

# UC Irvine

## UC Irvine Previously Published Works

### Title

Bayesian deep learning for reliable oral cancer image classification.

### Permalink

<https://escholarship.org/uc/item/9qm4f2t2>

### Journal

Biomedical Optics Express, 12(10)

### ISSN

2156-7085

### Authors

Song, Bofan  
Sunny, Sumsum  
Li, Shaobai  
et al.

### Publication Date

2021-10-01


### DOI

10.1364/boe.432365

Peer reviewed



# Bayesian deep learning for reliable oral cancer image classification

BOFAN SONG,<sup>1,9</sup> SUMSUM SUNNY,<sup>2</sup> SHAOBAI LI,<sup>1</sup> KEERTHI GURUSHANTH,<sup>3</sup> PRAMILA MENDONCA,<sup>4</sup> NIRZA MUKHIA,<sup>3</sup> SANJANA PATRICK,<sup>5</sup> SHUBHA GURUDATH,<sup>3</sup> SUBHASHINI RAGHAVAN,<sup>3</sup> IMCHEN TSUSENNARO,<sup>6</sup> SHIRLEY T. LEIVON,<sup>6</sup> TRUPTI KOLUR,<sup>4</sup> VIVEK SHETTY,<sup>4</sup> VIDYA R. BUSHAN,<sup>4</sup> ROHAN RAMESH,<sup>6</sup> TYLER PETERSON,<sup>1</sup>  VIJAY PILLAI,<sup>4</sup> PETRA WILDER-SMITH,<sup>7</sup> ALBEN SIGAMANI,<sup>4</sup> AMRITHA SURESH,<sup>2,4</sup> MONI ABRAHAM KURIAKOSE,<sup>8</sup> PRAVEEN BIRUR,<sup>3,4</sup> AND RONGGUANG LIANG<sup>1,10</sup>

<sup>1</sup>Wyant College of Optical Sciences, The University of Arizona, Tucson, Arizona 85721, USA

<sup>2</sup>Mazumdar Shaw Medical Centre, Bangalore, India

<sup>3</sup>KLE Society Institute of Dental Sciences, Bangalore, India

<sup>4</sup>Mazumdar Shaw Medical Foundation, Bangalore, India

<sup>5</sup>Biocon Foundation, Bangalore, India

<sup>6</sup>Christian Institute of Health Sciences and Research, Dimapur, India

<sup>7</sup>Beckman Laser Institute and Medical Clinic, University of California, Irvine, California 92697, USA

<sup>8</sup>Cochin Cancer Research Center, Kochi, India

<sup>9</sup>songb@arizona.edu

<sup>10</sup>rliang@optics.arizona.edu

**Abstract:** In medical imaging, deep learning-based solutions have achieved state-of-the-art performance. However, reliability restricts the integration of deep learning into practical medical workflows since conventional deep learning frameworks cannot quantitatively assess model uncertainty. In this work, we propose to address this shortcoming by utilizing a Bayesian deep network capable of estimating uncertainty to assess oral cancer image classification reliability. We evaluate the model using a large intraoral cheek mucosa image dataset captured using our customized device from high-risk population to show that meaningful uncertainty information can be produced. In addition, our experiments show improved accuracy by uncertainty-informed referral. The accuracy of retained data reaches roughly 90% when referring either 10% of all cases or referring cases whose uncertainty value is greater than 0.3. The performance can be further improved by referring more patients. The experiments show the model is capable of identifying difficult cases needing further inspection.

© 2021 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](https://doi.org/10.1364/OAOpenAccessPublishingAgreement)

## 1. Introduction

Deep neural networks have gained attention in the field of medical imaging and diagnosis [1] and have been proven to be successful in detecting skin cancer [2], breast cancer [3], lung cancer [4] and oral cancer [5]. These methods are now being considered for integration into diagnostic systems [6], especially for disease screening in resource-limited settings where there is a severe shortage of trained doctors and specialists [7,8]. The reliability of automated decisions must be high for practical clinical applications [9]. However, despite the promising performance, traditional deep learning-based classification still lacks the ability of human physicians to quantify decision uncertainty. Without uncertainty measurements, clinicians cannot rely on decisions from deep learning based automatic systems in practical clinic routines. No matter how powerful

a deep learning classifier is, difficult diagnostic cases are inevitable and may lead to potentially severe consequences for patients if the model does not refer them for further inspection.

Machine learning based automatic oral cancer diagnostic systems are typically evaluated by accuracy, sensitivity, specificity, receiver operating characteristic (ROC) curve, and area under ROC curve (AUC) [10–12], but less attention has been paid to assessing confidence in neural network decisions. Previous models fail to know how confident the model is about a specific output. In this work, we propose a deep learning oral cancer image classification framework that can quantify the model's output uncertainty and suggest that difficult cases with high uncertainty value be referred for further examination. This deep learning classifier with uncertainty estimations could be integrated to assist and accelerate traditional clinical workflows, not replace them. Deep learning approaches are often seen as a 'black-box', but the methods could become more reliable and trustable by providing uncertain information and potentially increasing overall performance. If so, the automatic classifier could be considered as a tireless front-line doctor that needs no rest, directly providing diagnoses when confident and referring difficult cases to experienced specialists when uncertain.

Bayesian deep learning integrates deep learning with Bayesian probability theory. Bayesian neural networks provide a prediction and an uncertainty value by imposing prior distributions on the model parameters (weights) to obtain a posterior distribution of these parameters [13]. The uncertainty is estimated through a probability density over outcomes, while traditional deep learning only produces a deterministic output. Bayesian neural networks (BNN) have been studied for a long time [14,15] but are not widely used because of massive training times and implementation difficulties. Gal et al. [16] proposed a new framework that takes Monte Carlo samples of the prediction by training a dropout network that approximates Bayesian inference in deep Gaussian processes. This approach produces the posterior distribution of the network weights and estimates uncertainty in a more straightforward way. Some pioneering works in the emerging field of Bayesian deep learning in imaging systems [17,18] demonstrate that uncertainty analysis is necessary for scientific imaging and diagnosis where critical assessment is essential.

Oral cancer is one of the most common cancers, especially in low- and middle-income countries like India [19]. Early detection is the most effective way to reduce the mortality rate [20], but unfortunately, most high-risk populations lack healthcare infrastructure and doctors. Therefore, there is an urgent need for a reliable automatic detection system for large-scale, high-risk population screening. In this work, we developed a Bayesian deep neural network for oral cancer detection and trained it using our intraoral image dataset collected from a high-risk population in India. The results show that meaningful uncertainty information can be obtained and used to identify difficult cases in need of further examination effectively. By referring these cases for review, the overall detection performance of the trained network can be further improved.

## 2. Material and methods

Bayesian deep learning is an effective method to add uncertainty handling in deep learning models. It combines Bayesian probability theory with deep learning to extend standard neural networks by assigning distributions to their weights. Traditional deep networks have fixed weights, while the weights of Bayesian networks are assigned a probability distribution. Therefore, a standard deep network with fixed weights will always give the same outputs, but a Bayesian neural network will deliver stochastic outputs.

When conventional deep neural networks trained with cross-entropy loss classify  $N$  instances  $\{x_1, x_2 \dots x_N\}$  with corresponding labels  $\{y_1, y_2 \dots y_N\}$  to  $K$  classes using the Softmax model, cross-entropy minimizes the distance between the true class label distribution and the approximating distribution from the model, resulting in a single best set of parameters  $\omega$  and a network function  $\varphi$  correlated with those parameters. The probability that the instance  $x_i$  belongs to class

k is calculated as:

$$p(y_i = k|x_i, \omega) = \frac{e^{\varphi(x_i)}}{\sum_{j=1}^k e^{\varphi(x_j)}} \quad (1)$$

The probability value generated by softmax could be considered as the multiplication of features extracted by the convolutional layers and the weight for the features of each category. The result compares how many features are found within each category, and the sum of all probability values is 1.

Bayesian deep neural networks consider a distribution over network parameters instead of a single best set. The predictive posterior probability distribution calculated using the Bayesian deep learning model for a new instance  $x^*$  is:

$$p(y|x^*, X, Y) = \int p(y|x^*, \omega)p(\omega|X, Y)d\omega \quad (2)$$

The probability that a case belongs to a specific class is a single value produced by Eq. (1), while the predictive posterior defined by Eq. (2) is a distribution that describes all possible predictions given the network weights and test instance  $x^*$ . The width of the predictive posterior distribution would be able to reflect the model's confidence about a specific prediction.

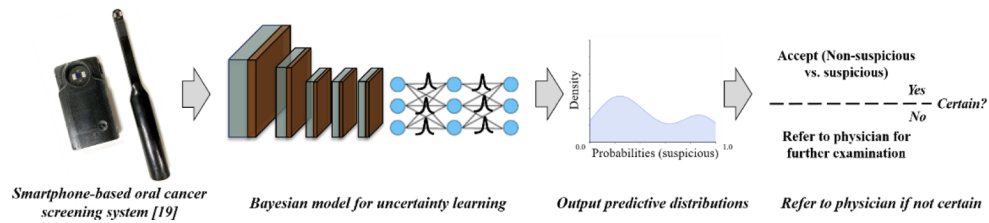
Dropout is a well-established procedure to regularize a deep learning model and reduce overfitting by randomly dropping neurons in a neural network layer. Since this technique can also be interpreted as a Bayesian approximation of a Gaussian process, model uncertainty can be obtained from dropout neural network models. This approach, called Monte Carlo (MC) dropout, mitigates the problem of representing uncertainty in deep learning without sacrificing either computational complexity or test accuracy. By running the Monte Carlo dropout model multiple ( $\rho$ ) times to obtain several stochastic outputs, the final prediction on a test instance (predictive mean) can be calculated using a Monte Carlo integration over  $\rho$  samples:

$$p(y|x^*, X, Y) \approx \frac{1}{\rho} \sum_{\rho=1}^{\rho} p(y|x^*, \omega^{(\rho)}) \quad (3)$$

The variance of the distributions from the ensembles will be considered as the model uncertainty on the prediction:

$$v = \frac{1}{\rho} \sum_{\rho=1}^{\rho} (p(y|x^*, \omega^{(\rho)}) - p(y|x^*, X, Y))^2 \quad (4)$$

The intraoral dataset used in this study contains 2350 cheek mucosa images that were captured using a smartphone-based intraoral screening device we developed [21], and among patients attending the outpatient clinics of the Department of Oral Medicine and Radiology at KLE Society Institute of Dental Sciences, Head and Neck Oncology Department of Mazumdar Shaw Medical Center (MSMC), and Christian Institute of Health Sciences and Research (CIHSR), India. Oral oncology specialists from MSMC, KLE and CIHSR labeled all the dual-modality image pairs and separated them into two categories: A) 'Normal' which contains normal and benign mucosal lesion images, and B) 'suspicious' which contains OPML and malignant lesion images. In a previous study, we show that oral oncology specialists' interpretation of classifying normal/benign vs. OPML/malignant has high accuracy with biopsy-confirmed cases [22]. The intraoral cheek mucosa dataset (normal/benign: 1510; suspicious: 840) were randomly split to training, validation, and standalone test. We used 1979 (normal/benign: 1272; suspicious: 707) images for training and validation (randomly split where 75% was allocated to training and 25% was allocated to validation) and another 371 (normal/benign: 238; suspicious: 133) images for the standalone test. The proposed Bayesian deep learning model for oral cancer detection is shown in Fig. 1.



**Fig. 1.** The proposed Bayesian deep learning model for oral cancer detection.

### 3. Experiments and results

We trained the Bayesian deep neural network for oral cancer images using VGG19 as a base network; the initial weights of the network were obtained from training the ImageNet dataset. Two dropout layers with a 0.5 rate were applied to the fully connected layers, which means half of the units in these layers were turned off during training and inference. The dropout in the network not only can be treated as a Bayesian approximation of a Gaussian process, but also as a means to reduce overfitting. Data augmentation was applied to the training set to over-sample the dataset based on the ratios of imbalanced classes by flipping horizontally and vertically, random rotating, and shearing. Our augmentation strategy increased the instances number of minority classes and could reduce the model bias toward the majority groups.

Adam optimizer was used to minimize the Focal loss function. Focal loss can reduce the class imbalance problem and hard sample issues. The initial learning rate was  $1e-4$  and decay 5 times by every 20 epochs. The experiment was run for 300 epochs, and the batch size was set to 32. The model with the best validation accuracy was saved. The model was trained on the high-performance computing platform of the University of Arizona [23]. The predictive means calculated by Eq. (3) will be used as final predictions by the Bayesian neural network. The predictive standard deviation calculated by Eq. (4) is the associated uncertainty value; we set  $\rho$  as 50 in our experiments. The structure of our proposed architecture and was shown in Fig. 2.

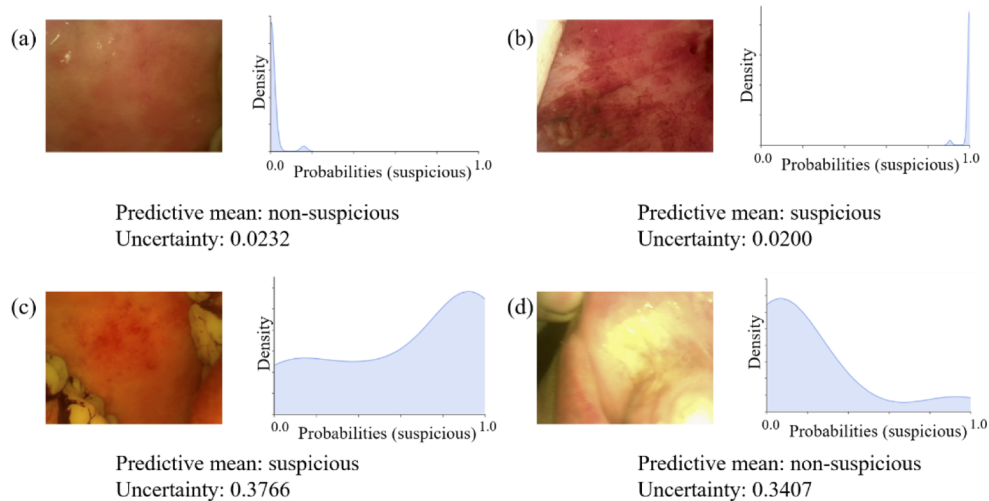


**Fig. 2.** Structure of our BNN architecture (VGG19 as baseNet).

We evaluated the Bayesian deep neural network's intraoral cancer image classification accuracy using a standalone dataset containing 371 intraoral cheek mucosa images in the test set. The results were compared with a conventional network with standard dropout. The BDNN achieved 85.6% accuracy, while the accuracy of the traditional network with standard dropout was 85.1%. The result shows good classification performance of the Bayesian network and indicates the BNN didn't sacrifice accuracy but slightly improved the performance by ensemble learning.

Figure 3 shows examples of the BNN prediction with uncertainty estimation. For a specific input case, the network can be certain (Fig. 3(a) and 3(b)) or uncertain (Fig. 3(c) and 3(d)) about its prediction, as indicated by the standard deviation of the predictive posterior distribution calculated with Eq. (4). The image of Fig. 3(a) was confidently classified as non-suspicious by the BNN since all the sampled probability (suspicious) predictions are near 0.0, and the uncertainty value (standard deviation) is 0.0232. The image of Fig. 3(b) was confidently classified as suspicious by the BNN since all the sampled probability (suspicious) predictions are near

1.0, and the uncertainty value is 0.0200. Whereas Fig. 3(c) shows an example of the model is uncertain: the predicted label (predictive mean) is correct, but with wider posterior distribution, and the uncertainty value is 0.3766, which may be because the lesion is not obvious in this case. Figure 3(d) shows another example where the BNN is highly uncertain: the uncertainty value is 0.3407, caused by image quality (over-exposed). These examples indicate the BNN can produce informative uncertainty estimation.

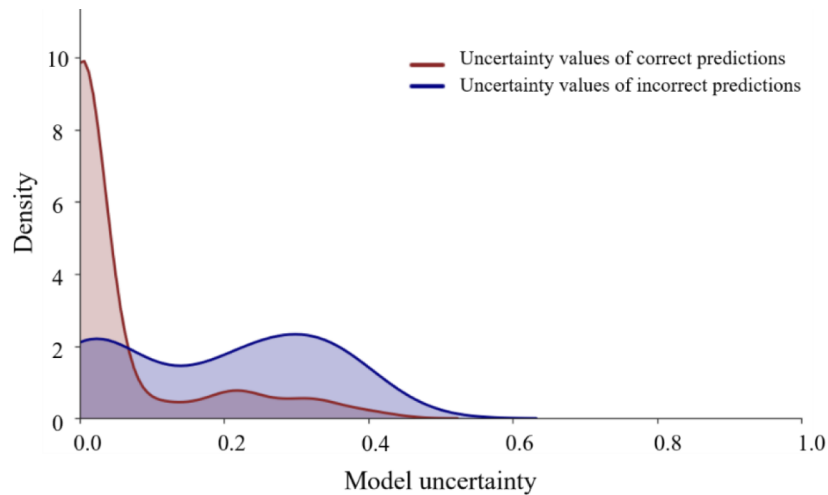


**Fig. 3.** Examples of intraoral images and corresponding approximate predictive probability (suspicious) posteriors distribution produced by the Bayesian neural network. (The density (vertical axis) is based on Gaussian kernel density estimates.)

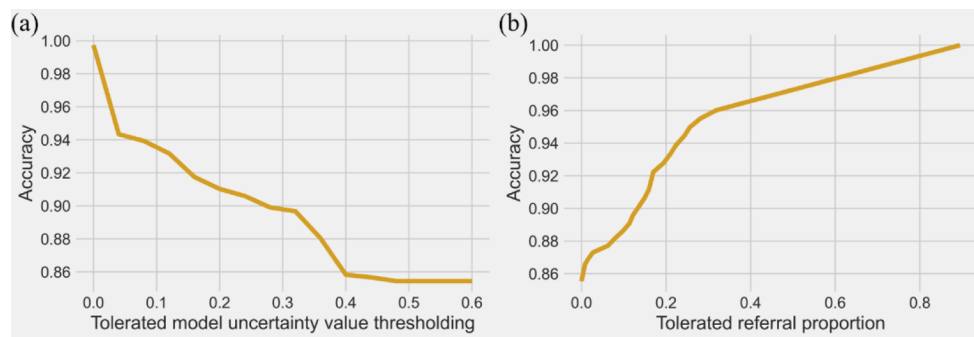
We were also curious if most misclassified cases have higher model uncertainty values than correct predictions, in which case they could be flagged for further inspection. The uncertainty values for all 371 standalone test images were plotted using Kernel Density Estimation with a Gaussian kernel and grouped by correct and incorrect predictions (see Fig. 4). The result shows incorrect cases have higher model uncertainty. Therefore, the correlated uncertainty value could be used to find difficult cases in need of further examination to improve the overall accuracy. The results also indicate the BNN model has the potential to mimic the human clinical workflow and be integrated into diagnostic systems.

We then sorted the predictions of the test images according to the uncertainty value from high to low and measured the change of accuracy when referring cases with uncertainty values higher than a specific level. By switching the level of uncertainty thresholding, we plotted the change of accuracy in Fig. 5(a). We used this experiment to test whether the proposed method could refer uncertain patients for further inspection and therefore mimic the clinic workflow. From the figure, we can see a continuous increase of accuracy in response to a decrease of uncertainty thresholding. The accuracy will be around 90% if the model refers cases with uncertainty higher than 0.3. We wanted to verify this increase in accuracy was not simply due to the model referring too many cases, which would render it effectively useless. Therefore, we monitored the change of accuracy when referring different proportions of cases (cases with larger uncertainty values were referred first). With the increase of referral proportion thresholding, we also observed an increase in prediction performance (see Fig. 5(b)). The figure shows the accuracy will be around 90% if the model refers 10% of the cases.

In addition, we noticed that there are cases where the probability value is distributed around 0.5 with low uncertainty, indicating that the network has extracted features correlated to both



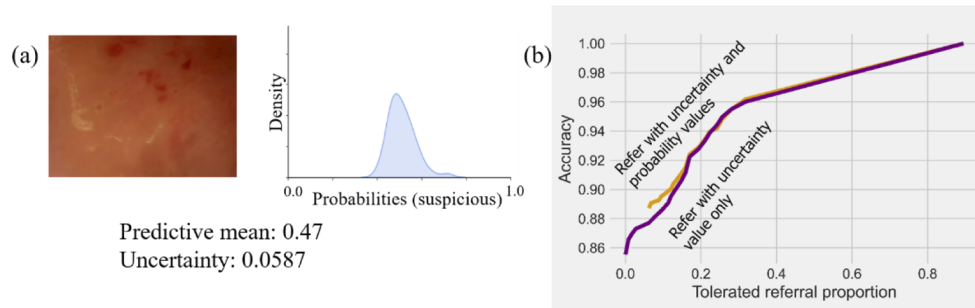
**Fig. 4.** Distribution of uncertainty value for all test data grouped by correct and incorrect model predictions. (The density (vertical axis) is based on Gaussian kernel density estimates.)



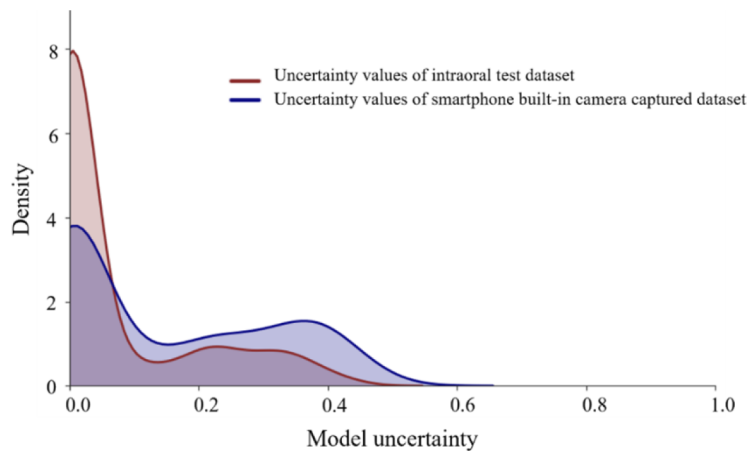
**Fig. 5.** The change of accuracy correlated to the tolerated (a) model uncertainty value, (b) referral proportion.

categories (See Fig. 6(a)). Although the network is certain about the result (low uncertainty value), it cannot provide a precise diagnosis (probability value around 0.5). Therefore, we tried to combine the uncertainty and mean probability values for the referral strategy and re-drew Fig. 5(b). We tried to refer cases with high uncertainty values and mean probability values between 0.4 to 0.6, and monitored the change of accuracy by referring varying proportions of cases. The results (Fig. 6(b)) show that this new referral strategy achieves higher accuracy for the same referral proportion. But no difference was observed with increasing proportion, because the cases referred due to the mean probability value are included in the uncertainty value referral.

Theoretically, the Bayesian neural network should produce higher uncertainty values on unfamiliar data. The model uncertainty should be able to identify data different from the training dataset to sort out abnormal or unusable data. We ran the model trained with intraoral cheek mucosa images on another cheek mucosa test dataset which was captured using a smartphone's built-in camera ( $n=351$ ). The uncertainty values for both the intraoral cheek mucosa test dataset and the dataset captured with smartphone built-in camera were plotted using Kernel Density Estimation with a Gaussian kernel in Fig. 7. We can see the intraoral cheek mucosa test dataset had lower average uncertainty. This result indicates the model was more confident with a familiar dataset.



**Fig. 6.** (a) Example of intraoral image and corresponding approximate predictive probability (suspicious) posteriors distribution with low uncertainty value and probability values are distributed around 0.5. (b) The change of accuracy correlated to the tolerated referral proportion with two referral strategies.



**Fig. 7.** Distribution of uncertainty value for familiar intraoral probe test dataset and unfamiliar test data captured with smartphone built-in camera. (The density (vertical axis) is based on Gaussian kernel density estimates.)

#### 4. Conclusion

We have presented a Bayesian deep learning-based framework to estimate model uncertainty for intraoral cancer images. The classification accuracy achieved by the proposed model was about 85% on the standalone test dataset, which is comparable to results from a traditional deep learning framework. The model was able to measure the uncertainty of each prediction and identify difficult cases in need of further examination. In other words, the reliability of BNN predictions could be assessed. Therefore, the reliability and overall performance of the model were improved. Our experimental results show that this model produced higher uncertainty values on incorrect predictions and achieved higher accuracy by referring cases with low confidence. We have monitored the change of accuracy with different levels of tolerated model uncertainty and different levels of referral proportion. In this process of imitating traditional medical workflow, we have observed a continuous improvement. In addition, we have shown that the model can sort out unfamiliar and difficult data. The method enables users to know when they can trust the output of a network. We believe this framework is a step forward in establishing accurate and reliable deep learning-based oral cancer detection and increasing the acceptance of deep learning integration into clinical practice in high-risk population screening.



**Funding.** National Institute of Biomedical Imaging and Bioengineering (UH2EB022623); National Institute of Cancer (UH3CA239682); National Institute of Dental and Craniofacial Research (R01DE030682) of the National Institutes of Health (NIH).

**Disclosures.** The authors declare that there are no conflicts of interest related to this article.

**Data availability.** Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

## References

1. J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O'Donoghue, 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. Tufail, 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," *Nat. Med.* **24**(9), 1342–1350 (2018).
2. 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 (2017).
3. B. Ehteshami Bejnordi, M. Veta, P. Johannes van Diest, B. van Ginneken, N. Karssemeijer, G. Litjens, and J. A. W. M. van der Laak, and C. C. and the, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA* **318**(22), 2199–2210 (2017).
4. N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A. L. Moreira, N. Razavian, and A. Tsirigos, "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning," *Nat. Med.* **24**(10), 1559–1567 (2018).
5. B. Song, S. Sunny, R. D. Uthoff, S. Patrick, A. Suresh, T. Kolar, G. Keerthi, A. Anbarani, P. Wilder-Smith, M. A. Kuriakose, P. Birur, J. J. Rodriguez, and R. Liang, "Automatic classification of dual-modality, smartphone-based oral dysplasia and malignancy images using deep learning," *Biomed. Opt. Express* **9**(11), 5318–5329 (2018).
6. F. Wang, L. P. Casalino, and D. Khullar, "Deep learning in medicine—promise, progress, and challenges," *JAMA Intern Med* **179**(3), 293–294 (2019).
7. L. Hu, D. Bell, S. Antani, Z. Xue, K. Yu, M. P. Horning, N. Gachuhi, B. Wilson, M. S. Jaiswal, B. Befano, L. R. Long, R. Herrero, M. H. Einstein, R. D. Burk, M. Demarco, J. C. Gage, A. C. Rodriguez, N. Wentzensen, and M. Schiffman, "An observational study of deep learning and automated evaluation of cervical images for cancer screening," *JNCI: Journal of the National Cancer Institute* **111**(9), 923–932 (2019).
8. K. de Haan, H. Ceylan Koydemir, Y. Rivenson, D. Tseng, E. Van Dyne, L. Bakic, D. Karınca, K. Liang, M. Ilango, E. Gumustekin, and A. Ozcan, "Automated screening of sickle cells using a smartphone-based microscope and deep learning," *npj Digit. Med.* **3**(1), 76 (2020).
9. D. S. Kermany, M. Goldbaum, W. Cai, C. C. S. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, J. Dong, M. K. Prasadha, J. Pei, M. Y. L. Ting, J. Zhu, C. Li, S. Hewett, J. Dong, I. Ziyar, A. Shi, R. Zhang, L. Zheng, R. Hou, W. Shi, X. Fu, Y. Duan, V. A. N. Huu, C. Wen, E. D. Zhang, C. L. Zhang, O. Li, X. Wang, M. A. Singer, X. Sun, J. Xu, A. Tafreshi, M. A. Lewis, H. Xia, and K. Zhang, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell* **172**(5), 1122–1131.e1129 (2018).
10. X. Zheng, Z. Yao, Y. Huang, Y. Yu, Y. Wang, Y. Liu, R. Mao, F. Li, Y. Xiao, Y. Wang, Y. Hu, J. Yu, and J. Zhou, "Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer," *Nat. Commun.* **11**(1), 1236 (2020).
11. N. Wentzensen, B. Lahrmann, M. A. Clarke, W. Kinney, D. Tokugawa, N. Poiras, A. Locke, L. Bartels, A. Krauthoff, J. Walker, R. Zuna, K. K. Grewal, P. E. Goldhoff, J. D. Kingery, P. E. Castle, M. Schiffman, T. S. Lorey, and N. Grabe, "Accuracy and efficiency of deep-learning-based automation of dual stain cytology in cervical cancer screening," *JNCI: Journal of the National Cancer Institute* **113**(1), 72–79 (2021).
12. M. Aubreville, C. Knipfer, N. Oetter, C. Jaremenko, E. Rodner, J. Denzler, C. Bohr, H. Neumann, F. Stelzle, and A. Maier, "Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning," *Sci. Rep.* **7**(1), 11979 (2017).
13. Z. Ghahramani, "Probabilistic machine learning and artificial intelligence," *Nature* **521**(7553), 452–459 (2015).
14. J. S. Denker and Y. LeCun, "Transforming neural-net output levels to probability distributions," in *Proceedings of the 3rd International Conference on Neural Information Processing Systems*, (Morgan Kaufmann Publishers Inc., 1990), pp. 853–859.
15. D. J. C. MacKay, "A practical Bayesian framework for backpropagation networks," *Neural Computation* **4**(3), 448–472 (1992).
16. G. Yarín and G. Zoubin, "Dropout as a Bayesian approximation: representing model uncertainty in deep learning," PMLR, pp. 1050–1059.
17. Y. Xue, S. Cheng, Y. Li, and L. Tian, "Reliable deep-learning-based phase imaging with uncertainty quantification," *Optica* **6**(5), 618–629 (2019).
18. R. Liu, S. Cheng, L. Tian, and J. Yi, "Deep spectral learning for label-free optical imaging oximetry with uncertainty quantification," *Light: Sci. Appl.* **8**(1), 102 (2019).

19. P. E. Petersen, "Oral cancer prevention and control—the approach of the World Health Organization," *Oral Oncology* **45**(4-5), 454–460 (2009).
20. A. G. Zygogianni, G. Kyrgias, P. Karakitsos, A. Psyri, J. Kouvaris, N. Kelekis, and V. Kouloulis, "Oral squamous cell cancer: early detection and the role of alcohol and smoking," *Head Neck Oncol* **3**(1), 2 (2011).
21. D. U. Ross, S. Bofan, S. Sumsum, P. Sanjana, S. Amritha, K. Trupti, G. Keerthi, W. Kimberly, G. Vishal, E. P. Mary, K. S. Anurag, W.-S. Petra, K. Moni Abraham, B. Praveen, and L. Rongguang, "Small form factor, flexible, dual-modality handheld probe for smartphone-based, point-of-care oral and oropharyngeal cancer screening," *J. Biomed. Opt.* **24**, 1–8 (2019).
22. N. Birur, K. Gurushanth, S. Patrick, S. Sunny, S. Raghavan, S. Gurudath, U. Hegde, V. Tiwari, V. Jain, M. Imran, P. Rao, and M. Kuriakose, "Role of community health worker in a mobile health program for early detection of oral cancer," *Indian J Cancer* **56**, 107–113 (2019).
23. <https://it.arizona.edu/service/high-performance-computing>.