Interface Events for Physiological and Pathological Changes.** This figure demonstrates the difference in response between physiological and pathological heart changes resulting from various types of events. Volume overload is when the heart needs to pump at a high volume flow rate. This causes physiological hypertrophy which strengthens the stroke volume of the heart. Pressure overload causes the heart to push against higher forces which increases wall thickness, therefore reducing stroke volume and reducing cardiovascular function.

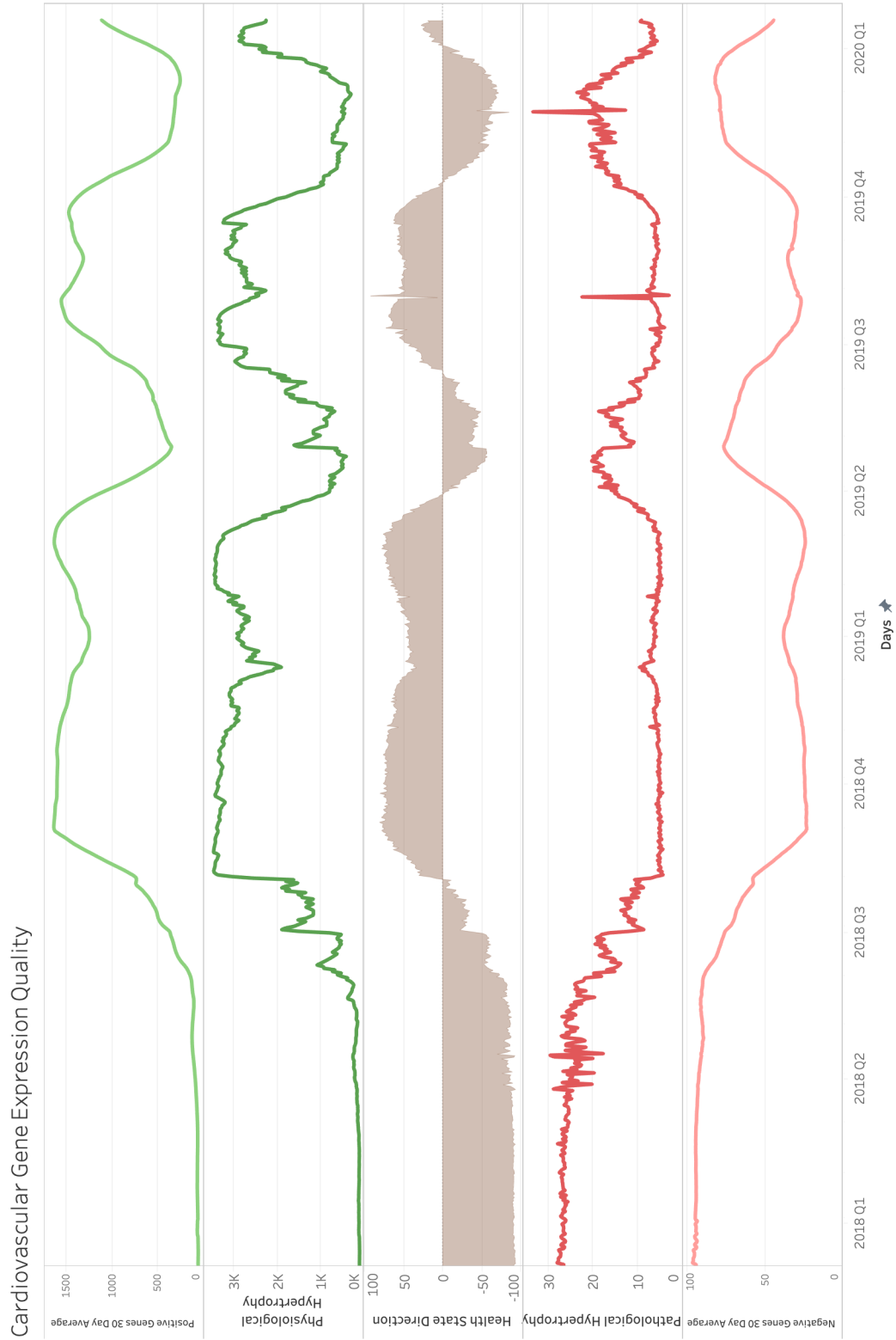


Figure 5.30: **Daily changes in Cardiovascular Biological Health.** This figure demonstrates the balance between healthy and pathological heart function on a day-by-day basis computed through the graph structure in Figure 5.32.

#### Q4 Doctor: How is my patient's left ventricular function?

Our test subject visits the doctor and wants to understand how their heart health is changing in regards to the left ventricular function. Again clinical use, organ and tissue level analysis is most common. The physician can use the GNB to see how the left ventricular volume is being updated on a daily basis. This state is shown in Figure 5.31. The subject can see there are periods where left ventricular health is improving or decreasing. A clinician can verify the stroke volume using ultrasound if they want a higher accuracy measurement.



Figure 5.31: **Daily Changes in Left Ventricular Stroke Volume.** This figure demonstrates the balance between healthy and pathological heart function on a day-by-day basis computed through the graph structure in Figure 5.32.

**Q5 Scientist: What are the molecular changes in this particular heart?**

The test subject participates in a research study where scientists are trying to understand what molecular pathways are potentially activated and deactivated in an individual. The scientist would like to know how cycling events are possibly changing molecular pathways that are related to heart health for potential further research and therapeutic targets. The scientist can zoom into the GNB level of the heart node molecular signaling and gene expression, as shown earlier in Figure 5.20. To see how the molecular signaling molecules are changing in the heart, a graph-based visualization is produced showing the pathway in each month of the year in 2019 (Figure 5.32).



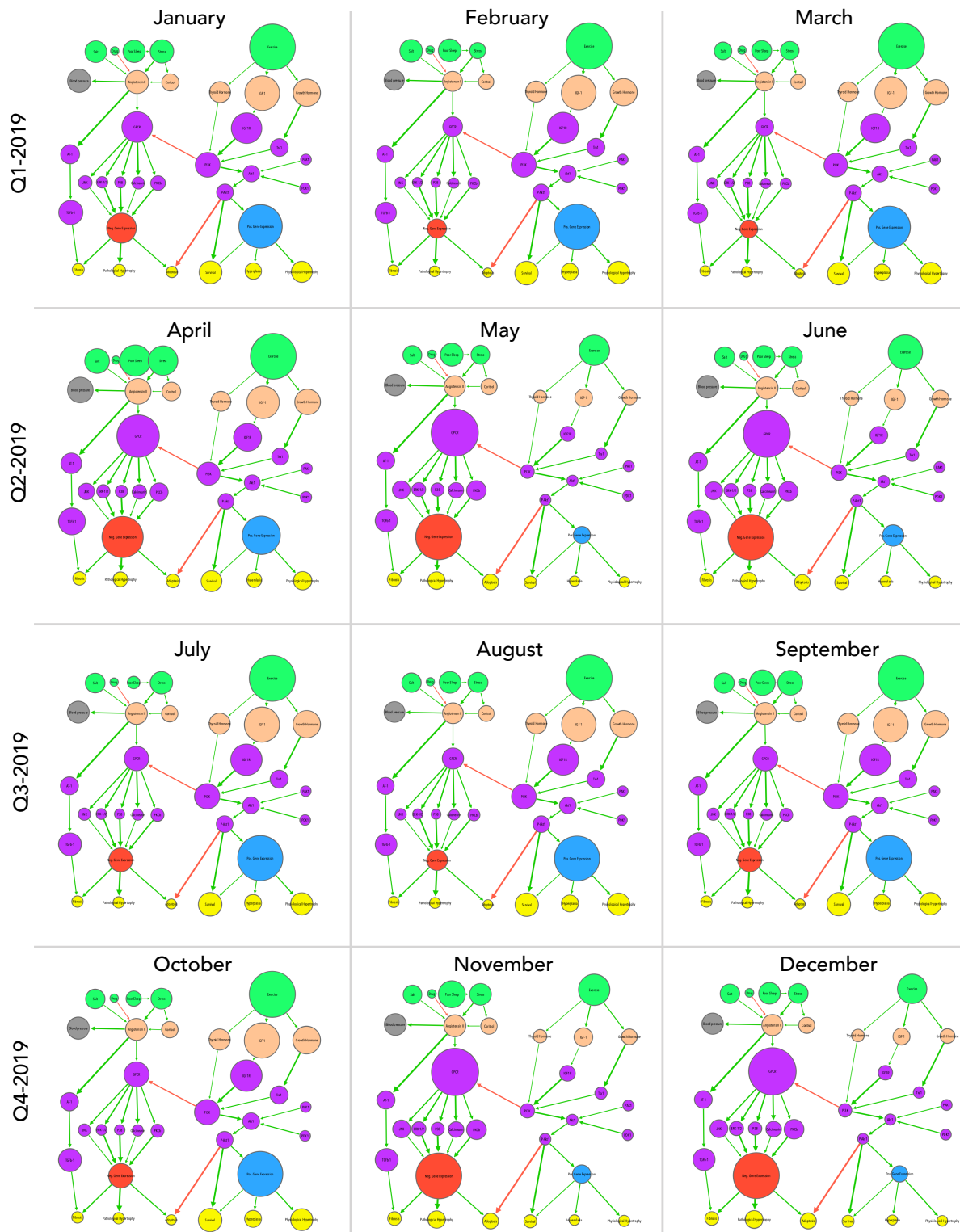
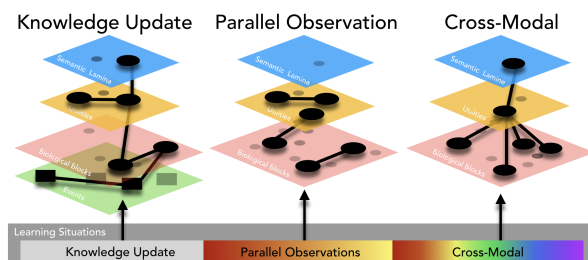


Figure 5.32: **Health State Graph Updates.** This figure demonstrates the estimated signaling pathway nodes for cardiovascular health change at the molecular gene expression level. An average for each month is shown in the year 2019. Legend is given in Figure 5.20.

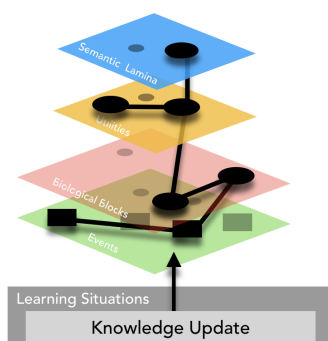
The various queries above demonstrate the usability of the health state estimation at different levels of abstraction. This outlines the framework by which any HSE system can be built. Concluding this section, we see how the health state updates can be used readily by the user, healthcare provider, and scientist to inquire how and why the health state is changing. Next, we discuss altering the HSE GNB structure to use incoming data to model the user better.

## 5.6 Personal Model Update



To update the GNB model for the individual, we take both a bottom-up and a top-down approach on the system. We start by using a top-down approach as a “Knowledge Update” on the model. Following this, we take two bottom-up approaches to better model the GNB. These data-driven approaches use two different methods of learning. One uses data incoming from two sources that are linked together to modify the edge. We call this parallel observation learning. The other uses many sources of incoming data to one node, to better estimate the values for that node. We call this approach, cross-modal learning.

### 5.6.1 Knowledge Update



Two forms of knowledge can inform the HSE system to update the GNB for the individual. The first form of knowledge is an update from the scientific literature on the structure and relationships between the entities. In this example, we update the relationship between an event category and biological nodes. Recently discovered knowledge in molecular biology shows how mitochondria increase in the muscles best from low-intensity exercise [311]. This finding is quite contrary to the commonly accepted research idea that high-intensity interval training (HIIT) was the most effective way to improve mitochondrial density in myocytes [123]. In light of this discovery, we increase the edge weights between the low-intensity exer-

cise interface events causing mitochondrial biogenesis and show the decreased edge weights for mitochondrial biogenesis from HIIT, as presented in Figure 5.33.

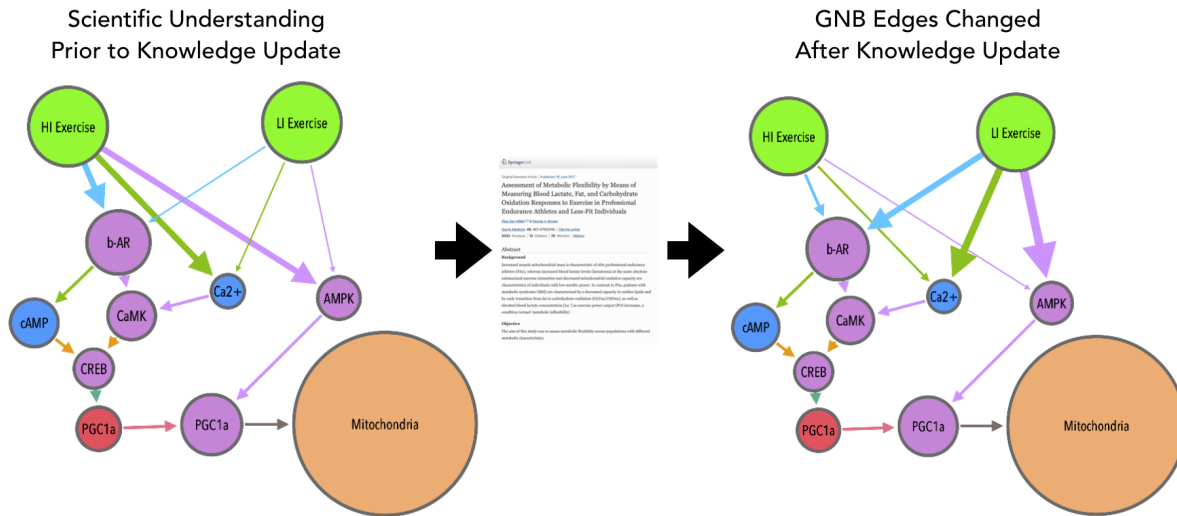


Figure 5.33: **Updating Model of Event to Biological Node.** This figure demonstrates the difference in the model between the original knowledge-based edge weights and after the knowledge update from the literature. High Intensity (HI) exercise was thought to be the primary causal pathway to increased mitochondrial biogenesis in the muscle and heart. New literature shows that Low Intensity (LI) exercise does this much more than HI exercise. Using this update, the edge weights on the right are changed.

The second approach to a knowledge update is the genetics of an individual. Most genetics studies establish the correlation or causation of a gene to either an event, another biological node, or a utility. In the case of this subject, we gathered the full 23andMe Single Nucleotide Polymorphism (SNP) of the individual and filtered out any genetics information that is pertinent to the GNB structure related to CRF. We find that the subjects have several genes associated with nutrition, exercise, and sleep that impact the graph structure, as shown in Table 5.7. We choose to focus the example on the exercise gene, Alpha-actinin-3 (ACTN3). ACTN3 is a binding protein that plays a role in skeletal muscle fibers. This gene regulates the fast-twitch muscle fibers, essential for regulating the coordination of muscle-fiber contraction [171].

Given this genetic makeup of one C and one T variant, the strength between low-intensity exercise and high power exercise will have a balanced effect on the user’s GNB biological nodes. If this user were to have two C copies of the ACTN3 gene SNP, then the edge between power exercise and muscle node attribute values would change. In this case, for a given high power interface event detected would increase the resulting change in the muscle fiber. However, for a low-intensity exercise event, the consequent change in muscle composition would be less.

Subject Genetics	Characteristic	Gene	Variant/Marker
<b>Nutrition</b>	Alcohol Flush	ALDH2	2G
	Caffeine	CYP1A2	rs2472297
	Caffeine	AHR	rs4410790
	Lactose Intolerance	LCT	rs4988235, 1A 1G
	Saturated Fat & Weight	APOA2	rs5082
<b>Exercise</b>	Muscle Composition	ACTN3	rs1815739, 1C 1T
<b>Sleep</b>	Deep Sleep	ADA	rs73598374, 2C
	Sleep Movement	BTBD9	rs3923809, 1G 1A

Table 5.7: **Genetics SNP.** Genetics SNP profile of subject related to the CRF GNB.

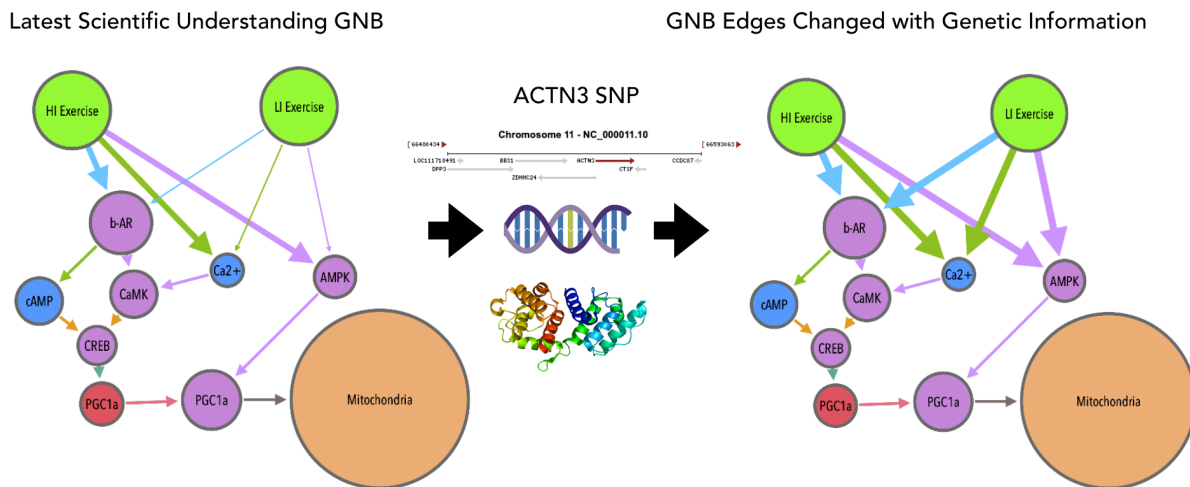
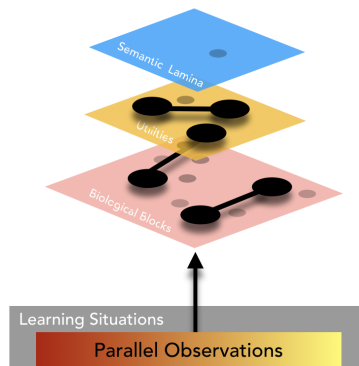


Figure 5.34: **Genetic Update of Model of Event to Biological Node.** This figure demonstrates the difference in the model between using the most updated general knowledge and knowledge, specifically about the user’s genetics. Because this user has two different types of ACTN3 gene, the impact of both HI Exercise and LI Exercise is higher.

This summarizes the knowledge-based approach to updating the graph. We continue next with data-driven model updates.

## 5.6.2 Parallel Observations



The first of two data-driven learning models we employ in this work is applicable when two nodes are observed with incoming data and are linked together. In the instance of our experiment, the two nodes that are linked are heart rate and power. Increases in heart rate result in increased power, and this is well understood both colloquially and in the biomedical literature. For each individual, this parameter is also always changing based on how their CRF state is changing. When plotting heart rate versus power, we notice the data ends up in two patterns. The first pattern is split with various linear clusters, as shown on image A in Figure 5.35. This pattern occurs due to a physiological phenomenon called heart rate drift, also known as aerobic decoupling. Aerobic decoupling is when the heart rate dramatically shifts for the same power output due to cardiac fatigue. Instances like image A show this effect happening. The green-colored portion of Figure 5.35 shows the beginning of the ride, and the red nodes indicate the values at the end of the ride. We can see that there is a shift in this twice reflecting initial fatigue followed by another step of extreme exhaustion. Because the relationship is changing so much within one ride, this instance is not a good source of parallel observation learning. The aerobic decoupling algorithm confirms this, with a decoupling of 40.3 percent. Image B, on the other hand, shows a single linear cluster that does not have any distinct separation. The aerobic decoupling algorithm confirms this, with a decoupling of 0.9 percent. Therefore the data from this ride can be used to give a quality relationship between the two variables. We use simple linear regression for this model, but more complex learning approaches can easily be applied if desired. To teach the system to look for high-quality relationships, we can compute the aerobic decoupling factor for each ride easily with the Algorithm 2. Using the latest value of this relationship, we can update

the edges between heart rate and power in the GNB. The latest value for the last month of data (December 2019) on the test subject is shown in Figure 5.36.

---

**Algorithm 2** Aerobic Decoupling

---

**Input:** Power and Heart Rate event streams ( $P, HR$ )

**Output:** Aerobic Decoupling Percentage ( $AD$ )'

**Begin**

1: for each activity

2:     if power and heart rate data present ( $P, HR$ )

3:         Compute power-heart rate ratio for starting half of event, ( $S$ ).

4:         Compute power-heart rate ratio for last half of event, ( $L$ ).

4:          $(D) = (S) - (L)$

5:          $(AD)' = (D)*100 / (S)$

6:     end for

7: end for

8: Return ( $AD$ )'

**End**

---



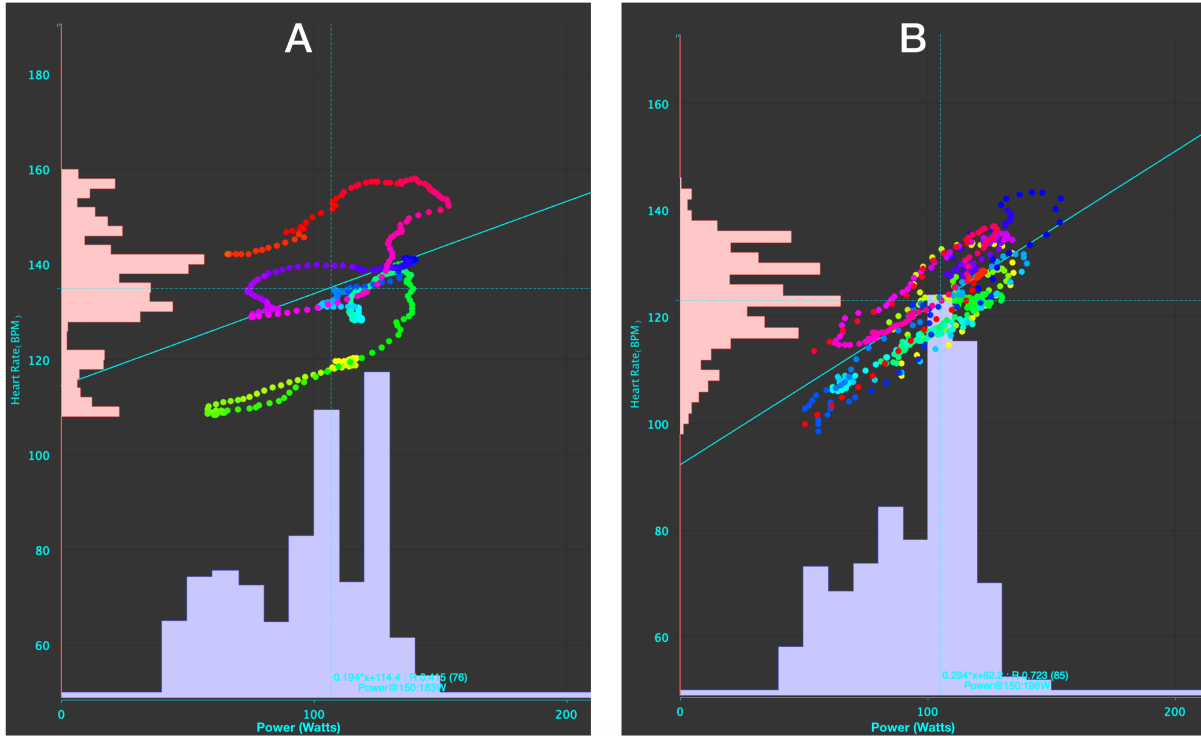


Figure 5.35: **Parallel Data Model Learning.** The above two graphs show heart rate data related to power data in two different activities. During activity A, we can see three distinct linear bands of relationships in red, magenta, and green. These clusters are separated due to a phenomenon called aerobic decoupling. We can check the causal agent of these multi clusters in the ride data by referencing the aerobic decoupling node. In ride A, it is 40.3 percent, versus Ride B is 0.9 percent. For this reason, the graph system can causally understand that activity B is a better sample. This allows for learning the relationship with higher quality.

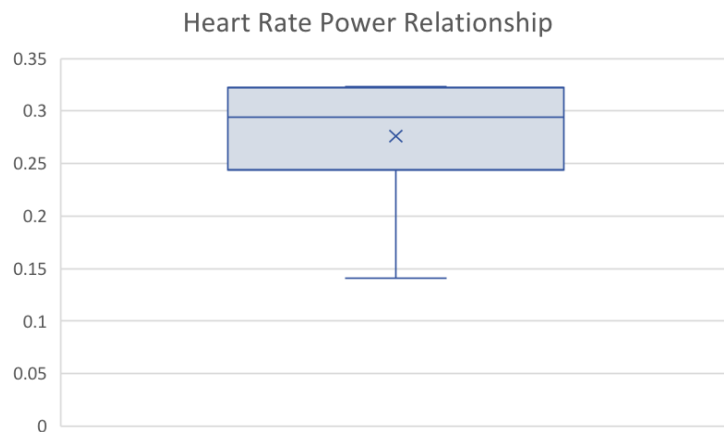
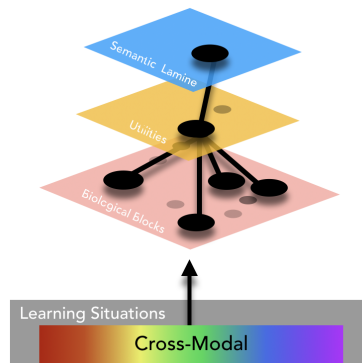


Figure 5.36: **Heart Rate and Power Relationship.** This box plot shows the relationship between heart rate and power for the test subject from various instances within a one month period.

### 5.6.3 Cross-Modal



Cross-modal is defined as understanding of a situation from multiple *types* of inputs [150]. In this case, cross-modal learning takes various inputs of the biological variables and events to estimate the value of a utility node. This set of experiments show that we can improve the quality of CRF estimation via this cross-modal approach. The estimations can improve by combining multiple cheap (in terms of monetary or energy costs) sensors. We use three features from the user’s event and data-streams that all have a direct impact on CRF: 1) Rate of vertical ascent in meters (VAM), 2) Training Impulse (TRIMP), and 3) Cadence-detected Active Time.

Per second power output values collected from the strain gauges serve as the ground truth in our experiments for bio-variable estimation. We used a rolling average of maximum 4-minute power output per day over 42 days to generate the ground truth for our experiments as the utility node to estimate. Forty-two days is a gold standard when understanding short-term changes in CRF [33]. Unsatisfied with this static parameter, we experimented to find out the optimum time to be considered for aggregating the metrics while estimating CRF values. We plotted the test error for the global models utilizing one metric in Figure 5.37. We can see that while there is some variation in mean error, the 95% confidence intervals overlap for all time windows, and we cannot find an optimum time window to use in our experiments based solely on the data. Therefore we have used the clinically recommended period of 42 days to aggregate the past exercise events.

We trained two sets of linear regression models for a global model and a personal model for each feature. The personal model for each individual was trained on with a 70 percent training subset of data for the subject and tested on the remaining 30 percent data.

**VAM** models are trained to predict average power output (normalized by body weight) in 4-minute windows in an activity using the VAM in the time window, and the maximum estimated power output is then used to compute a continuous daily estimate for VO2max [52]. We trained models with varying slope thresholds to identify the impact of slope on estimate accuracy. As the slope increases, the effect of other resistance factors (such as wind, rolling resistance) decreases, and the model performs better (Table 5.8). We choose which model to use based on the maximum slope observed in the 4-minute windows, for example, if in a ride the maximum slope observed in a 4-minute interval is 5.3%, we would choose the model trained on intervals where the slope is greater than 5%. Since we are predicting body weight normalized power using VAM, none of the two metrics are greatly influenced by individual parameters. This is reflected in similar global and individual model performances for VAM (Figure 5.38).

**TRIMP** is a method to quantify training load, and considers the intensity of exercise calculated by heart rate during exercise [109]. TRIMP captures the workload on an individual's heart in the last 42 days. We trained a linear regression model to predict an individual's VO2max value based on their total TRIMP score in the past 42 days. This metric proved to be more effective in a personal model than a global model as different individuals have different heart rate responses to the same exercise intensity, Fig 5.38

**Active time** is the actual amount of time the individual was actively putting in effort in the past 42 days. We obtained this metric using cadence values collected at per the second resolution. We also trained a linear regression model to predict an individual's VO2max value based on their total activity time in the past 42 days. Similar to TRIMP, this metric

performs better in a personalized model than in the global model as different individuals would have a different response to the same exercise volume (fig. 5.38).

**Combination models** outperform their constituent models in all our experiments, as shown by the error plots in Fig 5.38. The estimates from the previous models were combined using a weighted average, where weights for a model estimate are inverse of the model's training error. The error in estimates for these models is reported in Figure 5.38 and discussed in this section. We can see from the plot that the best model in terms of average error and variance in error utilizes all available data streams.

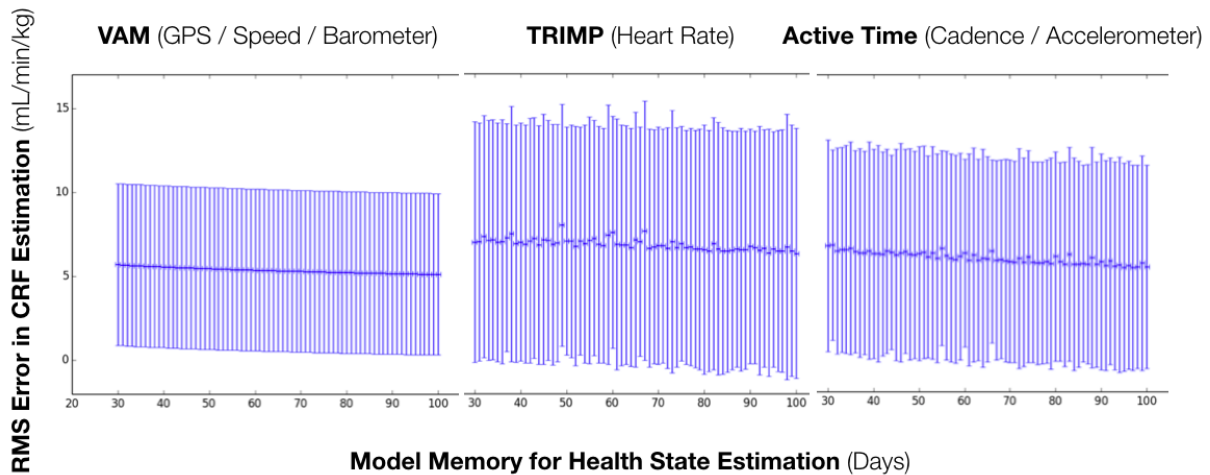


Figure 5.37: **Memory Parameter Optimization.** For each of the three sensor modes, we tested the model to see if there was an optimal amount of days to use as a memory for the CRF prediction. Any model memory of 10-100 days gave similarly accurate learning results, demonstrating there was no optimal learning memory for the model.

Table 5.8: **Slope optimization of VAM models.**

Slope threshold (%)	Test Set RMSE (Rel. Power)	Training Set RMSE (Rel. Power)	Training R Squared	Size of training set
0+	0.726	0.665	0.381	16810792
1+	0.620	0.537	0.527	10728610
2+	0.557	0.473	0.593	7952334
3+	0.488	0.424	0.655	6236507
4+	0.451	0.391	0.695	4884243
5+	0.420	0.363	0.732	3509724
6+	0.405	0.344	0.760	2294199
7+	0.395	0.328	0.781	1400488
8+	0.365	0.318	0.793	781841
9+	0.347	0.317	0.797	423315

Table 5.9: **VAM Parameter Optimization.** For the VAM based model, there is a beneficial effect to choose samples in which the slope gradient of the activity is steeper, which quantifies how much weight to assign to data samples through the quality metric of slope in our learning model.

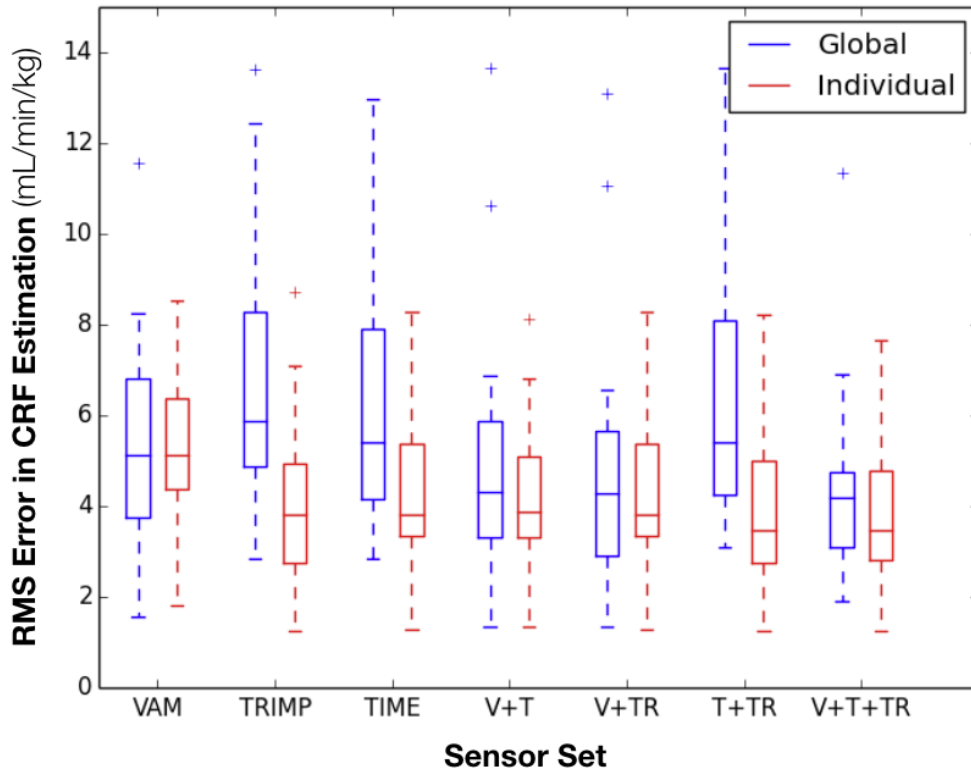
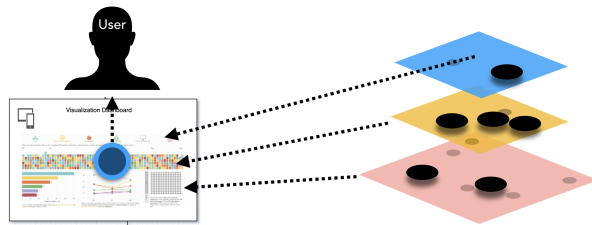


Figure 5.38: **Cross-Modal Sensor Fusion.** By using multiple nodes in our learning model, we can see a direct improvement in the CRF utility estimation, which shows the value of having a multi-modal sensor fusion approach.

In these experiments, we leverage cross-modal data to estimate a needed utility node. From our experimental results, we can see that increasing the number of data streams provide increased performance characteristics in achieving this goal.

The total health of an individual is much higher than any single node attribute, and that we need to understand better the total health state of a particular individual or organ system. We cover some actionable and semantically meaningful visualizations to see how to use the HSE.

## 5.7 Other Visualizations of HSE



How can we further assimilate HSE in a useful way for individuals and health providers? A first step is in the visual analysis of the HSE by users and health experts. Simple mobile applications such as the system

mocked in Figure 5.39 can give a quick overview of the HSE, similar to checking the financial market state or weather state. Expanding this idea to a population provides insight in a detailed manner that can be useful to scientists, researchers, and health experts. We go over an experiment to show such a system in the next section.

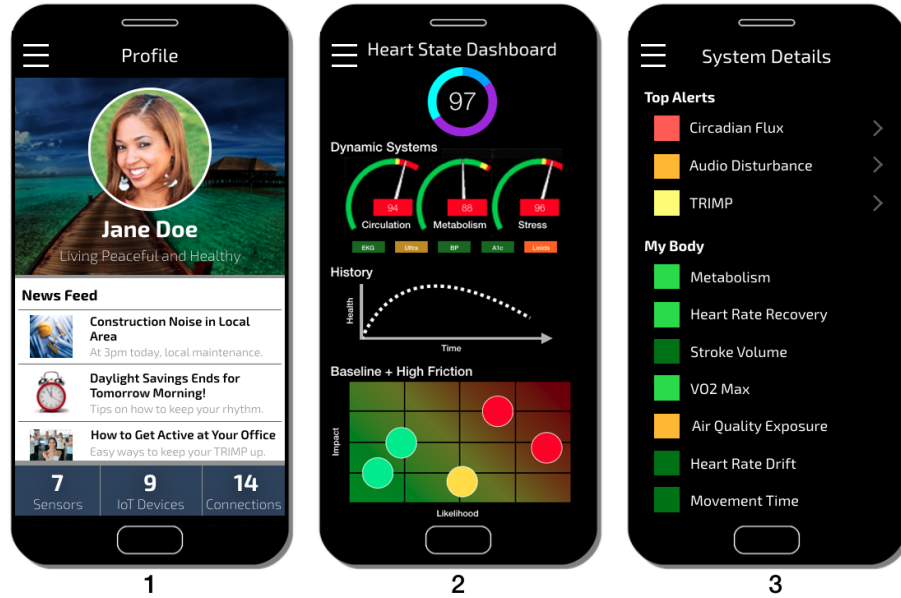


Figure 5.39: **Personal Mobile Dashboard View.** The personal mobile dashboard view is a concept design mock-up of how a user may access their health state. Screen 1 we show how health state can fuel recommendations. Inspired by jet engine dashboards, screen 2 can give a live snapshot of the health state. Screen 3 provides a comprehensive list of all bio-variables and health states being tracked, with a ranking system to provide relevant results at the top.

### 5.7.1 Multi-Person View

Using the same HSE approach for the test subject, we applied this to multiple subjects to produce a health state overview for population health analysis. The dataset used in these experiments includes the same activity sensor data streams at one-second resolution collected on 24 male cycling athletes over an average of 5 years in the United States. Athletes also had strain gauges installed on their bicycles to measure true physiologic power output. The total dataset includes 31,776 activities and 70,178 hours of exercise data (252.6 million data intervals). Social media outlets of Instagram and LinkedIn were also used to gather an image dataset of 50 images per athlete and general demographic background information. From this, we assimilate various CRF biological and environmental nodes through the health state estimation using this data for each of the 24 subjects, as shown in Figure 5.40. We find that even though most of these subjects are all cycling athletes, they have a wide range in these

nodes. Modern-day primary care doctors would not be able to see this when a patient visits. Looking at this data panel across the subject pool, we discover some interesting trends in Figure 5.40. As expected, as the age increases the overall heart health state decreases, since age is a large factor of cardiac health (age range in the panel is 18-57). As age increases, we also see a reduction in crime and noise pollution, suggesting that older individuals can afford to live in safer and quieter neighborhoods. We also see circadian rhythm disruption maximally in the middle ages (20-29), suggested a more erratic lifestyle for those in their twenties. Circulation and metabolism scores also trend (including VO2max / CRF) lower as age increases. Although this is a small sample size to make any strong conclusions, we can begin to see the usefulness of using this cross-modal analysis for clinical applications. This work was published at Association for Computing Machinery's (ACM) Multimedia 2018 in the Brave New Idea track [193].



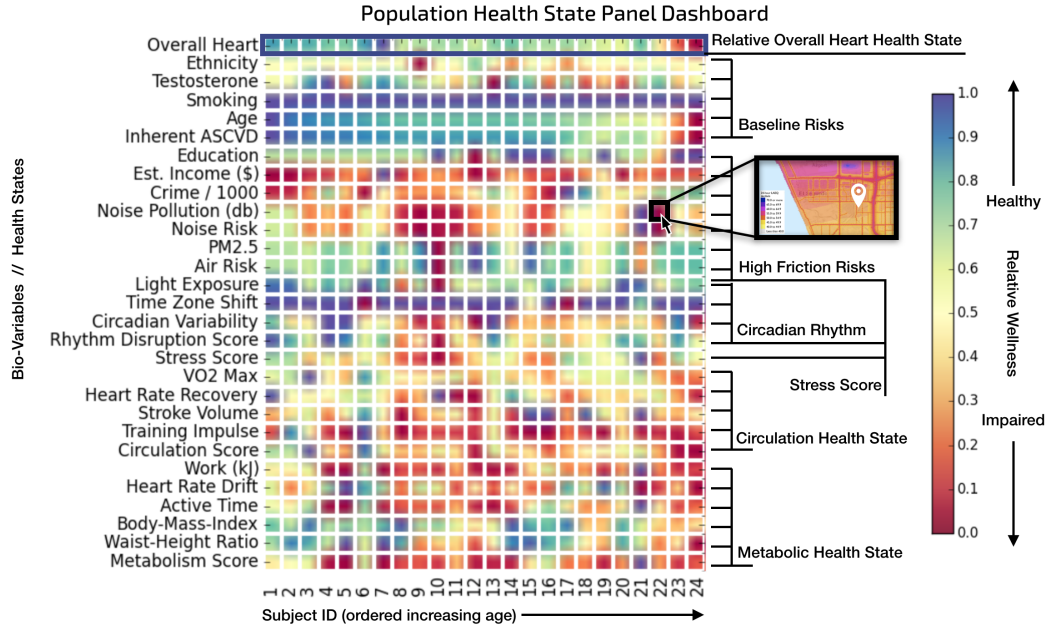


Figure 5.40: **Patient Dashboard View.** The patient dashboard view is a heat map of bio-variables and summary scores that affect each subject. Variables are normalized to a specific color for each box. The variables are graded on a scale ranging from blue being most healthy and red being least healthy. This type of visualization integrates cross-modal data in a manner that a clinician, hospital, public health agency, or any domain expert can use to monitor the health of a patient panel or population. Clicking on a particular box would pull up further insights and details. This highlighted box in this image shows the noise pollution insight for patient 22.

Concluding this chapter, we see how the framework in Chapter 4 can be implemented at a detailed level for estimating the health state of an individual, and how this can scale over time and across more individuals. This working framework is shown in Figure 5.41. The next chapter explores ideas within the original inspiration of using HSE for navigation and control. This exploration is intended to extend HSE beyond this dissertation.

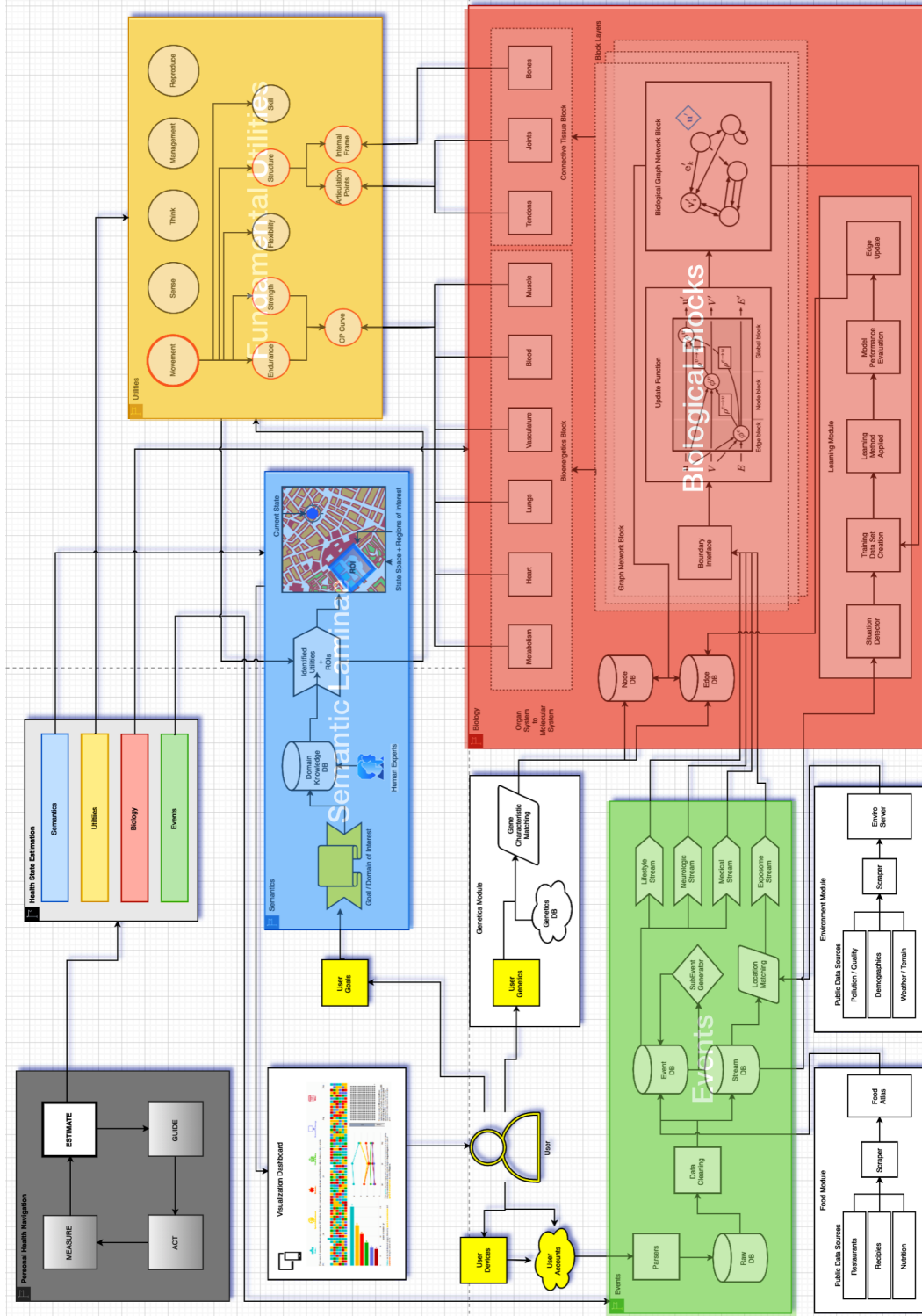


Figure 5.41: **Total Health State Estimation System.** Here we can see all the blocks of Chapter 5 in a system overview.

# Chapter 6

## Personal Health Navigation

Knowing is not enough; we must apply.

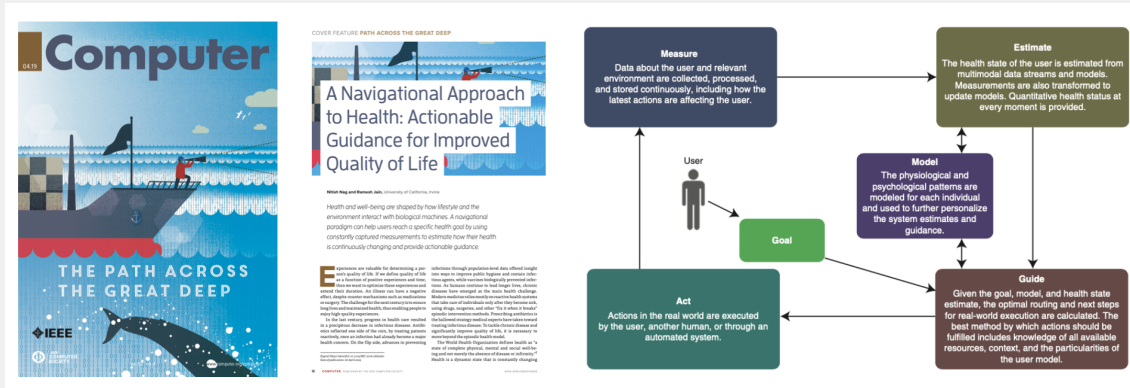
Willing is not enough; we must do.

---

*Johann Wolfgang von Goethe*

Knowing the health state provides information about one's health situation; however, individuals need to take action with this information. Akin to GPS for position, the possibilities to take action with this information are vast. Making decisions about events in one's life allow individuals to change their state towards their wishes. These chosen events become experiences, which are valuable for determining a person's quality of life. If we define quality of life as a function of positive experiences and time, then we want to optimize those experiences and extend their duration. As discussed, health is a dynamic state that is continually changing, based on biology, the environment, and lifestyle. Technology that senses a change in a user's position on the health continuum can drastically improve human health, especially once it is actionable. If people become aware of their biology, they could see how daily life is affecting their health and make informed decisions accordingly. Health decisions are too important to be episodic; they must be an intrinsic part of daily life. We propose

## Box 1: A Navigational Approach to Health



Health and well-being are shaped by how lifestyle and the environment interact with biological machines. A navigational paradigm can help users reach a specific health goal by using constantly captured measurements to estimate how their health is continuously changing and provide actionable guidance. *Cover Issue IEEE Computer April 2019* [186]

**Measure** : Data about the user and relevant environment are collected, processed, and stored continuously, including how the latest actions are affecting the user.

**Estimate** : The health state of the user is estimated from multi-modal data streams and models. Measurements are also transformed to update models. Quantitative health status at every moment is provided.

**Guidance** : Given the goal, model, and health state estimate, the optimal routing and next steps for real-world execution are calculated. The best method by which actions should be fulfilled includes knowledge of all available resources, context, and the particularities of the user model.

**Action** : Actions in the real world are executed by the user, another human, or through an automated system.

**Model** : The physiological and psychological patterns are modeled for each individual and used to further personalize the system estimates and guidance.

the navigation perspective, to make actionable guidance for improved quality of life [186].

This chapter discusses in detail the guidance and action components of navigation.

Navigation is different from recommendation systems and purely cybernetic control. This goal-based guidance system perpetually estimates the current state, computes the best route through intermediate states, and guides actions that lead to health goals. For success, this system requires robust continuous sensing, accurate estimation of the current state, control systems, actuation mechanisms, understanding of the outside world and context,

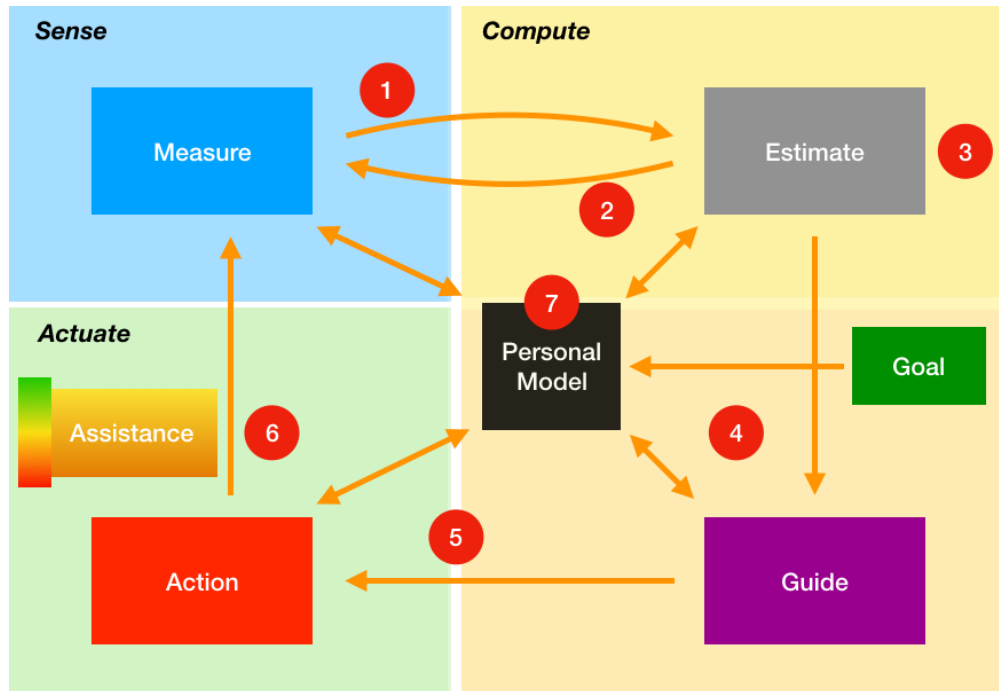


Figure 6.1: **Navigation Vision for Health.** Navigation for health has four key components: Measure, Estimate, Guide, Action. Step 1: Take measurements in the real world through a variety of sensors. Step 2: Additional measurements may be needed if they are not sufficient for a reasonable estimate. Step 3: From these measurements, we can produce an Estimate of the current health state. Step 4: Guide the user based on a reference goal and the user’s model. Step 5: Take actions in the real world through a variety of mechanisms. Step 6: Assisted mechanisms may contribute to the executing of these actions. Step 7: Retake measurements to understand how the person’s health state has changed, and this updates the personal model perpetually.

long-term planning, and goal decomposition at various levels of detail. Navigation must also manage stochastic, chaotic, and noisy environments effectively. This goal-driven, closed-loop sense-compute-then-act cycle combines the pioneering elements from cybernetic systems (well described in electromechanical systems and biology by Wiener et al. and the general problem solver by Newell et al. [292, 199]).

The purpose of personal health navigation (PHN) is to help an individual reach and maintain their desired health state. Systems that provide health navigation are broad, and humans may be integral components. Examples include a personal device such as a mobile phone and an entire Internet of Things (IoT) ecosystem within a building, a human trainer at a

gym, or an advanced cancer center. PHN starts with a user-specified health goal. From there, the system starts to collect measurements about the persons health that relate to his or her goal to estimate current health states. To most optimally reach the desired goal, the system breaks the steps from the current state and provides the next step to an actuation mechanism. All actions, whether advised by the system or not, are continuously measured to provide a new estimation of the individuals health state. A change in the state updates the next action. A cycle of these actions moves the users current health state closer to the target. Upon reaching the destination, the system continuously ensures minimized deviation from that state. The vast unknown nature of health means that we will obtain more diverse multi-modal data in the future and that this phenomenon requires combining these data sources.

While navigation is not the core of this dissertation work, it establishes the purpose of the health state estimation and how to best guide the better health of an individual. In this chapter, we will discuss the step by step process of personal health navigation and work that has been done to address certain challenges [186].

## 6.1 General Health State Space

The first step of navigation is establishing the general health state space, as described in Chapter 4. The “flight envelope” of an aircraft is the inspiration for the state space concept. The envelope encompasses measurements (load factor, atmospheric density, altitude, etc.), which determines the space in which it is safe for the plane to fly in, as shown in Figure 6.2. Similarly, the state space defines all possibilities and capabilities to understand the scope in which the navigation will occur. This state space includes all possible health states a human can exist in. For example, if the state space is the cardiovascular state, all possible

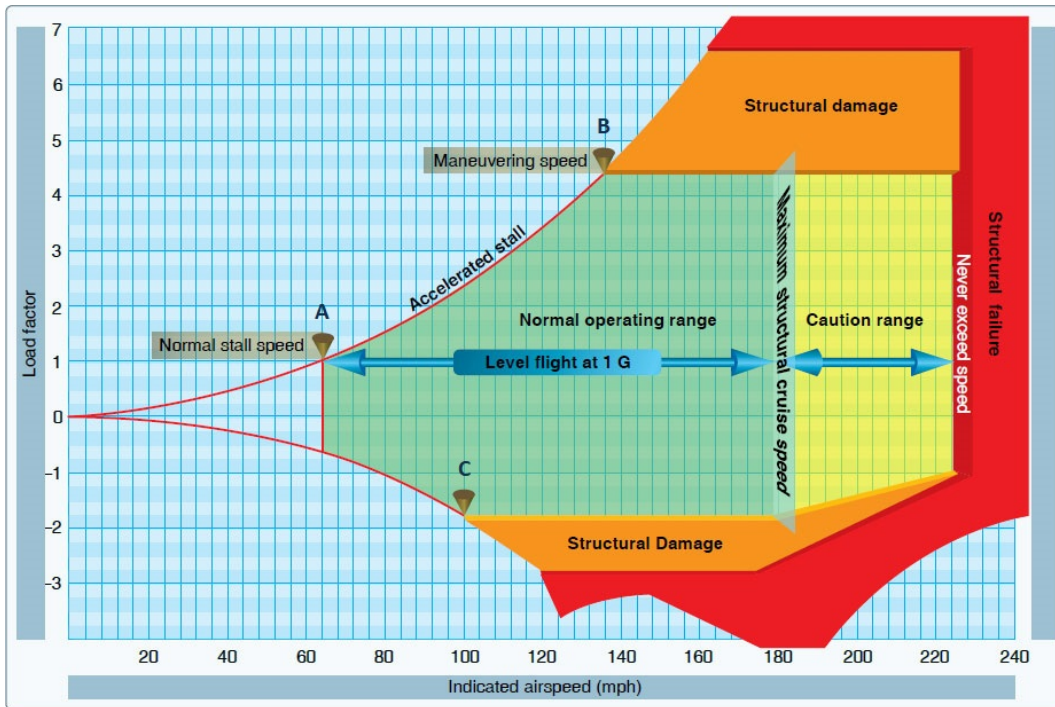


Figure 6.2: **Flight Envelope.** This graph shows the area of total possible states in which the plane is able to fly given its unique parameters.

measurements for the cardiovascular state will be considered, including all components of fitness (with specific measurements such as  $VO_{2max}$ ), cardiovascular disease, and more.

## 6.2 Personal Health State Space

When we apply the state space to an individual, we create a personal health state space, which includes all the possibilities for an individual given their situation. This is by definition a subset of the general health state space. Each person has a unique and finite possibility of biological states. Classic examples such as height, eye color, or fingerprints have a small set of states after reaching adulthood. However, many aspects of biology have an wide range of possible states, which change based on daily life. Consider how exercise changes the mitochondrial content of cells, how nutrition alters our metabolic processes, how medications modulate our blood pressure, and how ultraviolet light damages DNA. The changes in these

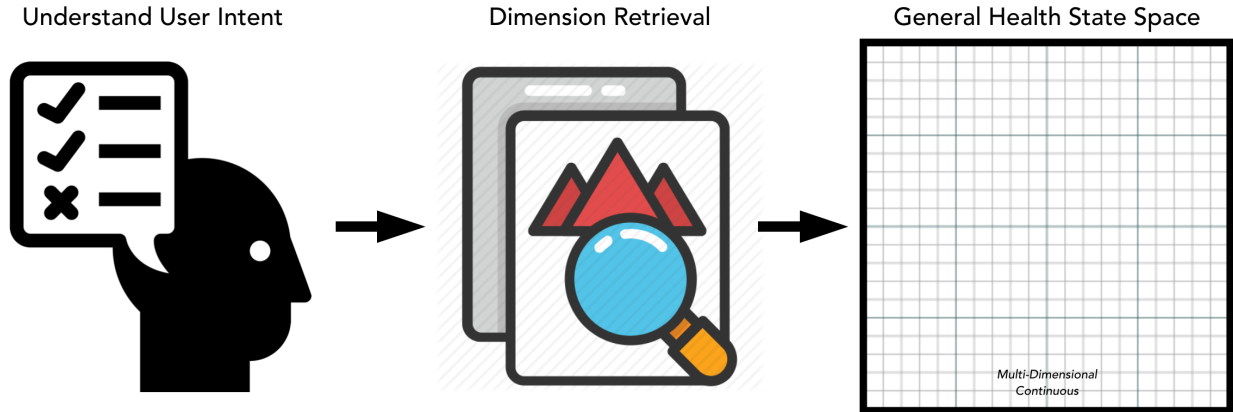


Figure 6.3: **General Health State Space.** The state space includes all the possible states that all humans can exist, with several dimensions. In order to interact with this state space effectively, a user may provide a particular intent, which can reduce the dimensionality to the relevant regions to visualize the general health state space.

states are what determines the movement in the state space. Mapping the personal health state space (PHSS) in totality is a complex, multidimensional challenge. We can apply this concept to what we observe in flight envelopes, in which specific models of planes have unique state spaces. Each type of plane will have its possibilities by which it can fly. Narrowing the state space to the individual's possibilities give a more specified scope best set up for personalized navigation.

## 6.3 State Transition Network

Within the personal health state space, there are connections made between each of the individual state possibilities. The connections themselves are the state transition network. Each edge within the network represents transitions that take the person from one node to another. With the cardiovascular state, for example, we determine which inputs will transition the heart state. The state transition takes into consideration all the inputs that will lead to the next state, therefore making connections or relationships between states. If one were to exercise, experience stress, or have a high salt meal, they are transitioning



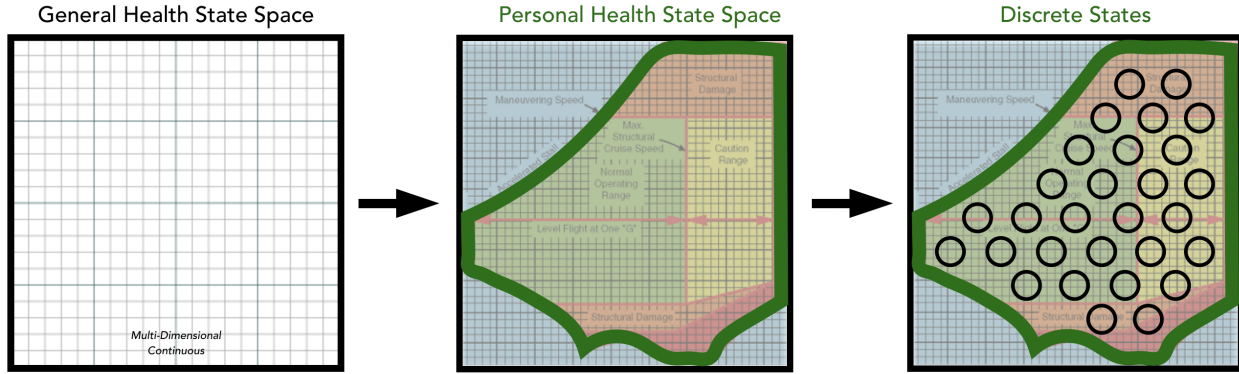


Figure 6.4: **Personal Health State Space.** The PHSS includes all the possible states that a specific human can exist at the current moment and at future time points. This state space is a subset of the general health state space. To convert this space to a computationally efficient world, we may apply a discrete layer on top of this state space, allowing for graph-based approaches for further applications.

between these states in their PHSS. To understand the PHSS in more detail, we must identify the knowledge layers and regions of interest within the space relevant to the topic of interest. Connections within a PHSS between two nodes will be unique based on the individual. For example, for person A and B to grow one pound of muscle, the inputs needed for A will be different than person B. Therefore the state transition network is also a unique part of the PHSS.

## 6.4 Knowledge Layers and Regions of Interest

Layers on top of the PHSS contains all relevant health domain knowledge, similar to physical maps that use latitude, longitude, and altitude to describe the globe independently of any knowledge layers. Information layers such as roads, oceans, country borders, and satellite imagery allow for navigation within the space, depending on the context (driving requires roads and traffic layers, while flight requires air class, way-points, and airport layers). Knowledge layers represent the real world, in which humans can make sense of their state concerning their interests. A semantic translation to the relevant PHSS will depend on the user's re-

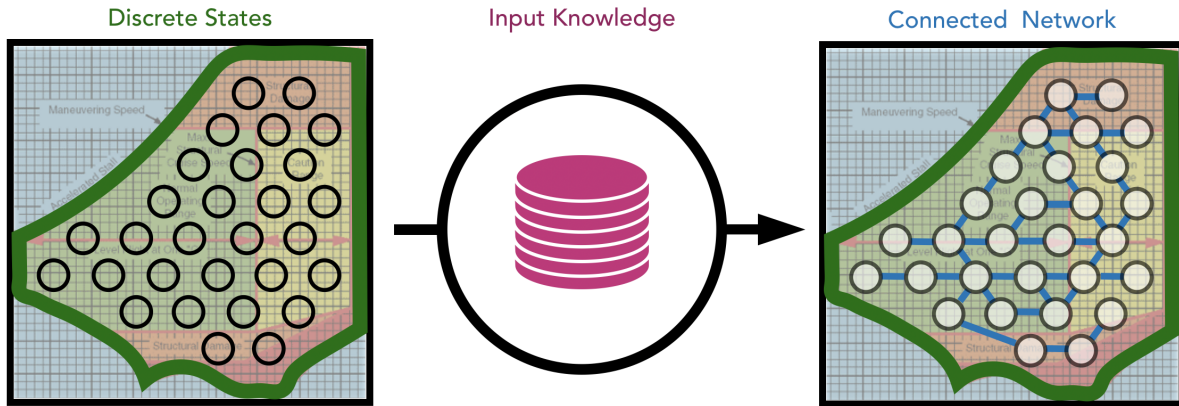


Figure 6.5: **Connected State Space Network.** The state space network includes all the possible transitions from one state to another for the specific individual. The connections contain the information of what inputs the matrix will cause a particular state to convert to the next state. The edges here are directed (although not visually shown). These connections are not static and are repopulated, given a reflection of the current situation.

gion of interest. For example, mapping the coordinates of a “cyclist” in the PHSS requires matching relevant coordinates within the desired region of interest in the state space. This region of interest is a particular area within the state space relative to where the individual currently is in their health status. A defined distance allows the individual can see how close or far they are from being in this region and the various nodes or states that would cause one to get there. Similarly, one can determine their risk for heart disease based on their current state’s distance from an undesirable region. Mapping current location distance from regions of poor health brings about a new approach to quantify one’s relationship to disease risk.

## 6.5 Current State Estimation

It is important to note that there is constant flux within an individual’s PHSS based on one’s location in the state space. For success, navigation must be able to assign an accurate location within the PHSS, just as GPS provides physical world navigation. Understanding this location is the HSE (all previous chapters). The same HSE tool may be useful for

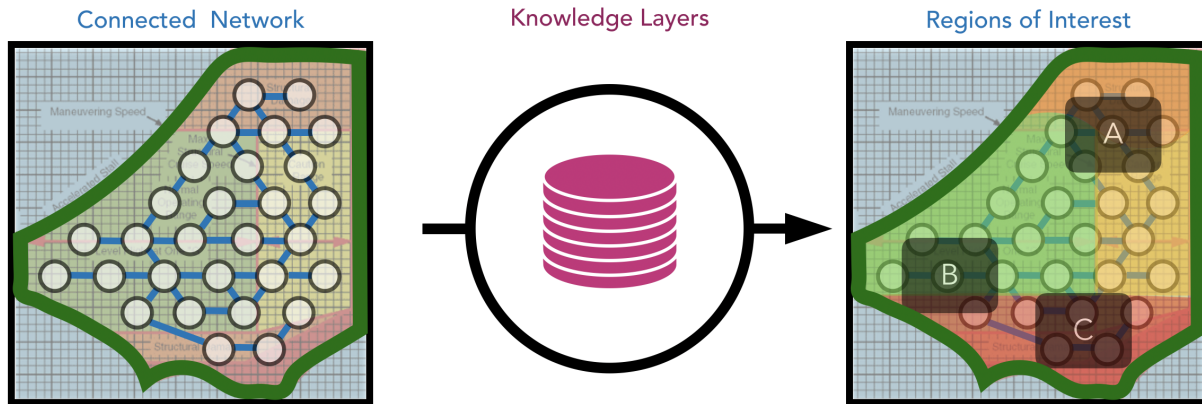


Figure 6.6: **Knowledge Layers.** Layering information on the state space allows for semantic interactions. Domain knowledge specifies these layers from the relevant database. We can gather specific regions of interest from knowledge sources that can label the state space so a user can understand the possibilities for their health in natural language or a relevant visualization. In this example, specific regions of interest are labeled with the colors of green, orange, yellow, and red. Red is an example of an undesirable region, while blue represents regions that are not within the person state space possibilities of the user. The black boxes represent points of interest, which is essentially a more confined region of interest.

many applications. For example, monitoring a cardiovascular health state is useful to both endurance athletes and heart disease patients. Estimation techniques have been of great interest in designing many applications, but health applications will require increasingly deep biological knowledge layers to define and estimate health states. This estimation could indicate its proximity to biological dysfunction or its capacity for the desired goal.

## 6.6 Goal Decomposition

For navigation to occur, an individual's requires specifying a goal state. Goal states can be specified as ROI or a specific point. If an individual has a region of interest for becoming a cyclist, they can more specifically choose a pointed goal such as competing in a 100-mile road cycling race. Goal decomposition is needed to translate the goal state into short-term

goals. This process occurs through breaking down the goal to identify the unique utilities relevant for the specified goal.

First, in this process, the system will need to translate the semantic goal to the PHSS. Because the PHSS is large, we must break down a macro health goal into several intermediate states and subgoals that the system can focus on reasonably. There may be a set of short-term goals at any given moment, each with its own set of intermediate states and subgoals. At the lowest atomic level for an application, PHN lists fulfillment tasks in priority order so that users can execute them in the real world to attain a subgoal or the next intermediate state.

To become a road cyclist, for example, there will need to be a process broken down into the key utilities (i.e., develop muscle strength, improve cardiovascular health, ensure body alignments). Each of these intermediate steps will have various subgoals required for the optimization process. If the individual's utility is to develop bone strength, they will need adequate nutrition and weight training. Each subgoal translates to a smaller state space. In the case of nutrition, models will list macro (carbohydrates, protein, fat) and micro (vitamins, minerals) nutrients. These components may then be the chosen atomic-level intermediate states for applying nutrition.

### **6.6.1 Timescales**

Goals and corresponding subgoals occur at various timescales. Take for example the cyclist whose race is in six months. The cyclist must maintain daily subgoals such as establish cardiovascular endurance (i.e., a two-hour bike ride, four days a week). We can then break it down into minute by minute subgoals to maintain a particular power output for interval training, to build a higher power to weight ratio. When looking at the biological level,

the subgoals may be positive gene expression for mitochondrial biogenesis and improved metabolism, which has a second by the second timescale.

### **6.6.2 Recurring Goals versus Achievement Goals**

There are two types of goals. One is a recurring goal that is an ongoing goal, such as a habit or way of living. For example, brushing teeth would be considered a recurring goal that an individual maintains over time. This requires constant navigation till the recurring goal is terminated by the user. There are also achievement-based goals, which are defined situations that once achieved terminate the navigation. Once the user meets the achievement (such as winning a race), they must then set a new goal.

### **6.6.3 Synergistic Goals versus Competing Goals**

We must note that there may be multiple goals that are either working in synergy or competing with one another. For example, if an individual has the goal for high performance in a road race and also would like to be an enduro mountain bike racer, there may be overlap in synergistic goals, such as the cardiovascular training required to perform in both types of events. However, as enduro mountain bike racing relies on high skill in demanding terrain and upper body strength (therefore increased weight), it will conflict with the training for high endurance activities, such as the 100-mile road race which require being at a lighter weight.

## 6.6.4 Identifying User Motivations for Goals

Identifying user motivations is critical for goal setting and, ultimately, the effectiveness of any lifestyle change. Engagement is dependent on providing recommendations that primarily fuel intrinsic motivation while using extrinsic motivation for further support. This understanding builds the foundation of sustainable behavioral adaptation for optimal personalized health benefits. **Intrinsic motivation** arises from internal enjoyment or satisfaction in a particular task, while **extrinsic motivation** refers to motivating factors from the outside world, such as rewards or punishment. While both can lead to lifestyle activities that benefit health, intrinsic motivation drives long term user engagement.

We can use multi-modal data to learn if a user is executing an action from intrinsic motivation or external factors. For instance, if a user repeats an activity without prompting and without external influence, this gives us an increased weighting that the motivation is intrinsic. If we detect themes and concepts in their social media consumption, and one takes part in an activity when a reward prompts them, we can consider this as more weighting towards the extrinsic motivation [187].

## 6.7 Guidance

The guidance process towards the defined goal includes creating a route and recommending the next steps along the way. Computers play a role in providing both short-term guidance and longer-term guidance; however, to predict the future, provide guidance, and understand the preferences and particularities of an individual, personal models are needed. These models, which establish the premise that each individual is a unique system, are needed to best estimate the users health state and how various inputs uniquely move them in their PHSS. Models are then accessed to provide precise guidance towards a goal state.

## Box 2: Intrinsic and Extrinsic Motivation Modeling Essential for Multi-Modal Health Recommender Systems

### Intrinsic and Extrinsic Motivation Modeling Essential for Multi-Modal Health Recommender Systems

**Nitish Nag**  
University of California, Irvine  
Irvine, California  
nagn@uci.edu

**Mathias Lux**  
Alpen-Adria-Universität Klagenfurt  
Klagenfurt, Austria  
mlux@uni-klu.ac.at

**Ramesh Jain**  
University of California, Irvine  
Irvine, California  
raj@uci.edu

**ABSTRACT**  
Managing health lays the core foundation to enabling quality life experiences. Modern computer science research, and especially the field of recommender systems, has enhanced the quality of experiences in fields such as entertainment, shopping, and advertising, yet lags in the health domain. We are developing an approach to leverage modern data for human health based on motivation modeling and recommendation of actions. Health is primarily a product of our everyday lifestyle actions, yet we have minimal health guidance on making everyday choices. Recommendations are the key to modern content consumption and decisions. Furthermore, long term engagement with recommender systems is key for true effectiveness. Distinguishing intrinsic and extrinsic motivations from multi-modal data is key to provide recommendations that primarily fuel the intrinsic motivations, while using extrinsic motivations to further support intrinsic motivation. This understanding builds the foundation of sustainable behavioral adaptation for optimal personalized lifestyle health benefits.

**CCS CONCEPTS**  
- Information systems -> Mobile information processing systems, -> Human centered computing, -> Ubiquitous computing, Mobile computing, Ambient intelligence.

**KEYWORDS**  
Health Informatics, Cross-Modal Data, Personal Health, Cybernetic Health, Machine Learning, Wellness, Health Recommender Systems

**1 INTRODUCTION**  
Health is essentially a product of our genome and lifestyle [3]. Lifestyle is the primary controllable aspect of our health. Although largely preventable, lifestyle-associated diseases continue to lead the rapid rise of non-communicable diseases across the globe [2]. Predicting changes in routine lifestyle habits is a tremendous psychological hurdle. Mobile platforms such as food analyzers from image, accelerometer based activity sensors, and life logging [3] gather data about lifestyle factors, but largely remain as isolated efforts. Personalized mobile digital health systems of all sorts are proposed or deployed as a general solution to provide the copious sea of data that we generate in our personal lives and to help us understand our health and to help us make better decisions. However, with multi-modal data, it is not clear how to best integrate this data to help us understand our health and to help us make better decisions. We are developing an approach to leverage modern data for human health based on motivation modeling and recommendation of actions. Health is primarily a product of our everyday lifestyle actions, yet we have minimal health guidance on making everyday choices. Recommendations are the key to modern content consumption and decisions. Furthermore, long term engagement with recommender systems is key for true effectiveness. Distinguishing intrinsic and extrinsic motivations from multi-modal data is key to provide recommendations that primarily fuel the intrinsic motivations, while using extrinsic motivations to further support intrinsic motivation. This understanding builds the foundation of sustainable behavioral adaptation for optimal personalized lifestyle health benefits.

A key concept in effective lifestyle recommendation will be the distinction between intrinsic and extrinsic motivations [1] (cf. Fig. 1). Intrinsic motivation arises from internal enjoyment or satisfaction in doing a certain task, while extrinsic motivation refers to motivating factors from the outside world such as rewards or punishment. While both can lead to lifestyle activities that benefit health, intrinsic motivation drives long term user engagement. Self-driven healthy activities and emotions (rather than rewards) healthy behavior are only sustainable and effective in long term developments. Extrinsic motivation, like recommendations from medical doctors or therapists, social pressure through groups of like-minded people, or an mobile phone app or fitness tracker device keeping users into action, on the other hand is a necessary part, but should only support intrinsic motivation. Intrinsic motivation is highly personalized, user like health and the reaction of the human body to different activities.

Multi-modal data can be used to learn if a user is exercising an action from intrinsic motivation or from external factors. For instance, if a user reports an activity without providing and without influence in their social media, this gives us an increased weighting that the motivation is intrinsic. If we detect fitness and energy in their social media consumption, and the activity is done when a friend prompts them, we can consider this as more weighting towards the extrinsic motivation.

We identify two potential inhibitor groups of users. These are not inhibitive starting points for our position. There is no formal distinction or definition available without further research. Members both groups can be highly successfully or externally motivated.

Figure 1: Intrinsic motivation relies on the user enjoying the task due to a sense of personal satisfaction. Materialistic rewards for positive reinforcement and punishment for negative reinforcement are examples of extrinsic motivations, which are applied to the user from the outside world, to collect, manage and provide ad hoc retrieval. The next step will be to use this data to give actionable decision recommendations to support individuals.

Distinguishing intrinsic and extrinsic motivations from multi-modal data is key to provide recommendations that primarily fuel the intrinsic intentions, while using extrinsic motivation to further support intrinsic motivation. This understanding builds the foundation of sustainable behavioral adaptation for optimal personalized lifestyle health benefits. This paper was written in collaboration with Mathias Lux of Alpen-Adria-Universität Klagenfurt, Austria. *Arxiv, August 2018* [187]

- Understand what components are intrinsic and extrinsic motivators for an individual are key to develop more intelligent recommendations to guide a person towards good health.
- Using intrinsic motivations as a basis for lifestyle change is much more effective for long term engagement.
- Extrinsic motivations are useful to engage in short bursts to push behaviors, but should be used judiciously.

The personal model can build knowledge and predict many aspects of an individual's life, such as how a person reacts to different stimuli under specific conditions or physiological change from an intervention. To model a person, a combination of long- and short-term information must interact. Long-term models of an individual can come from the genome or the history of event patterns and behavior. Biological or high-resolution continuous sensors can capture short-term models. These models are not static; they continuously change with

218

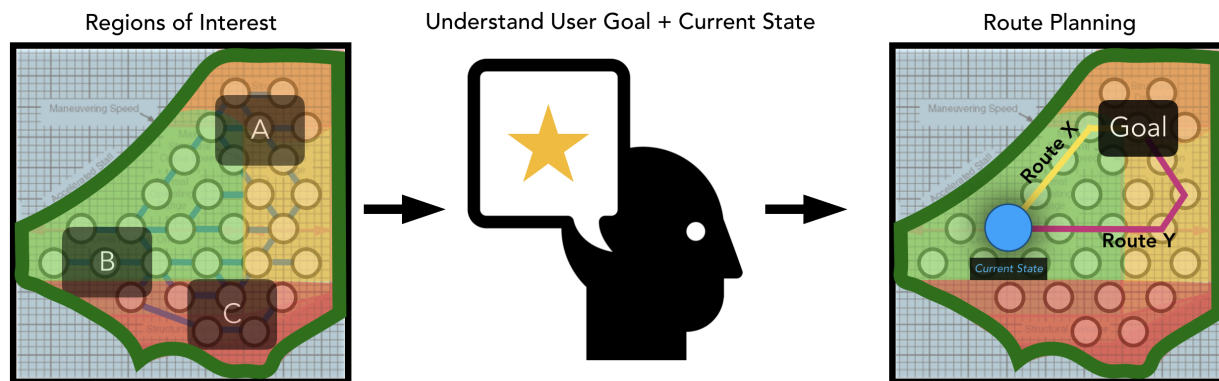


Figure 6.7: **Goal Selection and Route Planning.** Given a particular interaction layer, a user can select a goal of choice. We compare this goal to the current state of the user in order to determine routes that can lead to the goal state. Multiple routes may be generated based on optimization criteria that the user or the system can weigh.

an individuals age, life events, and other life parameters, which makes model-building a dynamic and causal understanding process.

After measuring, estimating, modeling the individual, and receiving a goal the user needs to receive guidance on the next step needed to reach that goal. For an individual to make decisions about events leading up to their goal state, the system routes through intermediate steps through the PHSS. At each moment in time instructions are given for the next appropriate actions.

### 6.7.1 Routing

Having a map, location, and goal (or destination) now sets the stage for routing. As we cannot go back to previous time points, we must move forward through intermediate states over time towards a desired goal state. Making a route on a map requires not only knowing the start and endpoints but also all the layers of roads and traffic. In the case of PHN, each set of intermediate states and sub-goals will have its layer of information, which is relevant for mapping, along with costs and constraints to transition among intermediate states. Interactions in PHN will be extremely complex due to its large dimensionality. We



must handle competing user goals, through methods such as prioritization or weighting. Meansends analysis or other problem-solving techniques, along with appropriate routing algorithms, will reveal the best intermediate states for the user to reach the goal. There may be multiple routes to get to the desired goal. However, route selection may be made by various optimization criteria that include user preferences, efficiency, speed, and resources.

For the example case of food routing for pregnancy. It is necessary to map available food options, nutrition, and locations, along with how each choice can fulfill the tasks demanded over a period of time of nine months. Consider models for breakfast, lunch, snack, and dinner in the daily sub-goal of a nutritional state. There must be coordination of total components to reach the sub-goal appropriately. For example, if breakfast does not provide enough iron, then it must be compensated for in other meals. Researchers must further consider the logistics of adding a dish into the equation. Hence, solving routing for various PHN components remains an open opportunity.

### **6.7.2 Actions and Cybernetic Control**

Actions of the person help lead them to a specified goal. Cybernetic control pairs the individual user and digital assistance to enact these real-world changes to transition the health state. Control mechanisms steer individuals towards the state transition on various timescales. Therefore, recommendation serves as a tool for the cybernetic control in a particular moment in time. Recommendation may help plan for a single moment in time, but it does not consider routing through the state space. This is a key difference between navigation and recommendation.

When applied to health, combining sub-population and individual data dynamically into a system gives real-time personalized evidence-based actions unique for each individual through sensors and specific feedback guidance loops. In the case of the cyclist, the control mechanism

## Box 3: Cybernetic Health

**Cybernetic Health**  
 Nitish Nag, Vaibhav Pandey, Hyungik Oh, Ramesh Jain  
 Department of Computer Science, University of California, Irvine / Irvine, California, 92617, USA  
 Prepared March 2017, Arxiv Version Accepted 21 May 2017  
 Corresponding Author: N. Nag, nag@uci.edu

**ABSTRACT**  
 Future health ecosystems demand the integration of emerging data technology with an increased focus on preventive medicine. Cybernetics extracts the full potential of data to serve the spectrum of health care, from acute to chronic problems. Building actionable, cybernetic navigation tools can greatly empower optimal health decisions, especially by quantifying lifestyle and environmental data. This data to decisions transformation is powered by intuitive event analysis to offer the best semantic abstraction of dynamic living systems. Achieving the goal of preventive health systems in the cybernetic model occurs through the flow of several components. From personalized models we can predict health status using perpetual sensing and data streams. Given these predictions we give precise recommendations to best suit the prediction for that individual. To enact these recommendations we use persuasive technology in order to deliver and execute targeted interventions.

**Author Keywords**  
 cybernetics; diabetes; quantified-self; objective-self; precision medicine; predictive health; prevention; event mining; context awareness; mobile health; persuasion; personalized

**INTRODUCTION**  
 Navigation systems, like Google Maps, became rapidly popular by providing expert knowledge and real-time personalized contextual information to guide travel. Personal lifestyle guidance in such a manner does not exist in health care systems. In order to shift the focus in health systems from temporary fixes to long-term solutions such a guidance system must be implemented [60] [48]. Commonly, health systems focus primarily on medical methods to manage health when a patient becomes ill. By evolutionary design, optimal health is universally desired (and should be provided) at all times. True health outcomes result from actions taken in every moment and place, not just medical interventions during sickness. Future advancements in health must continuously sense individual needs and rapidly provide the relevant resources to corrective actions ensure health stability. For example, optimal health for chronic diseases like type 2 diabetes (T2D) remains a challenge. Insulin

**Figure 1:** P<sup>2</sup> Cybernetic Health coordinates the elements of personalized, predictive and precision medicine through persuasion techniques that result in disease prevention.

**Figure 2:** Cybernetic Control pairs the individual user and digital health assistance to enact real-world changes to optimize health

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/> or send a letter to Creative Commons, PO Box 186, Mountain View, CA 94042, USA.

Cybernetics extracts the full potential of data to help control steps along the navigation pathway. This paper presents five connected components to enhance control. *Arxiv, May 2017* [192]

- Personalization must take into account the individual model.
- Prediction is possible with this model.
- Precise guidance can be given based on the goal set in the cybernetic system.
- Persuasion of the individual to execute the guided steps is key for lifestyle based intervention.
- Prevention of undesirable regions-of-interest further teaches the model to improve in the next iterative cycle.

for ensuring they are on target for getting to their next state is determined by the power out for each training interval from data collected from the power meter. If their power is too high or too low, the recommendation may be to adjust for the optimal power output in the live moment of training. Based on their biology and the effort during the training ride, they can have actionable next steps on what food they should eat next for optimal recovery [192].

Computing optimal control mechanisms, however, are not enough. For meaningful impact, we must implement the cybernetic control principles in the real world which includes noise

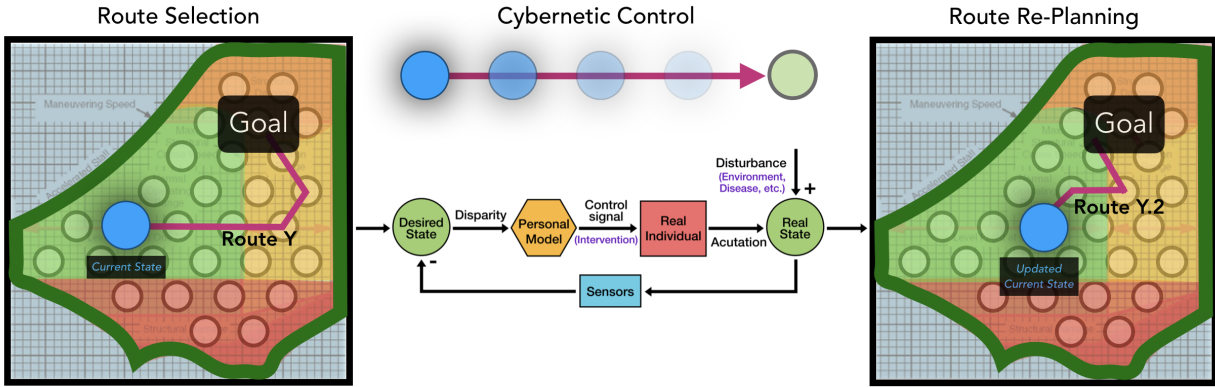


Figure 6.8: **Control and Re-Planning.** Once a specific route is selected, the user and system only need to worry about controlling the state transition for the current time iteration. Cybernetic loops are employed to ensure that the state transition is controlled successfully in the presence of disturbances and noise. Once the state has transitioned to a new state, the system re-plans the most optimal route again. Due to changes in resources, health state, or other factors, the route may be different at the next time iteration. Here we see route Y selected at time  $k$ , and at the next time iteration, the route plan has changed to Y.2 based on the updated state and resources at time  $k+1$ .

and disturbances. For the system to be effective in the physical work, we must consider the spectrum of actuation ranging from fully-automated systems to human-fulfilled actions.

## Automated Actions

Many machines automate tasks such as laundry or aircraft control, which are instances of completely automated control mechanisms that can make an action in the real world control a task. For a type 1 diabetic patient, continuous glucose monitoring through the artificial pancreas, paired with insulin pumps, can control blood sugar far better than humans can and acts in a minute by minute timescale. Heart pacemakers have electrodes that detect the heart's electrical activity and send data to generate electrical pulses to the heart to control and monitor the heartbeat at a second by the second timescale. In these closed-loop systems, the executed actions are entirely automated, and no human intervention or uncertainty is involved, so these cases can be considered deterministic actuators in PHN.

## **Partial Assistance**

Automation can assist humans in taking action. For example, building a Smart Home System helps modulate an individual's environment for circadian control, which ultimately assists in sleeping. Automated biological circadian rhythm control can be via connected lights, digital screen color variation, thermostats that mimic natural temperature fluctuations, and motorized curtains with smart white noise. While this may assist in circadian control and developing habits for sleep, the human needs to take action to ensure a consistent sleep schedule.

## **User Actions**

Humans are not always good at understanding complex instructions or remembering them at the right time and in the right situations. They also are easily influenced and distracted, which makes them non-deterministic actuators. Even when the user fully intends to reach a health goal, adhering to healthy routines can be difficult. The primary challenging task is to understand how to influence people at the right time and with the best medium to produce the desired action. For each individual, actions must be persuasive in the given context, practically feasible, and encourage participation. Richard Thaler won the 2017 Nobel Prize for understanding the psychological underpinnings and control of the decision-making process in humans [261]. Specifically, the following three high-impact contributions interweave directly with health navigation.

1. Limited rationality: Simplified decision-making focuses on the narrow impact of each decision rather than on its overall effect. Computing these complex effects can offload these tasks from the user.

2. Lack of self-control: A planner-doer model describes the relationship between long-term planning and short-term actions. Short-term temptations to make healthier lifestyle choices often fail. By quantifying the health cost or benefit in any situation, users can make more controlled and informed decisions.
3. Nudging and Probabilistic Nature of Inputs: By using subtle cues, users are pushed to make a confident decision without force, thus avoiding the pitfalls of self-control and limited rationality. Systems that employ these tactics can unobtrusively steer choices, leading to the goal state.

As PHN is able to model better the behavior of an individual, we can increase the probability of executing a given action using the considerations listed. The system explains why it suggests the nudge or action to users, and it offers alternate options to embolden the feeling of choice. Another avenue for human actuation will be to reduce the decision burden by eliminating the tediousness involved in being healthy. Even professional athletes do not measure every gram or determine every ingredient in their food due to the information and calculation burden. Translated into simple, step-by-step, context-aware guidance, users can better fulfill PHN actuation steps. PHN can start to apply control to the health state, either with fulfillment by a machine or with user participation, to move the health state toward the goal and close the loop.

Many of these actions are driven by daily decisions. Some of these decisions stem from hedonistic tendencies programmed in our biology from prehistoric times. Reward mechanisms were biologically encoded based on survival in a low resource environment with high uncertainty of the future. In the domain of food, these resources included factors such as caloric energy, water, and salt. Thus, humans have been optimized through evolution to accumulate these scarce resources in their body. Unfortunately the current era of civilization makes these types of problems largely irrelevant and much of our programmed biology is actually

maladaptive in this context. In order to move our health towards a positive direction, we need better actionable insight about our health than just our naturally encoded tendencies.

When a user is a final actuator in the system, there must be a probabilistic approach to knowing if a user will take a specific planned step. One way to increase the chance is to have high-quality recommendation systems.

### **6.7.3 Recommendation**

Recommendation systems play a role as a subset of cybernetic control. While the guidance is a full path to a goal, the recommendation systems increase the probability that a step in the process will take place for a particular state transition. A recommender system is also a filtering system that seeks to predict the rating or preference a user would give to an item. For example, services like Spotify have playlists generated based on user preferences. In the context of health, we use computed guidance and recommendation systems to guide an individual in lifestyle, nutrition, and other contextual components. Recommender systems have three components: 1) the user, 2) items to recommend, and 3) the context at a specific time and point that changes the relationship between the user and the items.

#### **User Profile**

The user health state is primarily a product of our everyday lifestyle actions, yet we have minimal access to recommendations on making everyday choices. Recommendations are the key to current content consumption and decisions. Cybernetic navigation principles that integrate health media sources can power dynamic recommendations to improve health decisions dramatically. Cybernetic components give real-time feedback on health status, while the navigational approach plots the health trajectory. These two principles coalesce

## Box 4: Health Multimedia: Lifestyle Recommendations Based on Diverse Observations

Oral Beave New Ideas ICMR'17, June 6-9, 2017, Bucharest, Romania

### Health Multimedia: Lifestyle Recommendations Based on Diverse Observations

Nitesh Nag  
University of California, Irvine  
Irvine, USA  
nagn@uci.edu

Vaibhav Pandey  
University of California, Irvine  
Irvine, USA  
vaibhap1@uci.edu

Ramesh Jain  
University of California, Irvine  
Irvine, USA  
rjain@uci.edu

**ABSTRACT**  
Managing health lays the core foundation to enabling quality life experiences. Modern multimedia research has enhanced the quality of experiences in fields such as entertainment, social media, and advertising; yet lags in the health domain. We are developing an approach to leverage multimedia systems for human health. Health is primarily a product of our everyday lifestyle actions, yet we have minimal health guidance on making everyday choices. Recommendations are the key to modern content consumption and decisions. Cybernetic navigation principles that integrate health media sources can power dynamic recommendations to dramatically improve our health decisions. Cybernetic components give real-time feedback on health status, while the navigational approach plots health trajectory. These two principles empower data to enable personalized, predictive, and precise health knowledge that can contextually disseminate the right actions to keep individuals on a path to wellness.

**KEYWORDS**  
cybernetic; navigation; quantified-self; objective-self; precision medicine; predictive health; prevention; event mining; context awareness; mobile health; persuasion; personalized

**ACM Reference format:**  
Nitesh Nag, Vaibhav Pandey, and Ramesh Jain. 2017. Health Multimedia: Lifestyle Recommendations Based on Diverse Observations. In *Proceedings of ICMR '17*, Bucharest, Romania, June 06-09, 2017, 8 pages. <https://doi.org/10.1145/3078971.3080455>

**1 INTRODUCTION**  
Inaccessibility and delayed dissemination of medical knowledge for lifestyle guidance persists in all health care systems, while global health burdens continue to rapidly rise. Health care is yearning for a major transformation, from reactive and hospital-centered to proactive prevention, from doctor-centered to patient-centered, from disease-centered to wellness promotion, and a shifting focus from temporary fixes to long-term solutions [25]. Commonly, physicians focus primarily on medical methods to manage health when a patient becomes ill. True health outcomes result from actions taken in every moment and place, not just medical intervention during sickness. Future advancements in health must continuously sense individual needs and regularly provide the relevant resources so corrective actions ensure health stability. Health is essentially a product of our genome and lifestyle [28]. Lifestyle is the primary controllable aspect of our health. Unfortunately, lifestyle associated diseases continue to lead the rapid rise of chronic diseases across the globe. Monitoring lifestyle factors has been difficult, and computational power with effective methods to address the needs of each individual have been limited. Trying to produce changes in routine lifestyle habits is also a tremendous psychological hurdle. Media platforms such as food analysis from images or accelerometer based activity sensors have started collecting data about lifestyle factors. These isolated efforts forge the integration of other media sources and back synchronization across time and space. Imagine how Google Maps on mobile phones alerts the driver of an upcoming turn. This is the product of real-time

ICMR 2017  
June 6-9, Bucharest, Romania

### Personal Health Space + Navigation

No Integrated Expert Knowledge  
No Quantified Effects  
No Navigation  
No Context

ACM International Conference  
on Multimedia Retrieval

### Traditional Health Search Space

No Integrated Expert Knowledge  
No Quantified Effects  
No Navigation  
No Context

Figure 1: Increased dimensionality (both time and space) of personalized health space can provide actionable navigation towards health goals. As a simple example, although jogging may have a higher effort at the moment (hence the uphill initial slope), it informs the user how this action will reach the goal node. By having the pizza, it may be easier in the moment, but the user will have to climb out of a sink in the map to reach their goal.

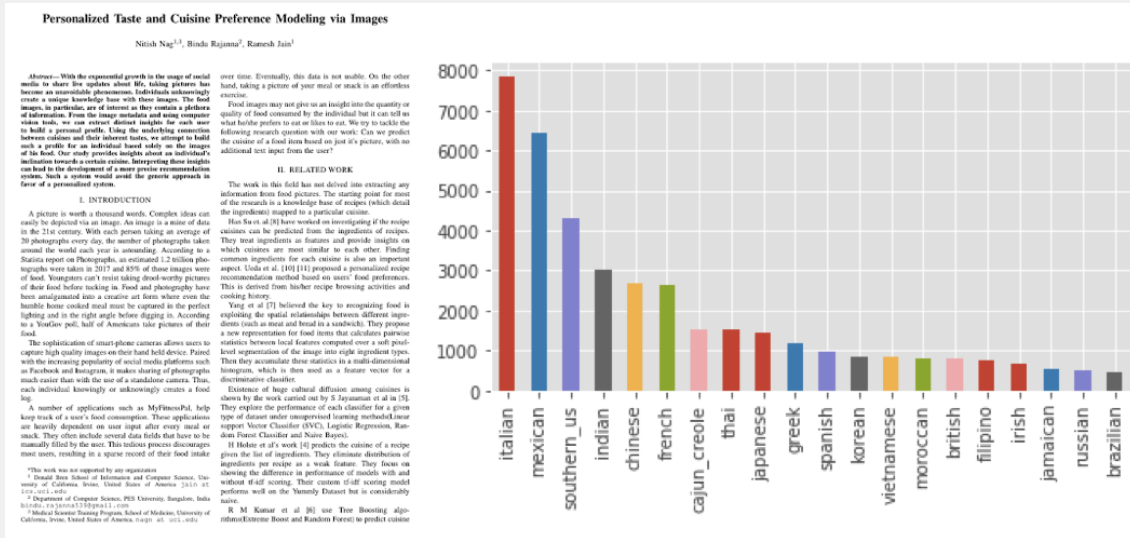
Cybernetic components give real-time feedback and control on immediate health status, while the navigational approach plots health trajectory for long-term planning. *Published in ACM International Conference on Multimedia Retrieval, June 06-09, 2017, Bucharest, Romania.* [188]

- Cybernetic loops and navigation systems need to understand user health state.
- Integration of health media sources can power dynamic recommendations to dramatically improve our health decisions.

data to enable personalized, predictive, and precise health knowledge that can contextually disseminate the right actions to keep individuals on their path to wellness [188].

Unique individual *preferences* are critical for determining the user profile, as preferences determine whether or not a user will proceed with the recommendation. Understanding a user's model allows us to build a more precise and personalized recommendation system. We have investigated the relationship between cuisines and their inherent taste by collecting data on an individual's food photos. We used image metadata and computer vision tools to extract distinct insights from each user to build their food profile (unpublished).

## Box 5: Personalized Taste and Cuisine Preference Modeling via Images



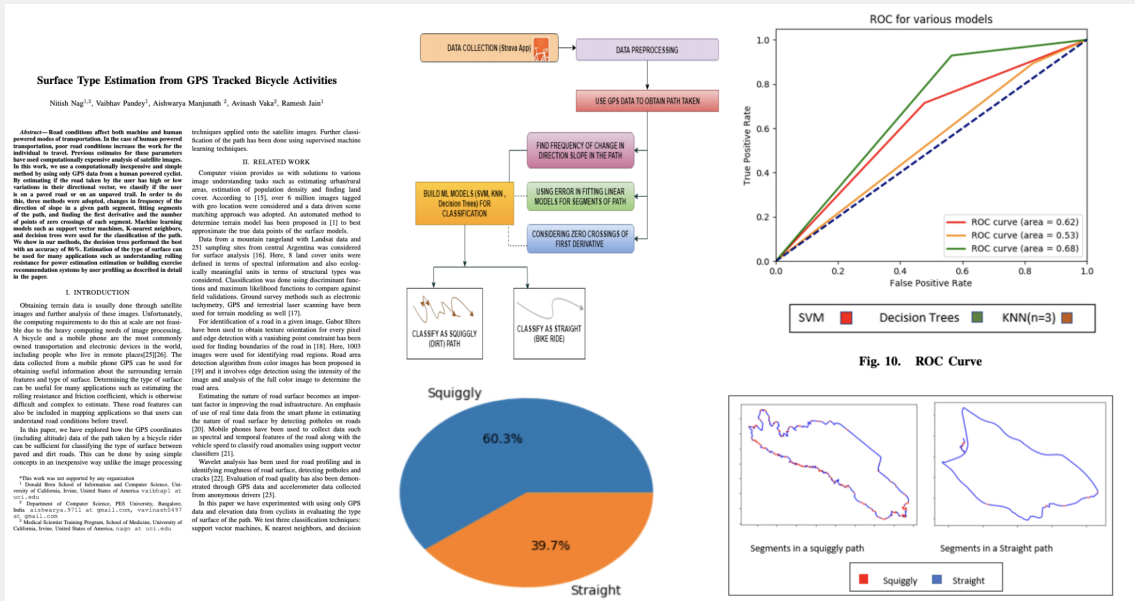
From the image metadata and using computer vision tools, we can extract distinct insights for each user to build a personal taste and cuisine profile. *Arxiv, 2020*

- Development of rank system for cuisines based on user's preferences.
- Weighting preferences based on event mining nutrition events.
- Identification of ingredients commonly preferred, allowing the system to generate recommendations that are novel and exciting to the user.

A combination of preferences for physical activity and conditions can also be applied in a recommender system to give an individual the most optimal plan for their exercise. For example, we have used a method for using GPS data from a human-powered cyclist to classify if the user is on a paved road or unpaved trail. Because road cyclists may prefer smooth roads, versus mountain bike riders who prefer dirt roads, we can begin to understand user preferences for bicycle travel paths. We can draw other relevant insights from the data we collect through Strava. We can collect social media platforms to determine the weather preference of the user, the kind of roads they prefer to cycle on, the time during which they regularly perform the activity, and much more. These can be combined in order to recommend suitable activities to improve the health state [191].



## Box 6: Surface Type Estimation from GPS Tracked Bicycle Activities



By estimating if the road taken by the user has high or low variations in their directional vector, we classify if the user is on a paved road or on an unpaved trail without accessing expensive satellite imagery or databases. *Arxiv, September 2018 [191]*

- Providing exercise recommendations by profiling the users path choices.
- Understanding rolling resistance for power estimation for cardiovascular applications in VO2max estimation.
- In the above figure, the orange slice highlights smooth roads, and the blue slice shows undulating terrain as part of the model regarding the test subjects preferences.

Just as *motivation* is critical for goal setting, it is also vital for one's willingness to follow through with a recommendation. Strong recommendations need to be set based on the intrinsic preferences of the user. Combining the right triggers with context, we can maximize the effectiveness of sustainable positive health recommendations. Over time, the recommender system should adapt to the user and identify triggers that are needed to support their motivations. For example, exercising every day should become habitual rather than a constant suggestion from the recommendation engine. In addition to one's health, *user enjoyment* is also a critical component of determining the user profile. Enjoyment leads to a higher quality of living, so it is critical to incorporate the best preferences.

## Box 7: Flavour Enhanced Food Recommendations

**Flavour Enhanced Food Recommendation**

<p><b>Nitish Nag</b> nags@uci.edu University of California, Irvine Irvine, California, United States of America 92697</p> <p><b>Kushal Samir Mehta</b> kushalmehta13@gmail.com FES University 100 Feet Ring Road Bangalore, Karnataka, India 560085</p> <p><b>Aditya Bharadwaj</b> aditya119@gmail.com FES University 100 Feet Ring Road Bangalore, Karnataka, India 560085</p>	<p><b>Aditya Narendra Rao</b> 13aditya13@gmail.com FES University 100 Feet Ring Road Bangalore, Karnataka, India 560085</p> <p><b>Nishant Bhattacharya</b> nishant51@gmail.com FES University 100 Feet Ring Road Bangalore, Karnataka, India 560085</p> <p><b>Dinkar Sitaram</b> dinkarsitaram@gmail.com FES University 100 Feet Ring Road Bangalore, Karnataka, India 560085</p>	<p><b>Akash Kulkarni</b> akash.kulkarni@gmail.com FES University 100 Feet Ring Road Bangalore, Karnataka, India 560085</p> <p><b>Pratul Ramikumar</b> pratul.ramkumar@gmail.com FES University 100 Feet Ring Road Bangalore, Karnataka, India 560085</p> <p><b>Ramesh Jain</b> rajuresh@uci.edu University of California, Irvine Irvine, California, United States of America 92697</p>
---	---	---

**ABSTRACT**  
We propose a mechanism to use the features of flavour to enhance the quality of food recommendations. An empirical method to determine the flavour of food is incorporated into a recommendation engine based on major gustatory nerves. Such a system has advantages of suggesting food items that the user is more likely to enjoy based upon matching with their flavour profile through use of the taste biological domain knowledge. This preliminary intends to spark more robust mechanisms by which flavour of food is taken into consideration as a major feature set into food recommendation systems. Our long term vision is to integrate this with health factors to recommend healthy and tasty food to users to enhance quality of life.

**CCS CONCEPTS**  
-Information systems → Collaborative filtering; Personalization; Collaborative filtering; Recommender systems; Personalization.

**KEYWORDS**  
food computing, personal health navigation, recommendation systems, flavor, flavor, taste, gustatory media

**ACM Reference Format:**  
Nitish Nag, Aditya Narendra Rao, Akash Kulkarni, Kushal Samir Mehta, Nishant Bhattacharya, Pratul Ramikumar, Aditya Bharadwaj, Dinkar Sitaram, and Ramesh Jain. 2019. Flavour Enhanced Food Recommendation. In *Proceedings of 5th International Workshop on Multimedia Assisted Dietary Management*. New York, NY, 2019 (MADiMa '19), 9 pages. DOI: 10.1145/3347463.3357338

**1 INTRODUCTION**  
Food is essential for human life. Beyond sustaining health by providing nourishment, it is also an integral part of human culture and quality of life. Modern information about nutrition and food fosters development of food computing methods [5]. The notion of food computing involves obtaining food data and identifying areas where it can be applied effectively, such as health, food science, culinary art, and human behavior. Food computing collects data from multiple sources and involves tasks such as perception, recognition, retrieval, recommendation, prediction and monitoring of

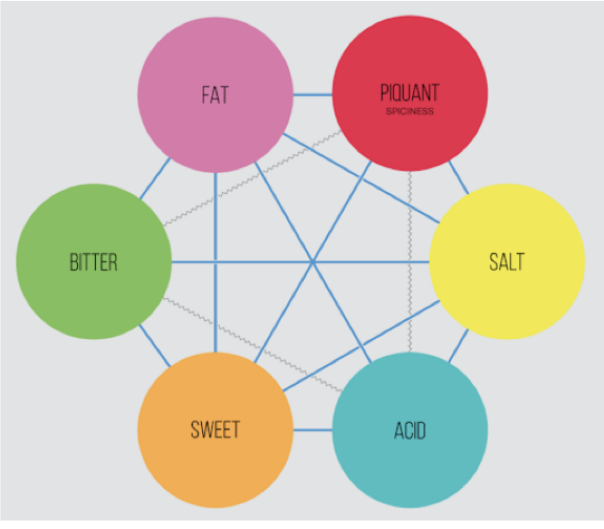


Figure 1: Mammalian taste receptors are the basic unit of biological taste. Understanding the features which impact these receptors is the fundamental start to profiling users and food in their culinary preferences. As biological discoveries in protein structure and function of taste are uncovered, the knowledge must be incorporated into the recommendation systems and health navigation paradigms [14].

Proceedings of 5th International Workshop on Multimedia Assisted Dietary Management, New York, October 21, 2019 (MADiMa '19), 9 pages. DOI: 10.1145/3347463.3357338

An empirical method to determine the flavour of food is incorporated into a recommendation engine based on major gustatory nerve signals. *MADiMa '19: Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management, October 2019* [195]

- Readily available data about nutrition can help in identifying item flavor.
- Our long term vision is to integrate health factors to recommend healthy and tasty food to users for a high quality dining experience.

## Items

We must capture in detail the items or characterization of specific objects that change the health state. By knowing the user profile, we can then match a user model with items.

For food recommendations, we propose a mechanism to determine the taste profile of items. We incorporate into a recommendation engine an empirical method to determine the flavor of food. Such a system has advantages of suggesting food items that the user is more likely to enjoy based upon matching with their flavor profile through the use of the taste biological domain knowledge. This preliminary intends to spark more robust mechanisms by which flavor of food can apply to recommendation systems [195].

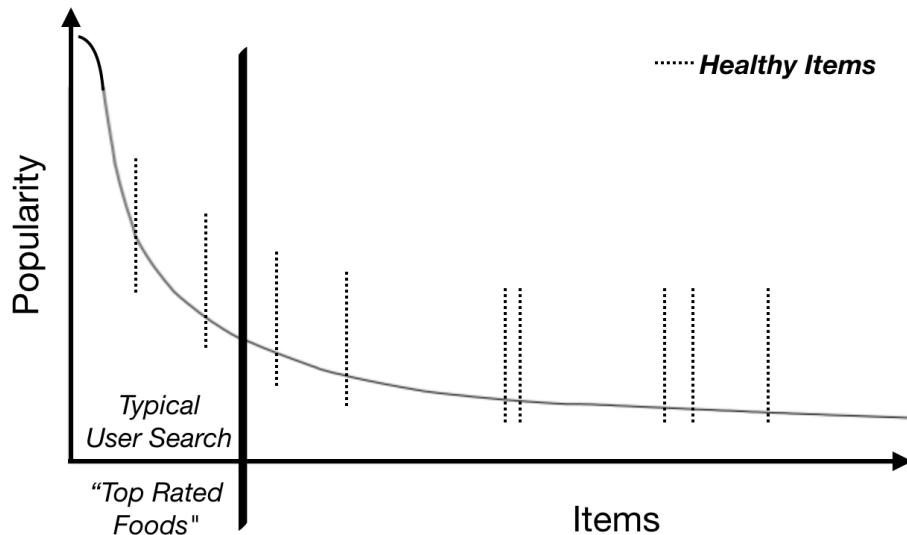


Figure 6.9: **The Long Tail Problem.** Usually users only see items that are top rated by other users. This leaves items that are highly relevant for the user far down in the search results, such as items that are healthy.

Along with the taste profile, we propose that food items should also have a health profile associated. Recommending items based on unique preferences comes challenges. For example, the Long Tail Problem highlights items that are not popular but highly relevant to the user, shown in Figure 6.9 [214]. However, by understanding the health status of the user, we can match health-relevant items that a user would not usually browse through. This matching process allows the user to discover healthy items that are relevant but would have been difficult to find through popularity based rating systems. Thus, a recommender system should provide relevantly matched items to reduce browsing burden. The goal of this system is to capture the healthy items in the long tail and match them to the personal needs of a user based on, for example, sodium needs. [190, 194, 189].

People want to know what foods they can consume at a specific place, alongside the taste, ingredients, allergens, nutrients, and visual and other sensory characteristics. Combining both taste profiles and health profiles for food items give us a basic *food atlas* [302]. This *food atlas* is a location-based database that has a live data base of nutritional items of food

## Box 8: Pocket Dietitian: Automated Healthy Dish Recommendations by Location

**Pocket Dietitian: Automated Healthy Dish Recommendations by Location**

Nitah Nag<sup>26)</sup>, Vaidya Pandey, Abhinav Sharma, Jonathan Lam, Ranyf Wang, and Ramesh Jain  
University of California, Irvine, USA  
nag@uci.edu

**Abstract.** A root cause of chronic disease is a lack of timely informed decision power in everyday lifestyle choices, such as in diets. Users are unable to clearly delineate and demand healthy food in a quantitative manner. To scale the benefit of health nutrition coaching in broad real-world scenarios, we need a technological solution that is constantly able to interpret nutrition information. We ingest nutritional facts about products to efficiently calculate which items are healthiest. We deliver these results to users based on their location context. Our ranking algorithm outperforms major nutrition score metrics, and is more consistent than human dietitians in real world scenarios. Most importantly, our system gives the user a rapid way to connect with healthy food in their vicinity, reducing the barriers to a healthy diet.

**Keywords:** Cybernetics · Precision medicine · Disease prevention  
Content awareness · Mobile health · Personalized · Nutrition  
Expert systems

### 1 Introduction

Health is essentially a product of our genome and lifestyle [12]. Dietary choices are a major component of lifestyle and can have great impact on an individual's long term health outcomes.

US federal nutrition guidelines (Fig. 1) are hard to translate into everyday life decisions. Although the knowledge exists, it is not in an actionable form. For example, typically patients with diabetes who are supervised by nutrition experts meet once every three months. Even if a nutritionist is available to guide a client, they don't usually have all information related to the appropriate nutrition in immediate context of the individual.

Real-time multimedia technologies will play a major role in powering recommendations to solve health issues. Leading medical professionals have advocated for a deeper integration of technology into health care [12]. Patients make better lifestyle choices that would combat diabetes if given guidance [11]. Ultimately, transforming data and knowledge to actionable lifestyle choices is the

© Springer International Publishing AG 2017  
R. Shrivastava et al. (Eds.): ICAP 2017 International Workshops, LNCS 10590, pp. 444–452, 2017.  
http://dx.doi.org/10.1007/978-3-319-70712-4\_33

Pocket dietitian refers to the application built that can rapidly locate healthy dishes on a mobile application. 2017 [194]

- Algorithms are needed to constantly interpret nutrition information.
- We have calculated nutritional facts about products to determine which products are healthiest for several types of patients.
- Our system is designed to give the user a rapid way to connect with healthy food in their vicinity, reducing barriers to taking a step in their health actions.

along with all parameters about each physical entity within inventory. The information regarding each food type can be globally available and creates standardization for food computing purposes. This idea has just recently been launched to the public as an open source project. “The World Food Atlas project is an open source effort to prepare a live atlas of specific food at a specific geographic location. World Food Atlas (WFA) aspires to provide information relevant to all dishes and food items for health and enjoyment that foods offers us, wherever you are in the world. We want to make it widely available so it can be converted to actionable information for a person in specific context” [302].

## Context

The next step is understanding the context (location and resources) for a given individual situation, which then determines the relationship between the user profile and items. We can continue to improve recommender systems with in-depth context provided. The utility function is the primary ranking method to predict what a given person would explicitly prefer. There are two layers of user and context modeling, coupled with traditional recommender system approaches. One is an endogenous context, which integrates data to estimate the physiologic or innate needs of the user (health state) or preferences, and second is the exogenous context, which incorporates factors outside of the person's body (i.e., points of interest, geospatial resources, social context) [190].

Through multi-modal sensor data integration applied to domain-specific utility functions, filters, and re-ranking weights, we can use context in recommendations. We showcase this concept through a nutrition guidance system focused on controlling sodium intake at a personalized level, dramatically improving upon the fixed recommendations. For a particular user, identifying contexts through continuous sensing is the best execution strategy. For one's nutrition, the system will need to decide the best meal choice to move the user's health state closer to his or her goal state based on situational constraints. This decision will require knowledge of all available meals, as well as their locations, hours, nutrition facts, and ingredients along with the personal sensor data modeling the physiologic needs.

A challenge that arises is the user's inability to delineate and demand healthy food in a quantitative manner. We have developed a technological solution that is continuously able to interpret nutrition information, ingest nutritional facts about products to efficiently calculate which items are healthiest, and deliver these results to users based on their location context. Our system gives the user a fast way to connect with healthy food in their vicinity, reducing the barriers to a healthy diet. Addressing the correct granularity of recommendations, such

## Box 9: Endogenous and Exogenous Multi-Modal Layers in Context Aware Recommendation Systems for Health

**ENDOGENOUS AND EXOGENOUS MULTI-MODAL LAYERS IN CONTEXT AWARE RECOMMENDATION SYSTEMS FOR HEALTH**  
Nitesh Nag, Vaibhav Pande, Ramesh Jain  
Department of Computer Science, University of California, Irvine - California, USA


**ABSTRACT**  
People care more about the solutions to their problems rather than data alone. Inherently, this means using data to generate a list of recommendations for a given situation. The rapid growth of multi-modal wearables and sensors have not made this jump effectively in the domain of health. Modern user content consumption and decision making in both cyber (e.g. entertainment, news) and physical (e.g. food, shopping) spaces rely heavily on targeted personalized recommender systems. The utility function is the primary ranking method to predict what a given person would explicitly prefer. In this work we describe two unique layers of user and context modeling that can be coupled to traditional recommender system approaches. The exogenous layer incorporates factors outside of the person's body (e.g. location, weather, social context), while the endogenous layer integrates data to estimate the physiologic or innate needs of the user. This is accomplished through multi-modal sensor data integration applied to domain-specific utility functions, filters and re-ranking weights. We showcase this concept through a nutrition guidance system focused on controlling sodium intake at a personalized level, dramatically improving upon the fixed recommendations.

**Index Terms**—Context Aware, Recommender Systems, Chemistry, Utility Functions, Expert Knowledge, Personal Sensors, Health, Multimodal Data

**1. INTRODUCTION**  
Recommendation systems are becoming an increasingly important part in various domains of daily life. The large sets from which users must make decisions from can be overwhelming from too much choice. In areas such as online shopping, entertainment, and information search there has been considerable effort to streamline choice through recommendations. These traditional recommendation systems consider two main components, the user and the items. In more recent work, context has become the third valuable component. Context is very complex and difficult to characterize, but usually has an effect to alter the recommendations to fit a given user situation. Building real-time user models from context has been a key component of advanced recommender

Generally, the contextual information of a given user is divided into 3 classes of fully observable, partially observable, or unobservable [1]. Increasing data to determine context is divided into another dimension of static or dynamic data. Wearable devices, sensors, multi-media, social networks, and many more multi-modal data streams carry information that can better characterize the user model to enhance recommendations in both static and dynamic data domains. Applying these multi-modal data sources for health applications remains a challenge. The growth momentum of wearable devices that monitor health metrics via sensors continues to promise better health for users. How will the data generated by these sensors become useful in daily life? Ideally an important application would be to improve personalized recommendations pertaining to health decisions. In this work we extend a new dimension to contextual data by separating the distinction between information in the environmental situation of a user (exogenous to the user) and the inherent biological needs of a user under the skin (endogenous to a user). After an explanation of the concept, we show this in real-time prediction of sodium needs for food recommendations.

**2. RELATED WORK**  
Context-Aware Recommendation Systems (CARS) in various domains have been used to improve recommendation performance increasingly over the past decade. Basic techniques of purpose filtering and context modeling have been well described in reviews and texts [1]. Furthermore, the contextual data has been usually split into dynamic and static components that can be observed at various fidelity levels (X and Y



**Fig. 1:** Increasing aggregation of multi-modal user data streams can more accurately predict the health status and needs of a user.

### Key Points August 2018 [190]

- The utility function is the primary ranking method to predict a person's preferences.
- We review both endogenous and exogenous layers determine user preferences, using multi-modal sensor data integration applied to domain-specific utility functions, filters, and re-ranking weights.
- Our approach is showcased with a nutrition guidance system at a personalized level.

as meals, is essential for simple decision making. We have made a decision support system using multi-modal data relying on timely, contextually aware, personalized data to find local restaurant dishes to satisfy a user's needs. Algorithms in this system take nutritional facts regarding products, efficiently calculate which items are healthiest, then re-rank and filter results to users based on their personalized health data streams and environmental context. The primary goal of lowering barriers to a healthy personalized choice drives our recommendation engine [189].

## Box 10: Live Personalized Nutrition Recommendation Engine

Session: Understanding and Promoting Personal Health  
MMHealth'17, October 23, 2017, Mountain View, CA, USA

### Live Personalized Nutrition Recommendation Engine

Nitish Nag  
University of California, Irvine  
Irvine, California  
nagn@uci.edu

Vaibhav Pandey  
University of California, Irvine  
Irvine, California  
vaibhapt@uci.edu

Ramesh Jain  
University of California, Irvine  
Irvine, California  
jain@ics.uci.edu

**ABSTRACT**  
Dietary choices are the primary determinants of prominent diseases such as diabetes, heart disease, and obesity. Human health care providers, such as dietitians, cannot be at the side of every user at all times to manually guide them towards optimal choices. Automated adaptive guidance fused with expert knowledge can use multimedia data to technologically scale health guidance without human intervention. Addressing the correct granularity of recommendations (in this case meal dishes) is essential for effortless decision making. Thus we make a decision support system using multi-modal data relying on timely, contextually aware, personalized data to find local restaurant dishes to satisfy a user's needs. Algorithms in this system take nutritional facts regarding products, efficiently calculate which items are healthier, then re-rank and filter results to users based on their personalized health data streams and environmental context. Our recommendation engine is driven by the primary goal of lowering the barriers to a personalized healthy choice when eating out, by distilling diabol suggestions to a single contextually aware and easily understood score.

**CCS CONCEPTS**  
• Information systems → Mobile information processing systems  
• Human-centered computing → Ubiquitous computing  
• Mobile computing, Ambient intelligence, Smartphones.

**KEYWORDS**  
Personalized Health, Cybernetics, Precision Medicine, Nutrition Science, Multi-modal data streams, Human Modeling, Resource-Needs Matching, Recommendation Engine

**1 INTRODUCTION**  
Mobile phone sensor technologies have created a vast amount of quantitative and qualitative multimedia regarding personal health. The next step in advanced health systems will be to effectively utilize this data to provide guidance for users. Since these data streams have different granularity levels, integration in the context of decision to make digital or better options at all or part of the work or personal or classroom use is a grand challenge for providers that experts are not made or distributed for public or commercial adoption and they require their own and the full attention on the page. Copyright for this work is reserved by others than the authors but by license, allowing with credit is permitted. In any other case, or to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a license from the publisher.

MMHealth'17, October 23, 2017, Mountain View, CA, USA  
© 2017 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.  
ACM ISBN 978-1-4503-5130-1/17/10...\$15.00  
https://doi.org/10.1145/3120253.312030

**Figure 1** User data and context is expertly matched with local physical resources of an application in an approach to extract utility from the data. Synchronized data streams can power recommendations for users to effectively manage their health at all times, locations, and contexts. We believe that recommendation systems, such as in Figure 1, that combine user personal information and context, along with local physical resources will drive the future of health behaviors. We dive into this principle by guiding users towards healthy food options that are personalized for their biological and contextual parameters. Improving health outcomes from lifestyle should be a normal part of life in every moment and place, not just medical intervention during illness. Health is essentially a product of our genome and lifestyle [24, 25] with lifestyle being the primary controllable aspect of our health. Recent computational technology has rapidly advanced quantifying and personalizing services such as advertising, entertainment, and shopping. These advances have put customers at the center of power in driving commercial success, such as through reviews on Amazon products or likes on Facebook. Unfortunately consumers still lack personalized quantitative power in decision making regarding their health. Diet is the most dangerous aspect of health risk factors in most western countries [2]. Patients make better lifestyle choices that would combat diabetes if given guidance, and many health conscious consumers demand healthy food [23]. Human health service providers want their patients to access expert information at all times yet they cannot be at the patient side at all times. This problem exists due to the obvious difficulty of scaling human dissemination of knowledge, like in hospitals. Translating this expert knowledge into everyday life decisions needs to be in a live actionable form. For example, typically patients with diabetes who are supervised by nutrition experts meet once every three months. This is an inappropriate amount of support for a patient, who is choosing what to eat multiple times a day. The patient should have guidance at all times. Even if a nutritionist is available to guide a client, they don't usually have all information related to the appropriate nutrition in immediate context of the individual. This is

### Key Points August 2018 [189]

- Automated adaptive guidance combined with expert knowledge can use multimedia data to technologically scale health guidance.
- We introduce a decision support recommendation system for finding local dishes satisfying a user's personalized needs.
- The recommendation engine aims to reduce barriers to healthy eating by suggesting dishes with a health score.

## Expert Guidance

There may be guidance mechanisms beyond the scope of computing. Human experts may be utilized for recommending change to an individual's health state in person or through various forms of telecommunication. These forms of guidance can be synchronous in time (real-time), or they can be asynchronous through various forms of communication and multi-media.

Through multi-modal sensor data integration applied to domain-specific utility functions, filters, and re-ranking weights, we can use context in recommendations. We showcase this concept through a nutrition guidance system focused on controlling sodium intake at a

personalized level, dramatically improving upon the fixed recommendations. For a particular user, identifying contexts through continuous sensing is the best execution strategy. For one's nutrition, the system will need to decide the best meal choice to move the user's health state closer to his or her goal state based on situational constraints. This decision will require knowledge of all available meals, as well as their locations, hours, nutrition facts, and ingredients along with the personal sensor data modeling the physiologic needs.

The end purpose of the PHSS is to ensure that the individual arrives closer to their end goal destination, which may mean high-quality intermediate events that determine one's experience in the navigation process. The quality of the experience provided is, therefore, one of the requirements of the system. In order to have a long-term engagement success with the navigation system, we must understand the individual's motivations, preferences, context, and other factors that may make the individual more engaged with the system.

The personal navigation system to health allows for concrete actions to be taken for the future performance of one's health. Future systems can leverage modern computing power with rich multi-modal data to understand movements on the health state space continuously and to contextually guide users toward wellness. The nexus of biomedical knowledge, sensor technology, computing power, mobile networks, and artificial intelligence will fuel significant growth in PHN systems.



# Chapter 7

## Conclusions and Future Work

If you're walking down the right path  
and you're willing to keep walking,  
eventually you'll make progress.

---

*Barack Obama*

Advancing the understanding of our health state is essential for improving human health. Throughout time, humans have primarily taken an episodic approach to interact with health. The next evolution in health will depend upon a continuous and high-resolution understanding of our health state to better control the future of our well-being.

This thesis introduces an open framework and concept to use a diverse array of data streams and knowledge to enhance the health state estimation of an individual. The product is graph modeled a digital twin that parallels the user's physical life. The first main research contribution provides a consolidation of defining health and it's role within the influence of control systems (Chapter 2). The second research contribution builds a computational framework to represent this health. The representation is based on connecting the user's life events, biology, and general utilities through a graph network block approach (Chapter

4). This work has the dual coupling of bottom up data-driven modeling and top-down domain knowledge. By consolidating systems into smaller blocks that can continuously be updated with new data or knowledge, the graph network can flexibly evolve and improve its representation of the user's health state. Demonstration of this framework with diverse high-resolution personal data is conducted with a focus on bioenergetics, cardiovascular disease, and performance (Chapter 5). Given this approach to health state estimation, we can enable personal health navigation systems that continuously guide an individual to a desired health state using a combination of planning intelligence and cybernetic control (Chapter 6). We hope individuals may be more motivated to track their health state, especially if it will be used in professional clinical decision making or influencing their daily actions through such navigational systems. Instant feedback on the health state can give insight on how lifestyle modification, medicine, environment and other inputs are causally changing the person's overall quality of life. The highest motivation of this work is to catalyze the research community to advance these ideas of health state estimation and personal health navigation. There remain significant challenges and opportunities in this endeavour. An non-exhaustive set of these challenges is given below.

## **7.1 Technical Challenges**

Improvements in various technical areas are necessary to bring further fruition of this framework to individuals. Many of the challenges need to be understood in depth.

### **7.1.1 Building Rich Knowledge Atlases**

The work presented here focuses only on one domain of cardiovascular and athletic performance. Experts must also address areas beyond cardiovascular health. To establish the

best cold-start graph representation of an individual, we need accessible domain knowledge. This accessibility needs to cooperate with both human and computational interfaces. There have been substantial efforts in both of these interfaces in general domains as well as health-specific fields. Examples of knowledge bodies that are widely used today include Wikipedia, PubMed, and Google Scholar [293, 198, 94]. For physicians and healthcare providers, the most commonly used atlas of knowledge is UptoDate [268]. Most of these knowledge atlases store information in static natural language, which is very useful to interface with humans. The computational interface with these knowledge sets will require robust Natural Language Processing techniques to produce a virtual representation of knowledge. Efforts have been made in specific domains to encode the knowledge in a more computable format. MDCalc is a website that allows domain experts to define formulas by which other users or applications can run domain knowledge-based calculations for precise biological estimations of physiology, risk, or medication dosing [177]. One step further for computational representation is the entire field of knowledge graphs. Open source platforms such as Cytoscape and HetNets have shown promise in taking the current domain knowledge and producing a connected system that can incorporate within further applications [244]. The ultimate ideology behind the knowledge atlases would be a dynamic reference database that applications can use to access the latest scientific understanding to build systems of navigation, estimation, and other services. These atlases would require the coordinated representation of both highly complex data and human knowledge structure and taxonomy.

### **7.1.2 Hardware**

Hardware improvements for health state estimation largely depend upon sensors that can unobtrusively gather data about events, biological metrics, and utility measures. Event sensors that can capture more attributes in detail can give insight on how the event is impacting the health state, and exactly what aspects of the event are relevant. Improved

biological sensors are needed to know what the observed values of biological health state components are to better estimate an all encompassing health state. State estimation may also become computationally demanding depending on the case. There remains a need to use computing hardware to include local, edge, and cloud based approaches for high performance health estimation. In various contexts, there will be a need for both online and offline methods to digest and compute the information. The online approaches need to synchronize at appropriate times to update models in the offline approaches.

### **7.1.3 Computing**

On the mathematical side, computing through modern modeling techniques can increase the computing efficiency and capability to handle higher loads of computational complexity with more efficiency. Black box artificial intelligence techniques to improve data driven relationships need to transition into explainable systems. If any health estimation is done on a battery powered device, energy management also becomes a key issue for constant updates of the health state. Given the high cost of accuracy and complexity, there needs to be better optimization of computing resources.

### **7.1.4 Accuracy and Precision**

Precision in estimation will be best suited for measuring relative change in an individual, as it refers to the closeness of measurements to one another. Accuracy refers to the closeness in measurements of to a specific value, giving a definite health measurement. Estimation in its initial iteration may not be accurate, but it is assumed to improve over time with refining of equations, algorithms, feature extraction methods, learning methods, as well as with improvements in hardware technology. We use basic methods in this paper as a starting point for further advancement with modern learning and predictive techniques. Because

of the relative large range in accuracy, wearable devices and similar low cost sensors are promoted for use as a screening tool to identify if a user is at risk or has a change in their health state [129]. Clinical gold standard testing, a more expensive alternative, may be used to confirm the true state of the user if the state estimation passes a certain screening threshold. More work will need to be done to ensure the robustness of the estimations are good enough for key decision-making, such as ensuring validated clinical and biomedical research.

### **7.1.5 N=1 Statistics**

Baseline comparisons are also difficult when studying individual subjects, and will require statistical methods for n of 1 studies [176, 46]. Even while referencing established clinical science research we still have no way to validate ground truth about an individual until decades into the future when a person becomes verified in their disease or die. Therefore, comparisons are not the best way to experimentally validate the estimation model. Performance comparisons for this type of experiment will need to be validated through large scale data collection and monitoring in prospective studies as mentioned in related works. Parallel models may compete for the best representation of the individual through an adversarial and recurrent network that is temporally dynamic. Ultimately, high quality validation techniques using only a single individual need to be advanced in order to definitively improve and verify the state estimation quality.

### **7.1.6 Event Mining**

Efficiently extracting knowledge from events and log data for an individual remains a challenge. Tracking of individual data may be limited, and barriers to efficient extraction of useful information exist. We see problems with the accessibility of data, more adequate

sensors developed for determining events, and event descriptions. There are also issues with semantic variability in the real-world, including multilevel event semantics, implicit semantics facets, and the influence of context [308]. The field of behavioral modeling can benefit greatly by having more data-driven personal models that can be fed into a navigation system to provide more engaging routing for the individual. Event mining can also help understand causal chain of events and discover edges in the graph that are unknown.

## 7.2 Real-World Adoption

For real-world adoption of this work into tangible services for society and individuals, we must consider various usability, economics, legal, ethical, and protection issues. For health state estimation to occur, users must have some reason to engage in collecting and sharing data about themselves. This estimation must be made useful through services back to the individual that is part of the broader health economy. All the sharing of information must also occur safely. Privacy refers to the rights to control personal information by the individual and how authorized agents use it. Security, on the other hand, refers to how personal information will be protected and defended against unauthorized and malicious access. Since health data inherently contains sensitive information about an individual, we must address these at both individual and societal levels.

### 7.2.1 Engagement

One of the greatest struggles in the health field is individual engagement with their health. In the current healthcare model, individuals predominantly interact with their health when there is a problem or urgent need (a reactive approach). Individuals must see value in state estimation and proactively engage in collecting their health data. We see a few challenges:

1) Health state estimation relies heavily on computing mechanisms and may provide complexities for humans to understand and use effectively. Humans need to be technologically savvy in interacting with sensors that are collecting their health information. Sensors must provide ease of use, and individuals must trust that the sensors are not harming them in any capacity.

2) It is not very easy to visualize the health state as it is so robust that a person may not be able to comprehend all the various components and relationships. There will be challenges with simplifying information to the semantic understanding of the individual.

3) Currently, individuals interact with their health through service providers (i.e., doctors, health professionals, insurance companies). Because of health's interconnectedness with external systems, service providers must be involved in the health estimation in order to help the individual in both understandings and taking steps towards better health.

4) Individuals may be disinclined to engaging with their health. It is necessary to consider behavioral psychology to see where there are gaps in engaging with health and having motivation for improving future health. There is the problem of temporal discounting, in which "value of a reward decreases as the time delay until its receipt increases" [91]. Humans are less inclined to make changes in the present day if the outcomes are far in the future.

### **7.2.2 User Autonomy, Authority, and Ownership**

As more personal health data is acquired to estimate the health state, there will arise issues with data ownership and utilization by external sources, introducing legal and ethical barriers and ambiguity in who will ultimately make use of the data. The privacy and security issues regarding sensitive health data can be at an individual level, organizational level, or national level for civilian protection. Emerging threats and vulnerabilities will continue to grow as

data becomes expansive, and it will be a challenge to balance innovation in health estimation with the regulation of health data.

### **7.2.3 Privacy**

A challenge with user health data is ownership of the information and the ability to release this data to external agents such as healthcare service providers. As the information is sensitive and robust, there must be controls and policies in place to ensure ownership of data by the individual and appropriately used. There are regulations in place today that help with data privacy issues such as the Health Insurance Portability and Accountability Act (HIPAA) passed by Congress in 1996. This law created rules to address disclosure of “protected health information” by organizations and created privacy rights standards for individuals to control the use of their health information [114]. Similarly, The European Union has established the General Data Protection Regulation (GDPR) in May 2018 to strengthen individual privacy rights by clarifying rules for companies and public bodies in the digital market [79].

We must carefully address information sharing as there can be external threats and harm to the individual with improper data management.

### **7.2.4 Security**

Security is a significant concern for personal health data as from breaches by external threats, hackers and ransomware are highly prevalent. Mishandling of information can stem from insider information misuse and lack of security for mobile devices. Healthcare providers may fail to evaluate risk when working with outside vendors. There becomes an issue with individuals sharing information voluntarily on the web or via apps, in which case attackers



can take disparate information pieces from many sources, including social networks, public records, and search engines [156].

Some methods for addressing security are being taken, such as blockchains used for storing and sharing information that is secure because of its transparency. Blockchain systems are more commonly used in healthcare to protect information but are susceptible to attacks [126]. Encryption methods take the data and convert it to encoded text. Data is unreadable unless an individual has the necessary key or code to decrypt it.

With the sensitivity of the information at hand, we must ensure further advancements to ensure all individual data is secure for the protection of the individual.

### **7.2.5 Health Economics**

Determining the widespread adoption of the health state estimation relies on external stakeholders and the nature of health economics. Healthcare today heavily relies on an infrastructure set by hospital systems, insurance agencies, and other health service providers. These agents are primarily focused on consumer demand, cost, and profitability, while individual health improvement may not be at the forefront. The incorporation of the health state estimation concept will depend on stakeholder alignment and building a model to incentive both providers and individuals. As the health economy continues to expand, there will inevitably arise challenges managing multiple stakeholders.

### **7.2.6 Standards and Interoperability**

Information and Communication Technologies (ICTs) play notable roles in the improvement of patient care via the exchange of information among healthcare providers. For complex systems to work together in an open way, the private and public sectors to collaborate and

build frameworks for building health tools and services, technical specifications, and content. It becomes a challenge to unite multiple stakeholders. The Standards and Interoperability Framework within the Office of the National Coordinator for Health Information Technology has made progress to take input from various stakeholders and create health IT specifications for us throughout the U.S. [115]. The Fast Healthcare Interoperability Resources (FHIR) is an international standard for exchanging healthcare information electronically, advancing interoperability in the healthcare community [200]. FHIR allows for the sharing of information in a standard way regardless of ways local electronic health records represent or store data, allowing for more access to real-time data. However, there remain challenges with the seamless exchange of patient information across different healthcare systems.

### **7.2.7 Research Community**

Collaboration between many different fields of medicine, bioscience, social science, computer science, statistics, and engineering must come together to have an inter-disciplinary approach towards these challenges. HealthMedia can be viewed as one response from the multimedia community to rise to this challenge. There is an increasing number of research work that shows how core multimedia research is becoming an essential enabler for solutions with applications and relevance for the societal health questions. Within this workshop, we continue to explore the relevance, contribution, and future directions of multimedia to health care and personal health. This workshop brings together researchers from diverse topics such as multimedia, tracking, life-logging, accessibility, HCI, but also health, medicine, and psychology to address challenges and opportunities of multimedia in and for health. To forward this mission through my contributions to the research community, I have led several projects, including leading a workshop committee chair for HealthMedia at the ACM Multimedia 2019 Conference in Nice, France [28], founding the Institute for Future Health at the

University of California, Irvine, and leading a session during an international workshop for Food Computing alongside Professor Ramesh Jain.

## **7.3 Final Words**

Health State estimation is just one piece of the total health continuum concept. Ultimately, there must be advancements to ensure the development of reliable, practical, and useful systems to guide people towards better health through a navigational paradigm. Additionally, privacy and security methods must evolve concurrently for such systems to function in the real world. This research invites others to participate in building the foundations of health state estimation and navigational approaches to health. This starting point sets the aim by which people can enjoy peak health by utilizing the combined advancements in computer science, medical research, and technology at large. We hope this builds research momentum towards providing a high-quality, healthy life for every person.

# Bibliography

- [1] 23andMe. DNA Genetic Testing & Analysis - 23andMe, 2020.
- [2] D. O. Abegunde, C. D. Mathers, T. Adam, M. Ortegón, and K. Strong. The burden and costs of chronic diseases in low-income and middle-income countries, 12 2007.
- [3] ACPM. What Is Preventive Medicine, 2019.
- [4] J. G. Adair. The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology*, 69(2):334–345, 5 1984.
- [5] R. Agrawal and R. Srikant. Mining sequential patterns. In *Data Engineering, 1995. Proceedings of the Eleventh International Conference on*, pages 3–14. IEEE, 1995.
- [6] F. Albinali, S. Intille, W. Haskell, and M. Rosenberger. Using wearable activity type detection to improve physical activity energy expenditure estimation. In *UbiComp'10 - Proceedings of the 2010 ACM Conference on Ubiquitous Computing*, volume 2010, pages 311–320. NIH Public Access, 2010.
- [7] M. Altini, P. Casale, J. Penders, and O. Amft. Personalized cardiorespiratory fitness and energy expenditure estimation using hierarchical Bayesian models. *Journal of Biomedical Informatics*, 56:195–204, 8 2015.
- [8] M. Altini, P. Casale, J. Penders, and O. Amft. Cardiorespiratory fitness estimation in free-living using wearable sensors. *Artificial Intelligence in Medicine*, 68:37–46, 3 2016.
- [9] S. Ananiadou, S. Pyysalo, J. Tsujii, and D. B. Kell. Event extraction for systems biology by text mining the literature, 7 2010.
- [10] Apple. Use Screen Time on your iPhone, iPad, or iPod touch - Apple Support, 2020.
- [11] H. J. Arevalo, F. Vadakkumpadan, E. Guallar, A. Jebb, P. Malamas, K. C. Wu, and N. A. Trayanova. Arrhythmia risk stratification of patients after myocardial infarction using personalized heart models. *Nature Communications*, 7:11437, 5 2016.
- [12] G. L. Armstrong, L. A. Conn, and R. W. Pinner. Trends in infectious disease mortality in the United States during the 20th century. *Journal of the American Medical Association*, 281(1):61–66, 1 1999.

- [13] S. P. Ashdown and L. Dunne. A study of automated custom fit: Readiness of the Technology for the apparel industry. *Clothing and Textiles Research Journal*, 24(2):121–136, 3 2006.
- [14] A. M. Axell, H. E. MacLean, D. R. Plant, L. J. Harcourt, J. A. Davis, M. Jimenez, D. J. Handelsman, G. S. Lynch, and J. D. Zajac. Continuous testosterone administration prevents skeletal muscle atrophy and enhances resistance to fatigue in orchidectomized male mice. *American Journal of Physiology - Endocrinology and Metabolism*, 291(3):E506–E516, 9 2006.
- [15] J. Ayres, J. Flannick, J. Gehrke, and T. Yiu. Sequential pattern mining using a bitmap representation. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 429–435. ACM, 2002.
- [16] J. F. Bach. Predictive medicine. *Bulletin et mémoires de l'Académie royale de médecine de Belgique*, 151(1), 1996.
- [17] I. Badash, N. P. Kleinman, S. Barr, J. Jang, S. Rahman, and B. W. Wu. Redefining Health: The Evolution of Health Ideas from Antiquity to the Era of Value-Based Care. *Cureus*, 2 2017.
- [18] S. Bandyopadhyay, J. Wolfson, D. M. Vock, G. Vazquez-Benitez, G. Adomavicius, M. Elidrisi, P. E. Johnson, and P. J. OConnor. Data mining for censored time-to-event data: a Bayesian network model for predicting cardiovascular risk from electronic health record data. *Data Mining and Knowledge Discovery*, 29(4):1033–1069, 7 2015.
- [19] E. W. Banister and T. W. Calvert. Planning for future performance: implications for long term training. *Canadian journal of applied sport sciences. Journal canadien des sciences appliquees au sport*, 5(3):170–176, 9 1980.
- [20] E. L. Barrett-Connor. Testosterone and risk factors for cardiovascular disease in men, 6 1995.
- [21] D. S. Bassett and E. T. Bullmore. Human brain networks in health and disease, 8 2009.
- [22] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, C. Gulcehre, F. Song, A. Ballard, J. Gilmer, G. Dahl, A. Vaswani, K. Allen, C. Nash, V. Langston, C. Dyer, N. Heess, D. Wierstra, P. Kohli, M. Botvinick, O. Vinyals, Y. Li, and R. Pascanu. Relational inductive biases, deep learning, and graph networks. *Arxiv*, pages 1–40, 2018.
- [23] D. J. Bem. Self-Perception Theory. *Advances in Experimental Social Psychology*, 6, 1972.
- [24] T. K. Bera. Bioelectrical impedance methods for noninvasive health monitoring: A review, 2014.

- [25] T. H. H. G. R. G. L. Bernard, C. *Lectures on the phenomena common to animals and plants*. Thomas, Springfield, IL, 1974.
- [26] W. S. Bernard and P. W. Christopher. *World cancer report 2014*. International Agency for Research on Cancer, Lyon, 2014.
- [27] R. R. Bini, P. A. Hume, and A. Cerviri. A comparison of cycling SRM crank and strain gauge instrumented pedal measures of peak torque, crank angle at peak torque and power output. In *Procedia Engineering*, volume 13, pages 56–61. Elsevier Ltd, 1 2011.
- [28] S. Boll, J. S. Lee, J. Meyer, N. Nag, and N. E. O’Connor. HealthMedia’19. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 2720–2721, New York, NY, USA, 10 2019. Association for Computing Machinery (ACM).
- [29] J. Borresen and M. Ian Lambert. The quantification of training load, the training response and the effect on performance, 2009.
- [30] W. M. Bortz. Biological basis of determinants of health, 3 2005.
- [31] A. Bowling. *Measuring Health*, 2005.
- [32] K. L. Brigham. Predictive health: The imminent revolution in health care, 10 2010.
- [33] G. Brooks, T. Austin, D. Fahey, and T. P. White. *Exercise physiology: Human bioenergetics and its applications. No. Ed. 2*. Mayfield Publishing Company, 1996.
- [34] H. Brüßow. What is health? *Microbial Biotechnology*, 6(4):341–348, 7 2013.
- [35] R. Budgett. Fatigue and underperformance in athletes: The overtraining syndrome. *British Journal of Sports Medicine*, 32(2):107–110, 6 1998.
- [36] A. C. o. Cardiology. ASCVD Risk Estimator +, 2020.
- [37] E. J. Caruana, M. Roman, J. Hernández-Sánchez, and P. Solli. Longitudinal studies. *Journal of Thoracic Disease*, 7(11):E537–E540, 2015.
- [38] C. J. Caspersen and G. M. Christenson. Dear editor : We submit our manuscript entitled Preparation of N , O -carboxymethyl chitosan ( NOCs ) coated alginate ( ALg ) microcapsules and application in Bifidobacterium longum BIOMA 5920 to Materials Science and Engineering C for publication . I. *Public Health Reports*, 100(April):126 – 131, 1985.
- [39] CBInsights. Global Healthcare Report Q3 2019, 2019.
- [40] CBInsights. WellnessTech\_MarketMap1.png (1333888), 2019.
- [41] CDC. CDC - Exposome and Exposomics - NIOSH Workplace Safety and Health Topic, 2014.

- [42] CDC. Heart Disease Facts — cdc.gov, 2020.
- [43] K. E. Center. Laser Cataract Surgery FAQs, 2020.
- [44] B. Chandrasekaran. Generic Tasks in Knowledge-Based Reasoning: High-Level Building Blocks for Expert System Design. *IEEE Expert-Intelligent Systems and their Applications*, 1(3):23–30, 1986.
- [45] Z. Chaochen, C. A. R. Hoare, and A. P. Ravn. A calculus of durations. *Information processing letters*, 40(5):269–276, 1991.
- [46] A. G. Chapple and J. W. Blackston. Finding Benefit in n-of-1 Trials. *JAMA International Medicine*, 2011.
- [47] J. T. Chen, J. Beckfield, P. D. Waterman, and N. Krieger. Can Changes in the Distributions of and Associations Between Education and Income Bias Temporal Comparisons of Health Disparities? An Exploration With Causal Graphs and Simulations. *American Journal of Epidemiology*, 2013.
- [48] W. Chen, J. L. Wilson, M. Khaksari, M. A. Cowley, and P. J. Enriori. Abdominal fat analyzed by DEXA scan reflects visceral body fat and improves the phenotype description and the assessment of metabolic risk in mice. *American Journal of Physiology - Endocrinology and Metabolism*, 303(5):E635–E643, 9 2012.
- [49] E. K. Choe, N. B. Lee, B. Lee, W. Pratt, and J. A. Kientz. Understanding quantified-selfers’ practices in collecting and exploring personal data. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 1143–1152. Association for Computing Machinery, 2014.
- [50] N. Chomsky. Methodological Preliminaries from Aspects. In *Aspects of the Theory of Syntax*, pages 3–62. The MIT Press, Cambridge, 1965.
- [51] M. Chu, T. Nguyen, V. Pandey, Y. Zhou, H. N. Pham, R. Bar-Yoseph, S. Radom-Aizik, R. Jain, D. M. Cooper, and M. Khine. Respiration rate and volume measurements using wearable strain sensors. *npj Digital Medicine*, 2(1), 12 2019.
- [52] P. Cintia, L. Pappalardo, and D. Pedreschi. ’Engine matters’: A first large scale data driven study on cyclists’ performance. In *Proceedings - IEEE 13th International Conference on Data Mining Workshops, ICDMW 2013*, pages 147–153. IEEE Computer Society, 2013.
- [53] R. E. Clark. The classical origins of Pavlov’s conditioning. *Integrative Physiological and Behavioral Science*, 39(4):279–294, 2004.
- [54] M. Clinic. Coronary angiogram, 2020.
- [55] L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, and F. Zhang. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121):819–823, 2 2013.

- [56] T. O. F. Contents. Automated Fitness Level ( VO 2 max ) Estimation with Heart Rate and Speed Data. *Firstbeat*, pages 1–9, 2014.
- [57] J. H. Coote and M. J. Joyner. Is precision medicine the route to a healthy world?, 2015.
- [58] R. Corive, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18(1):32–80, 2001.
- [59] F. H. CRICK. On protein synthesis. *Symposia of the Society for Experimental Biology*, 12:138–163, 1958.
- [60] S. Cycling. Technology — Stages Cycling North America, 2020.
- [61] T. H. Davenport and J. Dyché. Big Data in Big Companies. Technical report, International Institute for Analytics, 2013.
- [62] D. De Amici, C. Klersy, F. Ramajoli, L. Brustia, and P. Politi. Impact of the Hawthorne effect in a longitudinal clinical study: The case of anesthesia. *Controlled Clinical Trials*, 21(2):103–114, 4 2000.
- [63] J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. ODonoghue, D. Visentin, G. van den Driessche, B. Lakshminarayanan, C. Meyer, F. Mackinder, S. Bouton, K. Ayoub, R. Chopra, D. King, A. Karthikesalingam, C. O. Hughes, R. Raine, J. Hughes, D. A. Sim, C. Egan, A. Tu-fail, H. Montgomery, D. Hassabis, G. Rees, T. Back, P. T. Khaw, M. Suleyman, J. Cornebise, P. A. Keane, and O. Ronneberger. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, 24(9):1342–1350, 9 2018.
- [64] L. De Greef, M. Goel, M. J. Seo, E. C. Larson, J. W. Stout, J. A. Taylor, and S. N. Patel. BiliCam: Using mobile phones to monitor newborn jaundice. In *UbiComp 2014 - Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 331–342, New York, New York, USA, 2014. ACM Press.
- [65] D. Dias and O. Tchepel. Modelling of human exposure to air pollution in the urban environment: A GPS-based approach. *Environmental Science and Pollution Research*, 21(5):3558–3571, 3 2014.
- [66] Dictionary.com. Diopter — Definition of Diopter at Dictionary.com, 2020.
- [67] D. Diez, . M. Wheelock, S. Goto, J. Z. Haeggström, G. Paulsson-Berne, G. K. Hansson, U. Hedin, A. Gabrielsen, and C. E. Wheelock. The use of network analyses for elucidating mechanisms in cardiovascular disease, 1 2010.
- [68] H. M. Doerr. Stella ten years later: A review of the literature. *International Journal of Computers for Mathematical Learning*, 2004.



- [69] R. C. Dorf, R. H. Bishop, B. Columbus, I. New, Y. San, F. Upper, S. River, A. Cape, T. Dubai, L. Madrid, M. Munich, P. Montreal, T. Delhi, M. C. Sao, P. Sydney, H. Kong, S. Singapore, and T. Tokyo. *Modern Control Systems TWELFTH EDITION*. Technical report, University of California, Davis, 2000.
- [70] R. J. Dubos. *Mirage of Health: Utopias, Progress and Biological Change*, 1987.
- [71] H. L. DUNN. High-level wellness for man and society. *American journal of public health*, 49(6):786–792, 6 1959.
- [72] G. F. Dunton, E. Dzibur, and S. Intille. Feasibility and performance test of a real-time sensor-informed context-sensitive ecological momentary assessment to capture physical activity. *Journal of Medical Internet Research*, 18(6):e106, 6 2016.
- [73] T. Economist. Quality-of-life index 1. Technical report, The Economist, 2005.
- [74] L. Ehrlinger and W. Wöß. Towards a definition of knowledge graphs. In *CEUR Workshop Proceedings*, volume 1695, 2016.
- [75] W. Elmenreich. *Sensor Fusion in Time-Triggered Systems*. Technical report, Institute for Technical Computer Science, 2002.
- [76] U. EPA. *Human Health Risk Assessment Research Methods, Models, Tools, and Databases*, 2020.
- [77] C. Esposito, A. Castiglione, C. A. Tudorica, and F. Pop. Security and Privacy for Cloud-Based Data Management in the Health Network Service Chain: A Microservice Approach. *IEEE Communications Magazine*, 55(9):102–108, 2017.
- [78] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2 2017.
- [79] European Commission. *Data protection in the EU* — European Commission, 2016.
- [80] D. L. Evans. Cardiovascular adaptations to exercise and training. *The Veterinary clinics of North America. Equine practice*, 1(3):513–531, 1985.
- [81] A. Falcon. *Aristotle on Causality*, 2006.
- [82] J. Fan, W. Wu, and L. Hunter. *Clothing Appearance and Fit: Science and Technology*. Woodhead Publishing Limited, 2004.
- [83] F. C. Fang and A. Casadevall. Reductionistic and holistic science, 4 2011.
- [84] Firstbeat. *White Papers and Publications - Firstbeat*, 2020.
- [85] M. Fleischman and E. Hovy. Emotional Variation in speech-based Natural Language Generation. *Proceedings of the 2nd International Conference on Natural Language Generation (INLG'02)*, pages 57–64, 2002.

- [86] M. Gao, V. K. Singh, and R. Jain. EventShop: From heterogeneous web streams to personalized situation detection and control. In *Proceedings of the 4th Annual ACM Web Science Conference, WebSci'12*, volume volume, pages 105–108, New York, New York, USA, 2012. Association for Computing Machinery.
- [87] S. Gastinger, A. Sorel, G. Nicolas, A. Gratas-Delamarche, and J. Prioux. A comparison between ventilation and heart rate as indicator of oxygen uptake during different intensities of exercise. *Journal of sports science & medicine*, 9(1):110–8, 2010.
- [88] S. Gentry and P. Badrinath. Defining Health in the Era of Value-based Care: Lessons from England of Relevance to Other Health Systems. *Cureus*, 3 2017.
- [89] Global Wellness Institute. *Statistics & Facts* Global Wellness Institute, 2016.
- [90] S. Glouberman and J. Millar. Evolution of the determinants of health, health policy, and health information systems in Canada, 3 2003.
- [91] S. Goldstein and J. A. Naglieri. *Encyclopedia of Child Behavior and Development*. Springer Science & Business Media, 2011.
- [92] K. Goldston and A. J. Baillie. Depression and coronary heart disease: A review of the epidemiological evidence, explanatory mechanisms and management approaches, 2008.
- [93] S. N. Goodman, S. Goel, and M. R. Cullen. Machine learning, health disparities, and causal reasoning, 12 2018.
- [94] Google. Google Scholar, 2020.
- [95] Google. Google Trends, 2020.
- [96] B. J. Gordon. *Anatomy and Physiology*. OpenStax College, Rice University,, Houston, 2016.
- [97] I. Gormezano and E. James Kehoe. *Handbook of Learning and Cognitive Processes (Volume 2): Conditioning and ...* - Google Books, 1975.
- [98] Government of Ontario. Preventing and Managing Chronic Disease: Ontario’s Framework Ministry of Health and Long-Term Care. Technical Report May, Government of Ontario, 2007.
- [99] A. Greco, P. Steca, R. Pozzi, D. Monzani, M. D’Addario, A. Villani, V. Rella, A. Giglio, G. Malfatto, and G. Parati. Predicting depression from illness severity in cardiovascular disease patients: Self-efficacy beliefs, illness perception, and perceived social support as mediators. *International Journal of Behavioral Medicine*, 21(2):221–229, 2014.
- [100] P. Grigg. Peripheral neural mechanisms in proprioception, 1994.
- [101] T. K. GS Brush. DNA Replication Mechanisms. In *DNA Replication in Eukaryotic Cells*, page 21ff. Garland Science, 1996.

- [102] D. Guichard. Introduction to Combinatorics and Graph Theory, 2016.
- [103] D. Gunning. Explainable artificial intelligence (xai). Technical report, DARPA, 2017.
- [104] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin. A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing*, 240:98–109, 5 2017.
- [105] P. K. Gupta, H. Gupta, A. Sundaram, M. Kaushik, X. Fang, W. J. Miller, D. J. Esterbrooks, C. B. Hunter, I. I. Pipinos, J. M. Johanning, T. G. Lynch, R. A. Forse, S. M. Mohiuddin, and A. N. Mooss. Development and validation of a risk calculator for prediction of cardiac risk after surgery. *Circulation*, 124(4):381–387, 7 2011.
- [106] F. Gustafsson. Statistical Sensor Fusion, 2013.
- [107] G. H. Guyatt, J. L. Keller, R. Jaeschke, D. Rosenbloom, J. D. Adachi, and M. T. Newhouse. The n-of-1 randomized controlled trial: Clinical usefulness. Our three-year experience. *Annals of Internal Medicine*, 112(4):293–299, 2 1990.
- [108] A. Haag, S. Goronzy, P. Schaich, and J. Williams. Emotion recognition using biosensors: First steps towards an automatic system. In *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, volume 3068, pages 36–48. Springer Verlag, 2004.
- [109] S. L. Halson. Monitoring Training Load to Understand Fatigue in Athletes, 11 2014.
- [110] J. Han, J. Pei, B. Mortazavi-Asl, Q. Chen, U. Dayal, and M.-C. Hsu. FreeSpan: frequent pattern-projected sequential pattern mining. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 355–359. ACM, 2000.
- [111] J. L. Hargrove. Dynamic Modeling in the Health Sciences, 1998.
- [112] N. I. o. Health. Individuals and Families Models and Interventions - Health and Behavior - NCBI Bookshelf, 2001.
- [113] Helix. Helix - Population Genomics, 2020.
- [114] HHS. Summary of the HIPAA Privacy Rule — HHS.gov, 2013.
- [115] HHS. Standards and Interoperability Framework The Standards and Interoperability Framework Functions. Technical report, Health and Human Services, 2020.
- [116] R. Hill, D. Shadija, and M. Rezai. Enabling Community Health Care with Microservices, 2017.
- [117] D. S. Himmelstein, A. Lizee, C. Hessler, L. Brueggeman, S. L. Chen, D. Hadley, A. Green, P. Khankhanian, and S. E. Baranzini. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *eLife*, 6, 9 2017.

- [118] M. Hirsch, Irl B. *History of Glucose Monitoring*. American Diabetes Association, Arlington, 2018.
- [119] C. Holz and E. J. Wang. Glabella. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):1–23, 9 2017.
- [120] L. Hood and S. H. Friend. Predictive, personalized, preventive, participatory (P4) cancer medicine. *Nature Reviews Clinical Oncology*, 8(3):184–187, 3 2011.
- [121] B. Hookway. *Interface*. The MIT Press, 2014.
- [122] F. B. Horak. Clinical measurement of postural control in adults. *Physical Therapy*, 67(12):1881–1885, 1987.
- [123] D. Hoshino, Y. Yoshida, Y. Kitaoka, H. Hatta, and A. Bonen. High-intensity interval training increases intrinsic rates of mitochondrial fatty acid oxidation in rat red and white skeletal muscle. *Applied Physiology, Nutrition and Metabolism*, 38(3):326–333, 3 2013.
- [124] F. Z. Hou, F. W. Li, J. Wang, and F. R. Yan. Visibility graph analysis of very short-term heart rate variability during sleep. *Physica A: Statistical Mechanics and its Applications*, 458:140–145, 9 2016.
- [125] T. W. House. Precision Medicine Initiative, 2020.
- [126] X. Huang. Blockchain in Healthcare: A Patient-Centered Model. *Biomedical Journal of Scientific & Technical Research*, 20(3):15017, 8 2019.
- [127] T. K. Hui and R. S. Sherratt. Coverage of emotion recognition for common wearable biosensors. *Biosensors*, 8(2):30, 3 2018.
- [128] R. B. Hull, W. P. Stewart, and Y. K. Yi. Experience Patterns: Capturing the Dynamic Nature of a Recreation Experience. *Journal of Leisure Research*, 24(3):240–252, 7 1992.
- [129] IEEE Computer Society. Computer Magazine: Special Issue on Digital Health — IEEE Computer Society, 2019.
- [130] N. H. L. . B. Institute. Framingham Heart Study (FHS) — National Heart, Lung, and Blood Institute (NHLBI), 2018.
- [131] N. W. Institute. The Six Dimensions of Wellness - National Wellness Institute.
- [132] Institute of Medicine (US) Committee on Health and, P. Behavior: Research, and Policy. *Individuals and Families: Models and Interventions*. National Academies Press (US), Washington, DC, 2001.
- [133] T. Integrative. The integrative human microbiome project: Dynamic analysis of microbiome-host omics profiles during periods of human health and disease corresponding author, 9 2014.

- [134] J. P. Ioannidis. Neglecting Major Health Problems and Broadcasting Minor, Uncertain Issues in Lifestyle Science, 12 2019.
- [135] R. Jain and L. Jalali. Objective self. *IEEE MultiMedia*, 21(4):100–110, 10 2014.
- [136] L. Jal, H. Oh, S. Pongpaichet, R. Jain, L. Jalali, D. Huo, and M. Tang. Personicle: Personal Chronicle of Life Events Personal Health Navigator View project Smart city systems View project Personicle: Personal Chronicle of Life Events. Technical report, University of California, Irvine, 2014.
- [137] C. Jiang, X. Wang, X. Li, J. Inlora, T. Wang, Q. Liu, and M. Snyder. Dynamic Human Environmental Exposome Revealed by Longitudinal Personal Monitoring. *Cell*, 175(1):277–291, 9 2018.
- [138] K. W. Johnson, K. Shameer, B. S. Glicksberg, B. Readhead, P. P. Sengupta, J. L. Björkegren, J. C. Kovacic, and J. T. Dudley. Enabling Precision Cardiology Through Multiscale Biology and Systems Medicine, 6 2017.
- [139] A. M. Jones and A. Vanhatalo. The Critical Power Concept: Applications to Sports Performance with a Focus on Intermittent High-Intensity Exercise, 3 2017.
- [140] J. E. Joseph, W. von Humboldt, and P. Heath. On Language: The Diversity of Human Language-Structure and Its Influence on the Mental Development of Mankind. *Language*, 67(4):843, 1991.
- [141] D. Kahneman, A. B. Krueger, D. A. Schkade, N. Schwarz, and A. A. Stone. A survey method for characterizing daily life experience: The day reconstruction method. *Science*, 306(5702):1776–1780, 12 2004.
- [142] S. V. Kasl and S. Cobb. Health behavior, illness behavior, and sick role behavior. I. Health and illness behavior., 1966.
- [143] Y. Kawano and K. Yanai. Food image recognition with deep convolutional features. In *UbiComp 2014 - Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 589–593, New York, New York, USA, 2014. Association for Computing Machinery, Inc.
- [144] J. Kim, H. R. R. Woo, and H. G. G. Nam. Toward Systems Understanding of Leaf Senescence: An Integrated Multi-Omics Perspective on Leaf Senescence Research, 6 2016.
- [145] M. Kimball, R. Willis, G. Akerlof, T. Antonucci, R. Barsky, S. Basu, D. Benjamin, K. Charles, F. Conrad, M. Couper, M. Elsby, G. Fisher, B. Frey, C. House, M. Hurd, H. Levy, C. Manski, R. Nesse, F. Ohtake, A. Rangel, L. Rayo, M. Shapiro, D. Silverman, A. Stutzer, Y. Tsutsui, and J. Yellen. Utility and Happiness. Technical report, University of Michigan, 2006.
- [146] D. A. Kindig and G. Stoddart. What is population health?, 3 2003.

- [147] S. Kumar, G. Abowd, W. T. Abraham, M. Al’Absi, D. H. Chau, E. Ertin, D. Estrin, D. Ganesan, T. Hnat, S. M. Hossain, Z. Ives, J. Kerr, B. M. Marlin, S. Murphy, J. M. Rehg, I. Nahum-Shani, V. Shetty, I. Sim, B. Spring, M. Srivastava, and D. Wetter. Center of Excellence for Mobile Sensor Data-to-Knowledge (MD2K). *IEEE Pervasive Computing*, 16(2):18–27, 2017.
- [148] C. K. H. J. L. T.-W. Kwon, Oh Wook. Eurospeech 2003 Abstract, 2003.
- [149] K. K. Kwong, J. W. Belliveau, D. A. Chesler, I. E. Goldberg, R. M. Weisskoff, B. P. Poncelet, D. N. Kennedy, B. E. Hoppel, M. S. Cohen, R. Turner, H. M. Cheng, T. J. Brady, and B. R. Rosen. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, 89(12):5675–5679, 1992.
- [150] C. Lalanneab. Crossmodal integration for perception and action. *Journal of Physiology*, 2004.
- [151] E. C. Larson, M. Goel, G. Boriello, S. Heltshe, M. Rosenfeld, and S. N. Patel. SpiroSmart: Using a microphone to measure lung function on a mobile phone. In *UbiComp’12 - Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 280–289, New York, New York, USA, 2012. ACM Press.
- [152] D. c. Lee, E. G. Artero, X. Sui, and S. N. Blair. Mortality trends in the general population: the importance of cardiorespiratory fitness., 2010.
- [153] J. Lee. Odds Ratio or Relative Risk for Cross-Sectional Data? *Journal of Epidemiology*, 1994.
- [154] S. Leng, R. S. Tan, K. T. C. Chai, C. Wang, D. Ghista, and L. Zhong. The electronic stethoscope, 7 2015.
- [155] C.-S. Li, P. S. Yu, and V. Castelli. Hierarchyscan: A hierarchical similarity search algorithm for databases of long sequences. In *Data Engineering, 1996. Proceedings of the Twelfth International Conference on*, pages 546–553. IEEE, 1996.
- [156] F. Li, X. Zou, P. Liu, and J. Y. Chen. New threats to health data privacy. *BMC Bioinformatics*, 12(S12):S7, 12 2011.
- [157] T. Li. *Event Mining: Algorithms and Applications*. CRC Press, 2015.
- [158] E. O. Lillie, B. Patay, J. Diamant, B. Issell, E. J. Topol, and N. J. Schork. The n-of-1 clinical trial: The ultimate strategy for individualizing medicine? *Personalized Medicine*, 8(2):161–173, 3 2011.
- [159] C. Liu, Y. Bai, X. Xu, L. Sun, A. Wang, T. Y. Wang, S. K. Maurya, M. Periasamy, M. Morishita, J. Harkema, Z. Ying, Q. Sun, and S. Rajagopalan. Exaggerated effects of particulate matter air pollution in genetic type II diabetes mellitus. *Particle and Fibre Toxicology*, 11(1):1–14, 5 2014.

- [160] D. M. Lloyd-Jones, L. T. Braun, C. E. Ndumele, S. C. Smith, L. S. Sperling, S. S. Virani, and R. S. Blumenthal. Use of Risk Assessment Tools to Guide Decision-Making in the Primary Prevention of Atherosclerotic Cardiovascular Disease: A Special Report From the American Heart Association and American College of Cardiology. *Journal of the American College of Cardiology*, 73(24):3153–3167, 6 2019.
- [161] R. Lozano, M. Naghavi, K. Foreman, S. Lim, K. Shibuya, V. Aboyans, J. Abraham, T. Adair, R. Aggarwal, S. Y. Ahn, M. A. AlMazroa, M. Alvarado, H. R. Anderson, L. M. Anderson, K. G. Andrews, C. Atkinson, L. M. Baddour, S. Barker-Collo, D. H. Bartels, M. L. Bell, E. J. Benjamin, D. Bennett, K. Bhalla, B. Bikbov, A. B. Abdulhak, G. Birbeck, F. Blyth, I. Bolliger, S. Boufous, C. Bucello, M. Burch, P. Burney, J. Carapetis, H. Chen, D. Chou, S. S. Chugh, L. E. Coffeng, S. D. Colan, S. Colquhoun, K. E. Colson, J. Condon, M. D. Connor, L. T. Cooper, M. Corriere, M. Cortinovis, K. C. de Vaccaro, W. Couser, B. C. Cowie, M. H. Criqui, M. Cross, K. C. Dabhadkar, N. Dahodwala, D. De Leo, L. Degenhardt, A. Delossantos, J. Denenberg, D. C. Des Jarlais, S. D. Dharmaratne, E. R. Dorsey, T. Driscoll, H. Duber, B. Ebel, P. J. Erwin, P. Espindola, M. Ezzati, V. Feigin, A. D. Flaxman, M. H. Forouzanfar, F. G. R. Fowkes, R. Franklin, M. Fransen, M. K. Freeman, S. E. Gabriel, E. Gakidou, F. Gaspari, R. F. Gillum, D. Gonzalez-Medina, Y. A. Halasa, D. Haring, J. E. Harrison, R. Havmoeller, R. J. Hay, B. Hoen, P. J. Hotez, D. Hoy, K. H. Jacobsen, S. L. James, R. Jasrasaria, S. Jayaraman, N. Johns, G. Karthikeyan, N. Kassebaum, A. Keren, J.-P. Khoo, L. M. Knowlton, O. Kobusingye, A. Koranteng, R. Krishnamurthi, M. Lipnick, S. E. Lipshultz, S. L. Ohno, J. Mabweijano, M. F. MacIntyre, L. Mallinger, L. March, G. B. Marks, R. Marks, A. Matsumori, R. Matzopoulos, B. M. Mayosi, J. H. McAnulty, M. M. McDermott, J. McGrath, Z. A. Memish, G. A. Mensah, T. R. Merriman, C. Michaud, M. Miller, T. R. Miller, C. Mock, A. O. Mocumbi, A. A. Mokdad, A. Moran, K. Mulholland, M. N. Nair, L. Naldi, K. M. V. Narayan, K. Nasser, P. Norman, M. O'Donnell, S. B. Omer, K. Ortblad, R. Osborne, D. Ozgediz, B. Pahari, J. D. Pandian, A. P. Rivero, R. P. Padilla, F. Perez-Ruiz, N. Perico, D. Phillips, K. Pierce, C. A. Pope, E. Porrini, F. Pourmalek, M. Raju, D. Ranganathan, J. T. Rehm, D. B. Rein, G. Remuzzi, F. P. Rivara, T. Roberts, F. R. De León, L. C. Rosenfeld, L. Rushton, R. L. Sacco, J. A. Salomon, U. Sampson, E. Sanman, D. C. Schwebel, M. Segui-Gomez, D. S. Shepard, D. Singh, J. Singleton, K. Sliwa, E. Smith, A. Steer, J. A. Taylor, B. Thomas, I. M. Tleyjeh, J. A. Towbin, T. Truelsen, E. A. Undurraga, N. Venketasubramanian, L. Vijayakumar, T. Vos, G. R. Wagner, M. Wang, W. Wang, K. Watt, M. A. Weinstock, R. Weintraub, J. D. Wilkinson, A. D. Woolf, S. Wulf, P.-H. Yeh, P. Yip, A. Zabetian, Z.-J. Zheng, A. D. Lopez, and C. J. Murray. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, 380(9859):2095–2128, 12 2012.
- [162] A. LUCIA, J. HOYOS, A. SANTALLA, M. PEREZ, and J. CHICHARRO. Kinetics of  $\dot{V}O_2$  in professional cyclists. *Medicine and Science in Sports and Exercise*, 34(2):320–325, 2 2002.

- [163] Z. Lv, L. Feng, H. Li, and S. Feng. Hand-free motion interaction on google glass. In *SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applications, SA 2014*. Association for Computing Machinery, Inc, 11 2014.
- [164] C. C. Macadam. Understanding and modeling the human driver. *Vehicle System Dynamics*, 40(1-3):101–134, 9 2003.
- [165] H. Maezawa, K. Yoshida, T. Nagamine, J. Matsubayashi, R. Enatsu, K. Bessho, and H. Fukuyama. Somatosensory evoked magnetic fields following electric tongue stimulation using pin electrodes. *Neuroscience Research*, 62(2):131–139, 10 2008.
- [166] L. Maitre and M. Vrijheid. Health Impacts of Developmental Exposure to Environmental Chemicals. In *Exposomics: The Exposome in Early Life*, pages 463–484. Springer, Singapore, 2019.
- [167] S. Majumder and M. J. Deen. Smartphone sensors for health monitoring and diagnosis, 5 2019.
- [168] I. Maker. Insight Maker — Free Simulation and Modeling in your Browser.
- [169] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky. The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, 2015.
- [170] F. Martin-Sanchez, G. Lopez-Campos, and K. Gray. Biomedical Informatics Methods for Personalized Medicine and Participatory Health. In *Methods in Biomedical Informatics: A Pragmatic Approach*, pages 347–394. Elsevier Inc., 10 2013.
- [171] M. Massidda, N. Miyamoto, S. Beckley, N. Kikuchi, and N. Fuku. Genetics of flexibility. In *Sports, Exercise, and Nutritional Genomics*, pages 273–293. Elsevier, 1 2019.
- [172] Max Roser and Hannah Ritchie. Cancer - Our World in Data, 2019.
- [173] A. McAfee and E. Brynjolfsson. Big Data: The Management Revolution. *Harvard Business Review*, 2012.
- [174] R. McCarney, J. Warner, S. Iliffe, R. Van Haselen, M. Griffin, and P. Fisher. The Hawthorne Effect: A randomised, controlled trial. *BMC Medical Research Methodology*, 7(1):30, 12 2007.
- [175] J. McDuffie. Why the Military Released GPS to the Public. *Popular Mechanics*, 2017.
- [176] H. J. McQuay, D. Carroll, A. R. Jadad, C. J. Glynn, T. Jack, R. A. Moore, and P. J. Wiffen. Dextromethorphan for the treatment of neuropathic pain: a double-blind randomised controlled crossover trial with integral n-of-1 design. *Pain*, 59(1):127–133, 10 1994.
- [177] MDCalc. MDCalc - Medical calculators, equations, scores, and guidelines, 2020.



- [178] Mercer. Quality of Living City Ranking — Mercer, 2020.
- [179] M. Meyer, M. Sedlmair, P. S. Quinan, and T. Munzner. The nested blocks and guidelines model, 2015.
- [180] MIT Media Lab. Frequently asked questions, 2020.
- [181] H. P. Moravec. Sensor Fusion in Certainty Grids for Mobile Robots. In *Sensor Devices and Systems for Robotics*, pages 253–276. Springer Berlin Heidelberg, 1989.
- [182] T. Morzy, M. Wojciechowski, and M. Zakrzewicz. Efficient Constraint-Based Sequential Pattern Mining Using Dataset Filtering Techniques. In *Databases and Information Systems II*, pages 297–309. Springer, 2002.
- [183] O. J. Muensterer, M. Lacher, C. Zoeller, M. Bronstein, and J. Kübler. Google Glass in pediatric surgery: An exploratory study. *International Journal of Surgery*, 12(4):281–289, 4 2014.
- [184] T. B. Murdoch and A. S. Detsky. The inevitable application of big data to health care, 4 2013.
- [185] C. J. Murray, J. Abraham, M. K. Ali, M. Alvarado, C. Atkinson, L. M. Baddour, D. H. Bartels, E. J. Benjamin, K. Bhalla, G. Birbeck, I. Bolliger, R. Burstein, E. Carnahan, H. Chen, D. Chou, S. S. Chugh, A. Cohen, K. E. Colson, L. T. Cooper, W. Couser, M. H. Criqui, K. C. Dabhadkar, N. Dahodwala, G. Danaei, R. P. Dellavalle, D. C. Des Jarlais, D. Dicker, E. L. Ding, E. R. Dorsey, H. Duber, B. E. Ebel, R. E. Engell, M. Ezzati, D. T. Felson, M. M. Finucane, S. Flaxman, A. D. Flaxman, T. Fleming, M. H. Forouzanfar, G. Freedman, M. K. Freeman, S. E. Gabriel, E. Gakidou, R. F. Gillum, D. Gonzalez-Medina, R. Gosselin, B. Grant, H. R. Gutierrez, H. Hagan, R. Havmoeller, H. Hoffman, K. H. Jacobsen, S. L. James, R. Jasrasaria, S. Jayaraman, N. Johns, N. Kassebaum, S. Khatibzadeh, L. M. Knowlton, Q. Lan, J. L. Leasher, S. Lim, J. K. Lin, S. E. Lipshultz, S. London, R. Lozano, Y. Lu, M. F. MacIntyre, L. Mallinger, M. M. McDermott, M. Meltzer, G. A. Mensah, R. Micha, C. Michaud, T. R. Miller, C. Mock, T. E. Moffitt, A. A. Mokdad, A. H. Mokdad, A. E. Moran, D. Mozaffarian, T. Murphy, M. Naghavi, K. M. Venkat Narayan, R. G. Nelson, C. Olives, S. B. Omer, K. Ortblad, B. Ostro, P. M. Pelizzari, D. Phillips, C. A. Pope, M. Raju, D. Ranganathan, H. Razavi, B. Ritz, F. P. Rivara, T. Roberts, R. L. Sacco, J. A. Salomon, U. Sampson, E. Sanman, A. Sapkota, D. C. Schwebel, S. Shahrzad, K. Shibuya, R. Shivakoti, D. Silberberg, G. M. Singh, D. Singh, J. A. Singh, D. A. Sleet, K. Steenland, M. Tavakkoli, J. A. Taylor, G. D. Thurston, J. A. Towbin, M. S. Vavilala, T. Vos, G. R. Wagner, M. A. Weinstock, M. G. Weisskopf, J. D. Wilkinson, A. D. Lopez, and A. Zabetian. The State of US health, 1990-2010: Burden of diseases, injuries, and risk factors. *JAMA - Journal of the American Medical Association*, 310(6):591–608, 8 2013.
- [186] N. Nag and R. Jain. A Navigational Approach to Health: Actionable Guidance for Improved Quality of Life. *Computer*, 52(4):12–20, 4 2019.

- [187] N. Nag, M. Lux, and R. C. Jain. Intrinsic and Extrinsic Motivation Modeling Essential for Multi-Modal Health Recommender Systems. *Arxiv*, 8 2018.
- [188] N. Nag, V. Pandey, and R. Jain. Health multimedia: Lifestyle recommendations based on diverse observations. In *ICMR 2017 - Proceedings of the 2017 ACM International Conference on Multimedia Retrieval*, pages 99–106, New York, New York, USA, 6 2017. Association for Computing Machinery, Inc.
- [189] N. Nag, V. Pandey, and R. Jain. Live personalized nutrition recommendation engine. In *MMHealth 2017 - Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care, co-located with MM 2017*, pages 61–68, New York, New York, USA, 10 2017. Association for Computing Machinery, Inc.
- [190] N. Nag, V. Pandey, and R. C. Jain. Endogenous and Exogenous Multi-Modal Layers in Context Aware Recommendation Systems for Health. *Arxiv*, 8 2018.
- [191] N. Nag, V. Pandey, A. Manjunath, A. Vaka, and R. Jain. Surface Type Estimation from GPS Tracked Bicycle Activities. *Arxiv*, 9 2018.
- [192] N. Nag, V. Pandey, H. Oh, and R. Jain. Cybernetic Health. *Arxiv*, 2017.
- [193] N. Nag, V. Pandey, P. J. Putzel, H. Bhimaraju, S. Krishnan, and R. Jain. Cross-modal health state estimation. In *MM 2018 - Proceedings of the 2018 ACM Multimedia Conference*, pages 1993–2002, 2018.
- [194] N. Nag, V. Pandey, A. Sharma, J. Lam, R. Wang, and R. Jain. Pocket Dietitian: Automated Healthy Dish Recommendations by Location. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 10590 LNCS, pages 444–452. Springer Verlag, 2017.
- [195] N. Nag, A. N. Rao, A. Kulhalli, K. S. Mehta, N. Bhattacharya, P. Ramkumar, A. Bharadwaj, D. Sitaram, and R. Jain. Flavour enhanced food recommendation. In *MADiMa 2019 - Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management, co-located with MM 2019*, pages 60–66, New York, New York, USA, 10 2019. Association for Computing Machinery, Inc.
- [196] National Center for Chronic Disease Prevention and Health Promotion. Chronic Diseases in America, 2019.
- [197] Nature.com. Bioenergetics - Latest research and news — Nature, 2020.
- [198] NCBI. Home - PubMed - NCBI, 2020.
- [199] A. Newell and H. A. Simon. GPS, a program that simulates human thought. In *Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence*, pages 453–460. Morgan Kaufmann Pub, 2013.
- [200] E. NHS. Fast Healthcare Interoperability Resources - NHS Digital, 2019.

- [201] J. K. Nicholson, E. Holmes, and I. D. Wilson. Gut microorganisms, mammalian metabolism and personalized health care. *Nature Reviews Microbiology*, 3(5):431–438, 5 2005.
- [202] NIH. What is the Precision Medicine Initiative? - Genetics Home Reference, 2020.
- [203] D. Noble. Modeling the heart - From genes to cells to the whole organ, 3 2002.
- [204] D. D. Odoneal, M. T. Marois, D. Ward, C. H. Schmid, R. Cabrera, I. Sim, Y. Wang, B. Wilsey, N. Duan, S. G. Henry, and R. L. Kravitz. Effect of Mobile Device-Assisted N-of-1 Trial Participation on Analgesic Prescribing for Chronic Pain: Randomized Controlled Trial. *Journal of General Internal Medicine*, 35(1):102–111, 1 2020.
- [205] H. Oh and R. Jain. From multimedia logs to personal chronicles. In *MM 2017 - Proceedings of the 2017 ACM Multimedia Conference*, pages 881–889, 2017.
- [206] S. L. Oh, Y. Hagiwara, U. Raghavendra, R. Yuvaraj, N. Arunkumar, M. Murugappan, and U. R. Acharya. A deep learning approach for Parkinsons disease diagnosis from EEG signals. *Neural Computing and Applications*, pages 1–7, 8 2018.
- [207] N. M. Oliver, B. Rosario, and A. P. Pentland. A bayesian computer vision system for modeling human interactions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):831–843, 2000.
- [208] S. J. Olshansky. From Lifespan to Healthspan, 10 2018.
- [209] W. H. Organization. International Statistical Classification of Diseases and Related Health Problems. *CWL Publishing Enterprises, Inc., Madison*, 2(Second Edition):352, 2004.
- [210] W. H. Organization. WHO — Metrics: Disability-Adjusted Life Year (DALY), 2014.
- [211] W. H. Organization. Facing the Facts #1 chronic diseases and their common risk factors. Technical report, World Health Organization, 2020.
- [212] D. Otasek, J. H. Morris, J. Bouças, A. R. Pico, and B. Demchak. Cytoscape Automation: Empowering workflow-based network analysis. *Genome Biology*, 20(1), 9 2019.
- [213] C. Pang, G. Y. Lee, T. I. Kim, S. M. Kim, H. N. Kim, S. H. Ahn, and K. Y. Suh. A flexible and highly sensitive strain-gauge sensor using reversible interlocking of nanofibres. *Nature Materials*, 11(9):795–801, 2012.
- [214] Y. J. Park and A. Tuzhilin. The long tail of recommender systems and how to leverage it. In *RecSys’08: Proceedings of the 2008 ACM Conference on Recommender Systems*, pages 11–18, New York, New York, USA, 2008. ACM Press.

- [215] J. Pärkkä, M. Ermes, K. Antila, M. Van Gils, A. Mänttari, and H. Nieminen. Estimating intensity of physical activity: A comparison of wearable accelerometer and gyro sensors and 3 sensor locations. In *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, pages 1511–1514, 2007.
- [216] G. Patents. Flexible tactile sensor, 2002.
- [217] Pathway Commons. Pathway Commons: A Resource for Biological Pathway Analysis, 2020.
- [218] R. S. Patil. REVIEW OF CAUSAL REASONING IN MEDICAL DIAGNOSIS. In *Proceedings - Annual Symposium on Computer Applications in Medical Care*, pages 11–16, 1986.
- [219] J. Pearl. A Probabilistic Calculus of Actions. *Arxiv*, 2013.
- [220] J. Pearl. *The Book of Why*. Basic Books, 2018.
- [221] J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M.-C. Hsu. Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth. In *icccn*, page 215. IEEE, 2001.
- [222] J. Pei, J. Han, and W. Wang. Mining sequential patterns with constraints in large databases. In *Proceedings of the eleventh international conference on Information and knowledge management*, pages 18–25. ACM, 2002.
- [223] Physiopedia. Muscle spindles, 2020.
- [224] D. C. Poole, M. Burnley, A. Vanhatalo, H. B. Rossiter, and A. M. Jones. Critical power: An important fatigue threshold in exercise physiology. *Medicine and Science in Sports and Exercise*, 48(11):2320–2334, 11 2016.
- [225] K. Qu, L. C. Zaba, P. G. Giresi, R. Li, M. Longmire, Y. H. Kim, W. J. Greenleaf, and H. Y. Chang. Individuality and Variation of Personal Regulomes in Primary Human T Cells. *Cell Systems*, 1(1):51–61, 7 2015.
- [226] S. Rajagopalan and R. D. Brook. Air pollution and type 2 diabetes: Mechanistic insights, 12 2012.
- [227] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G. Z. Yang. Deep Learning for Health Informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1):4–21, 1 2017.
- [228] B. Reeves, T. Robinson, and N. Ram. Time for the Human Screenome Project, 1 2020.
- [229] Retul. Retul Bike Fit Sydney - Bicycle Fitting Sydney - Onebody Bike Fitters, 2020.
- [230] R. Ritcher. Precision health: Predicting and preventing disease not just treating it, 2015.

- [231] S. L. Robbins, J. H. Robbins, and D. G. Scarpelli. Human disease - Classifications of diseases — Britannica.com, 2011.
- [232] W. E. Roberts. Skin Type Classification Systems Old and New, 10 2009.
- [233] K. Romstedt. Modeling the Cardiovascular System using STELLA <sup>®</sup> A module for Computational Biology. Technical report, Capital University, 2003.
- [234] R. Ross, S. N. Blair, R. Arena, T. S. Church, J. P. Després, B. A. Franklin, W. L. Haskell, L. A. Kaminsky, B. D. Levine, C. J. Lavie, J. Myers, J. Niebauer, R. Sallis, S. S. Sawada, X. Sui, and U. Wisløff. Importance of Assessing Cardiorespiratory Fitness in Clinical Practice: A Case for Fitness as a Clinical Vital Sign: A Scientific Statement from the American Heart Association. *Circulation*, 134(24):e653–e699, 2016.
- [235] N. P. Russell, Stuart. No Artificial Intelligence: A Modern Approach. In *Artificial Intelligence: A Modern Approach*. Pearson, 2003.
- [236] S. Sachdeva. Fitzpatrick skin typing: Applications in dermatology. *Indian Journal of Dermatology, Venereology and Leprology*, 75(1):93, 1 2009.
- [237] M. Sagner, A. McNeil, P. Puska, C. Auffray, N. D. Price, L. Hood, C. J. Lavie, Z. G. Han, Z. Chen, S. K. Brahmachari, B. S. McEwen, M. B. Soares, R. Balling, E. Epel, and R. Arena. The P4 Health Spectrum A Predictive, Preventive, Personalized and Participatory Continuum for Promoting Healthspan, 3 2017.
- [238] N. Salkind. *Encyclopedia of Research Design*. SAGE Publications, Thousand Oaks, 2010.
- [239] R. Sandhu, T. Georgiou, E. Reznik, L. Zhu, I. Kolesov, Y. Senbabaoglu, and A. Tanenbaum. Graph curvature for differentiating cancer networks. *Scientific Reports*, 5(1):1–13, 7 2015.
- [240] S. Santini, B. Ostermaier, and A. Vitaletti. First experiences using wireless sensor networks for noise pollution monitoring. In *REALWSN 2008 - Proceedings of the 2008 Workshop on Real-World Wireless Sensor Networks*, pages 61–65, 2008.
- [241] N. Sartorius. The meanings of health and its promotion, 2006.
- [242] A. Sebok, C. Wickens, and R. Sargent. Using meta-analyses results and data gathering to support human performance model development. In *Proceedings of the Human Factors and Ergonomics Society*, pages 783–787. SAGE PublicationsSage CA: Los Angeles, CA, 9 2013.
- [243] M. S. Setia. Methodology series module 3: Cross-sectional studies. *Indian Journal of Dermatology*, 61(3):261–264, 5 2016.
- [244] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–2504, 11 2003.

- [245] S. M. Shieber. Lessons from a Restricted Turing Test. *Communications of the ACM*, 37(6):70–78, 4 1994.
- [246] S. Shiffman, A. A. Stone, and M. R. Hufford. Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, 4(1):1–32, 4 2008.
- [247] M. M. Shores, M. L. Biggs, A. M. Arnold, N. L. Smith, W. T. Longstreth, J. R. Kizer, C. H. Hirsch, A. R. Cappola, and A. M. Matsumoto. Testosterone, dihydrotestosterone, and incident cardiovascular disease and mortality in the cardiovascular health study. *Journal of Clinical Endocrinology and Metabolism*, 99(6):2061–2068, 6 2014.
- [248] J. G. Simon, J. B. De Boer, I. M. Joung, H. Bosma, and J. P. Mackenbach. How is your health in general? A qualitative study on self-assessed health. *European Journal of Public Health*, 15(2):200–208, 4 2005.
- [249] V. Siroux, L. Agier, and R. Slama. The exposome concept: A challenge and a potential driver for environmental health research. *European Respiratory Review*, 25(140):124–129, 6 2016.
- [250] J. Smolander, T. Juuti, M. L. Kinnunen, K. Laine, V. Louhevaara, K. Männikkö, and H. Rusko. A new heart rate variability-based method for the estimation of oxygen consumption without individual laboratory calibration: Application example on postal workers. *Applied Ergonomics*, 39(3):325–331, 5 2008.
- [251] P.-E. Sottas, N. Robinson, O. Rabin, and M. Saugy. The Athlete Biological Passport. *Clinical Chemistry*, 57(7):969–976, 7 2011.
- [252] J. Stokes, J. Noren, and S. Shindell. Definition of terms and concepts applicable to clinical preventive medicine. *Journal of Community Health*, 8(1):33–41, 9 1982.
- [253] Strava. Strava — Run and Cycling Tracking on the Social Network for Athletes, 2017.
- [254] T. Sullivan. The future of personalized health is scientific wellness, 2018.
- [255] P. Suravajhala, L. J. Kogelman, and H. N. Kadarmideen. Multi-omic data integration and analysis using systems genomics approaches: Methods and applications In animal production, health and welfare, 4 2016.
- [256] A. L. Svalastog, D. Donev, N. J. Kristoffersen, and S. Gajović. Concepts and definitions of health and health-related values in the knowledge landscapes of the digital society. *Croatian Medical Journal*, 58(6):431–435, 12 2017.
- [257] M. Szklo. Population-based Cohort Studies. Technical Report 1, Oxford Academic, 1998.
- [258] I. Tait. DUBOS’ MIRAGE OF HEALTH. *The Journal of the Royal College of General Practitioners*, 32(237):248, 1982.
- [259] C. Tantucci and D. Modena. Lung function decline in COPD, 2012.

- [260] L. Tarassenko and E. J. Topol. Monitoring Jet Engines and the Health of People, 12 2018.
- [261] R. H. Thaler. Behavioral economics: Past, present, and future, 7 2016.
- [262] M. C. Tissot van Patot, A. E. Hill, C. Dingmann, L. Gaul, K. Fralick, U. Christians, B. Honigman, and M. D. Salman. Risk of impaired coagulation in warfarin patients ascending to altitude ( $\geq 2400$  m), 3 2006.
- [263] G. Toffetti, S. Brunner, M. Blöchlinger, F. Dudouet, and A. Edmonds. An architecture for self-managing microservices. In *Proceedings: AIMC 2015 - Automated Incident Management in Cloud, 1st International Workshop, in conjunction with EuroSYS 2015*, pages 19–24, New York, New York, USA, 4 2015. Association for Computing Machinery, Inc.
- [264] C. Tomasetti and B. Vogelstein. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*, 347(6217):78–81, 1 2015.
- [265] J. W. Travis and R. S. Ryan. *Wellness Workbook: How to Achieve Enduring Health and Vitality*. Celestial Arts, 2004.
- [266] P. D. Treffers, A. W. Goedhart, J. W. Waltz, and E. Koudijs. The systematic collection of patient data in a Centre for Child and Adolescent Psychiatry. *British Journal of Psychiatry*, 157(NOV.):744–748, 1990.
- [267] A. M. Turing. Computing machinery and intelligence. In *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, pages 23–65. Springer Netherlands, 2009.
- [268] UpToDate. Evidence-Based Clinical Decision Support at the Point of Care — UpToDate, 2016.
- [269] L. Valiant. *Probably approximately correct : nature’s algorithms for learning and prospering in a complex world*. Basic Books, 2013.
- [270] L. G. Valiant. A theory of the learnable. In *Proceedings of the Annual ACM Symposium on Theory of Computing*, pages 436–445, 1984.
- [271] L. G. Valiant. Knowledge infusion: In pursuit of robustness in artificial intelligence. In *Leibniz International Proceedings in Informatics, LIPIcs*, volume 2, pages 415–422, 2008.
- [272] I. I. Vankov and J. S. Bowers. Training neural networks to encode symbols enables combinatorial generalization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1791), 2 2020.
- [273] Verily. X - Verily, 2019.

- [274] M. V. Villarejo, B. G. Zapirain, and A. M. Zorrilla. A stress sensor based on galvanic skin response (GSR) controlled by ZigBee. *Sensors (Switzerland)*, 12(5):6075–6101, 5 2012.
- [275] J. Vincent. Apple boasts about sales; Google boasts about how good its AI is - The Verge, 2016.
- [276] E. Vlahu-Gjorgievska and V. Trajkovik. Personal healthcare system model using collaborative filtering techniques. *Advances in Information Sciences and Service Sciences*, 3(3):64–74, 4 2011.
- [277] M. Vlastelica, A. Paulus, V. Musil, G. Martius, and M. Rolínek. Differentiation of Blackbox Combinatorial Solvers. *Arxiv*, 12 2019.
- [278] F. R. Vogenberg, C. I. Barash, and M. Pursel. Personalized medicine - Part 1: Evolution and development into theranostics, 10 2010.
- [279] M. Vrijheid. EThe exposome: A new paradigm to study the impact of environment on health. *Thorax*, 69(9):876–878, 9 2014.
- [280] M. Wainberg, D. Merico, A. Delong, and B. J. Frey. Deep learning in biomedicine. *Nature Biotechnology*, 36(9):829–838, 10 2018.
- [281] L. K. Wallace, K. M. Slattery, and A. J. Coutts. A comparison of methods for quantifying training load: Relationships between modelled and actual training responses. *European Journal of Applied Physiology*, 114(1):11–20, 1 2014.
- [282] E. J. Wang, W. Li, D. Hawkins, T. Gernsheimer, C. Norby-Slycord, and S. N. Patel. HemaApp: Noninvasive blood screening of hemoglobin using smartphone cameras. In *UbiComp 2016 - Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 593–604, New York, New York, USA, 2016. ACM Press.
- [283] F. Wang, N. Lee, J. Hu, J. Sun, S. Ebadollahi, and A. F. Laine. A framework for mining signatures from event sequences and its applications in healthcare data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):272–285, 2013.
- [284] Y. Wang, C. Gan, M. Siegel, Z. Zhang, J. Wu, and J. Tenenbaum. A Computational Model for Combinatorial Generalization in Physical Perception from Sound. In *Creative Commons Attribution 3.0*, 2019.
- [285] S. Warshall. A Theorem on Boolean Matrices. *Journal of the ACM (JACM)*, 9(1):11–12, 1 1962.
- [286] U. o. Washington. The Gut Microbiome and Brain Health, 2018.
- [287] D. M. Wegner. A Computer Network Model of Human Transactive Memory. *Social Cognition*, 13(3):319–339, 9 1995.



- [288] M. Wentink, L. P. Stassen, I. Alwayn, R. J. Hosman, and H. G. Stassen. Rasmussen’s model of human behavior in laparoscopy training. *Surgical Endoscopy and Other Interventional Techniques*, 17(8):1241–1246, 8 2003.
- [289] U. Westermann and R. Jain. Toward a common event model for multimedia applications. *IEEE Multimedia*, 14(1):19–29, 1 2007.
- [290] P. K. Whelton, R. M. Carey, W. S. Aronow, D. E. Casey, K. J. Collins, C. Dennison Himmelfarb, S. M. DePalma, S. Gidding, K. A. Jamerson, D. W. Jones, E. J. MacLaughlin, P. Muntner, B. Ovbiagele, S. C. Smith, C. C. Spencer, R. S. Stafford, S. J. Taler, R. J. Thomas, K. A. Williams, J. D. Williamson, and J. T. Wright. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Journal of the American College of Cardiology*, 71(19):e127–e248, 5 2018.
- [291] G. Wickstrom and T. Bendix. The Hawthorne effect - What did the original Hawthorne studies actually show?, 2000.
- [292] N. Wiener. *Cybernetics, or Control and Communication in the Animal and the Machine*, 1962.
- [293] Wikipedia. Wikipedia, 2020.
- [294] C. P. Wild. Complementing the genome with an ”exposome”: The outstanding challenge of environmental exposure measurement in molecular epidemiology, 8 2005.
- [295] D. Wiley and A. Cory. Health Belief Model. In *Encyclopedia of School Health*. SAGE Publications, 2013.
- [296] P. T. Williams. Physical fitness and activity as separate heart disease risk factors: A meta-analysis. *Medicine and Science in Sports and Exercise*, 33(5):754–761, 2001.
- [297] D. L. Williamson and J. Carr. Health as a resource for everyday life: Advancing the conceptualization. *Critical Public Health*, 19(1):107–122, 2009.
- [298] M. Wojciechowski. Interactive constraint-based sequential pattern mining. In *Advances in Databases and Information Systems*, pages 169–181. Springer, 2001.
- [299] J. D. Wolchok, A. Hoos, S. O’Day, J. S. Weber, O. Hamid, C. Lebbé, M. Maio, M. Binder, O. Bohnsack, G. Nichol, R. Humphrey, and F. S. Hodi. Guidelines for the evaluation of immune therapy activity in solid tumors: Immune-related response criteria. *Clinical Cancer Research*, 15(23):7412–7420, 12 2009.
- [300] K. Wolf, A. Schmidt, A. Bexheti, and M. Langheinrich. Lifelogging: You’re wearing a camera? *IEEE Pervasive Computing*, 13(3):8–12, 2014.
- [301] World and Health Org. Preamble to the Constitution, 1946.

- [302] WorldFoodAtlas.org. World Food Atlas A resource for Healthy and Happy Culinary Experience, 2020.
- [303] B. Wu, Z. Tian, and M. Chen. Condition-based maintenance optimization using neural network-based health condition prediction. *Quality and Reliability Engineering International*, 29(8):1151–1163, 12 2013.
- [304] P. Y. Wu, C. W. Cheng, C. D. Kaddi, J. Venugopalan, R. Hoffman, and M. D. Wang. -Omic and Electronic Health Record Big Data Analytics for Precision Medicine. *IEEE Transactions on Biomedical Engineering*, 64(2):263–273, 2 2017.
- [305] S. Wu, W. Zhu, P. Thompson, and Y. A. Hannun. Evaluating intrinsic and non-intrinsic cancer risk factors, 12 2018.
- [306] C. M. Wylie. The definition and measurement of health and disease. *Public health reports*, 85(2):100–104, 2 1970.
- [307] L. Xiao, S. Boyd, and S. Lall. A Scheme for robust distributed sensor fusion based on average consensus. In *2005 4th International Symposium on Information Processing in Sensor Networks, IPSN 2005*, volume 2005, pages 63–70, 2005.
- [308] L. Xie, H. Sundaram, and M. Campbell. Event mining in multimedia streams. *Proceedings of the IEEE*, 96(4):623–647, 4 2008.
- [309] K. Yanai and Y. Kawano. Food image recognition using deep convolutional network with pre-training and fine-tuning. In *2015 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2015*. Institute of Electrical and Electronics Engineers Inc., 7 2015.
- [310] X. Yang, T. Robinson, N. Ram, and B. Reeves. Using screenshots to predict task switching on smartphones. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 1–6, New York, New York, USA, 5 2019. Association for Computing Machinery.
- [311] S. Yeager. LISS Cardio — Low-Intensity Cardio for Speed.
- [312] C.-C. Yu and Y.-L. Chen. Mining sequential patterns from multidimensional sequence data. *Knowledge and Data Engineering, IEEE Transactions on*, 17(1):136–140, 2005.
- [313] L. Zadeh and C. Desoer. *Linear System Theory: The State Space Approach* -. Dover Publications, Inc., 2008.
- [314] Z. Zheng, X. Xu, X. Zhang, A. Wang, C. Zhang, M. Hüttemann, L. I. Grossman, L. C. Chen, S. Rajagopalan, Q. Sun, and K. Zhang. Exposure to ambient particulate matter induces a NASH-like phenotype and impairs hepatic glucose metabolism in an animal model. *Journal of Hepatology*, 58(1):148–154, 1 2013.

- [315] J. Zhu, A. Pande, P. Mohapatra, and J. J. Han. Using Deep Learning for Energy Expenditure Estimation with wearable sensors. In *2015 17th International Conference on E-Health Networking, Application and Services, HealthCom 2015*, pages 501–506. Institute of Electrical and Electronics Engineers Inc., 2015.

# Appendices

## A Abbreviations

Table A.1: Abbreviations

<b>Abbreviation</b>	<b>Expansion</b>
ABMS	American Board of Medical Specialties
ACM	Association for Computing Machinery
AALM	All Ages Lead Model
ABP	Athlete Biological Passport
ACC	American College of Cardiology
ACTN3	Alpha-actinin-3
AHA	American Heart Association
API	Application Programming Interface
ASCVD	Atherosclerotic Cardiovascular Disease
ATP	Adenosine Triphosphate
ATP	Association of Teachers of Preventive Medicine Foundation
BAC	Blood Alcohol Content
CDC	Centers for Disease Control and Prevention
CEDH	Center for Educational Development Health
cMRI	Cranial Magnetic Resonance Imaging

CO	Cardiac Output
COX	Cyclooxygenase
CP	Critical Power
CRF	Cardio-respiratory Fitness
CTL	Chronic Training Load
CVD	Cardiovascular Disease
DALY	Disability Adjusted Life Year
DB	Database
DEXA	Dual-energy X-ray absorptiometry
EEG	Electroencephalogram
EKG	Electrocardiogram
EMA	Ecological Momentary Assessment
EPA	Environmental Protection Agency
FHIR	Fast Healthcare Interoperability Resources
fMRI	Functional Magnetic Resonance Imaging
GDPR	General Data Protection Regulation
GIS	Geographical Information Systems
GNB	Graph Network Block
GOVVS	Gravity Ordered Velocity Stress Score
GPS	Global Positioning System
GSR	Galvanic Skin Resistance
HCI	Human-Computer-Interaction
HIPAA	Health Insurance Portability and Accountability Act Regulation
HIIT	High-Intensity Interval Training
HPA	High Power Activity
HR	Heart Rate

---

HRV	Heart Rate Variability
HSE	Health State Estimation
ICD	International Classification of Disease
ICTs	Information and Communication Technologies
ICU	Intensive Care Unit
IoT	Internet of Things
MRI	Magnetic Resonance Imaging
NIH	National Institute of Health
NLP	Natural Language Processing
NSAID	Non-Steroidal Anti Inflammatory Drugs
O <sub>2</sub>	Oxygen
PHSS	Personal Health State Space
QoL	Quality of Life
ROI	Region of Interest
SCD	Sudden Cardiac Death
SDK	Software Development Skit
SNP	Single Nucleotide Polymorphism
SV	Stroke Volume
TCA Cycle	Tricarboxylic Acid Cycle
TRIMP	Training Impulse
VAM	Vertical Ascent in Meters
VO <sub>2</sub> max	Maximum Volume of Oxygen
WHO	World Health Organization
YLL	Years of Life Lost

---

## B Resources

See next page for Table.

Table B.2: Hardware, Software, Knowledge Sources, Data Tools, and Visualization Resources

Hardware	Description	Website
Garmin	GPS multi-sport smartwatch	<a href="https://www.garmin.com/en-US/">https://www.garmin.com/en-US/</a>
Garmin	Heart rate monitor with chest strap	<a href="https://www.garmin.com/en-US/">https://www.garmin.com/en-US/</a>
Garmin	Pedaling cadence wireless sensor	<a href="https://www.garmin.com/en-US/">https://www.garmin.com/en-US/</a>
23andMe	Health and ancestry genetic reports	<a href="https://www.23andme.com/">https://www.23andme.com/</a>
Concept2	Rowing ergometer, indoor rowing machine	<a href="https://www.concept2.com/indoor-rowers">https://www.concept2.com/indoor-rowers</a>
Stages	Bicycle crank with strain gauge	<a href="https://stagescycling.com/">https://stagescycling.com/</a>
Elite	Cycling trainer that includes a power meter.	<a href="https://www.elite-it.com/it">https://www.elite-it.com/it</a>
AccuFitness	Measuring tape	<a href="https://www.accufitness.com/">https://www.accufitness.com/</a>
Elite Medical	Protractor Goniometer	<a href="https://www.elitemedicaemi.com/">https://www.elitemedicaemi.com/</a>
Plumb Bob	A perpendicular reference line to gravity	<a href="https://en.wikipedia.org/wiki/Plumb_bob">https://en.wikipedia.org/wiki/Plumb_bob</a>
<b><u>Software</u></b>		
Strava	Tracking physical activity using GPS data	<a href="https://www.strava.com/">https://www.strava.com/</a>
Zwift	Online cycling/running physical training program	<a href="https://zwift.com/">https://zwift.com/</a>
Sleep Cycle	Sleep pattern analysis and REM-based alarm clock	<a href="https://www.sleepcycle.com/">https://www.sleepcycle.com/</a>
Nomie	Mood tracking app	<a href="https://nomie.app/">https://nomie.app/</a>
Balance Health	Blood pressure data analysis	<a href="https://greatergoods.com/">https://greatergoods.com/</a>



Weight Gurus	Weight tracking app connecting data from weight scale	<a href="https://weightgurus.com/">https://weightgurus.com/</a>
Kinsa Smart	Smart body temperature readings and tracking	<a href="https://www.kinsahealth.co/">https://www.kinsahealth.co/</a>
Nuvoair	Lung function data collection and analysis	<a href="https://www.nuvoair.com/">https://www.nuvoair.com/</a>
Life Cycle	Tracking time on phone automatically	<a href="http://www.northcube.com/lifecycle/">http://www.northcube.com/lifecycle/</a>
Apple Health Kit	Health informatics mobile application and integration	<a href="https://developer.apple.com/healthkit/">https://developer.apple.com/healthkit/</a>
Sensor Push	Monitors humidity, temperature, etc.	<a href="https://www.sensorpush.com/">https://www.sensorpush.com/</a>
Lumen	Measuring metabolism through gas analyzer	<a href="https://www.lumen.me/">https://www.lumen.me/</a>
RescueTime	Tracking time on phone and computer	<a href="https://www.rescuetime.com/">https://www.rescuetime.com/</a>
Concept2	Rowing data collection application	<a href="https://www.concept2.com/">https://www.concept2.com/</a>
Amazon Alexa	Virtual assistant AI technology	<a href="https://www.amazon.com">https://www.amazon.com</a>
Nest Thermostat	Smart thermostat	<a href="https://nest.com/">https://nest.com/</a>
Nest Protect	Smart smoke detector	<a href="https://nest.com/">https://nest.com/</a>
IFTTT	Chains of simple conditional statements	<a href="https://ifttt.com/">https://ifttt.com/</a>
Phillips Hue	Wireless LED lamps	<a href="https://www2.meethue.com/">https://www2.meethue.com/</a>
Amazon	WiFi smart plugs	<a href="https://www.amazon.com">https://www.amazon.com</a>
Calorie Mama	Food image recognition app	<a href="https://www.caloriemama.ai/">https://www.caloriemama.ai/</a>
Google Maps	Satellite mapping and routing by GPS	<a href="https://www.google.com/maps">https://www.google.com/maps</a>
Epic	Medications, tests, appointments, medical bills, pricing	<a href="https://www.mychart.com/">https://www.mychart.com/</a>
BAC Track	Estimated blood alcohol level	<a href="https://www.bactrack.com/">https://www.bactrack.com/</a>

---

Hypoxico	Pulse oximeter with Bluetooth connectivity	<a href="https://hypoxico.com/">https://hypoxico.com/</a>
One Touch	Blood glucose monitor	<a href="https://www.onetouch.com/">https://www.onetouch.com/</a>
Withings Health Mate	Tracking and analysis of activity, sleep, weight, etc.	<a href="https://www.withings.com/">https://www.withings.com/</a>
Time Out	Time tracking software	<a href="https://cws-software.com/">https://cws-software.com/</a>

### **Knowledge Sources**

Cytoscape	Open source platform for visualizing complex networks	<a href="https://cytoscape.org/">https://cytoscape.org/</a>
UptoDate	Point-of-care medical, evidence-based clinical resource	<a href="https://www.uptodate.com/home">https://www.uptodate.com/home</a>
Nature Reviews	Publishing Agency	<a href="https://www.nature.com/">https://www.nature.com/</a>
PubMed	Accessing database references/abstracts on life sciences.	<a href="https://www.ncbi.nlm.nih.gov/pubmed/">https://www.ncbi.nlm.nih.gov/pubmed/</a>
DarkSky	Information on weather	
In the Sky	Information on daily sunset times	<a href="https://in-the-sky.org/sunrise.php">https://in-the-sky.org/sunrise.php</a>
Google Scholar	Web search engine of scholarly literature	<a href="https://scholar.google.com/">https://scholar.google.com/</a>
Mendeley	Reference Manager	<a href="https://www.mendeley.com/">https://www.mendeley.com/</a>
Aqicn	World-wide air quality monitoring data coverage	<a href="https://aqicn.org/sources/">https://aqicn.org/sources/</a>
EPA	Agency of the United States federal government	<a href="https://www.epa.gov/">https://www.epa.gov/</a>

### **Data Analysis**

QGIS	Open Source Geographic Information System	<a href="https://www.qgis.org/en/site/">https://www.qgis.org/en/site/</a>
------	---	---

---

Golden Cheetah	Open Source Activity Data Analysis Platform	<a href="https://www.goldencheetah.org/">https://www.goldencheetah.org/</a>
Google Sheets	Spreadsheet program	<a href="https://www.google.com/sheets/about/">https://www.google.com/sheets/about/</a>
Apple Photos	Photo download, editing platform	<a href="https://www.apple.com/ios/photos/">https://www.apple.com/ios/photos/</a>
Calorie Mama	Food image recognition and analysis	<a href="https://www.caloriemama.ai/">https://www.caloriemama.ai/</a>
Insight Maker	Creating dynamic simulation models	<a href="https://insightmaker.com/">https://insightmaker.com/</a>

### **Visualization**

Tableau	Data visualization software	<a href="https://www.tableau.com/">https://www.tableau.com/</a>
Gephi	Open-source network visualization software	<a href="https://gephi.org/">https://gephi.org/</a>
Draw.io	Network drawing software	<a href="https://www.draw.io/">https://www.draw.io/</a>
Keynote	Drawing and Presentation Software	<a href="https://www.apple.com/keynote/">https://www.apple.com/keynote/</a>

---