An Ensemble Deep Neural Network Approach for Oral Cancer Screening

https://doi.org/10.3991/ijoe.v17i02.19207

Nanditha B R (✉), Geetha Kiran A, Chandrashekar H S
Malnad College of Engineering, Karnataka, India
brn@mcehassan.ac.in

Dinesh M.S.
PET Research Centre, Karnataka, India

Murali S
Maharaja Institute of Technology, Karnataka, India

Abstract—One of the ways to reduce oral cancer mortality rate is diagnosing oral lesions at initial stages to classify them as precancerous or normal lesions. During routine oral examination, oral lesions are normally screened manually. In a low resource setting area where there is lack of medical facilities and also medical expertise, an automated mechanism for oral cancer screening is required. The present work is an attempt towards developing an automated system for diagnosing oral lesions using deep learning techniques. An ensemble deep learning model that combines the benefits of Resnet-50 and VGG-16 has been developed. This model has been trained with an augmented dataset of oral lesion images. The model outperforms other popularly used deep learning models in performing the classification of oral images. An accuracy of 96.2%, 98.14% sensitivity and 94.23% specificity was achieved with the ensemble deep learning model.

Keywords—Oral lesions, deep learning, ResNet-50, VGG-16, ensemble model, benign, malignant

1 Introduction

Cancer has become a serious global health disorder that has a high incidence of mortality. Oral cancer takes the sixth position among other types of cancers like skin, lung, and prostate cancer. Worldwide, nearly 6.5 lakh people are diagnosed with oral cancer and about 3 lakh deaths are reported annually [1]. There are many factors that cause cancer, some of which include smoking, alcohol consumption, tobacco chewing and radiation [2]. The USA [3] has 67% more survival rate in the last five years and India [4] has 37%. Detection and treatment at an early stage has resulted in an increase in oral cancer survival rate [5] in India. The 5-year survival rate of patients diagnosed with oral cancer at an early stage is 82%, and that of patients diagnosed in later stages is 27% in India [6]. In a resource limited country like India, where people in rural and remote areas are not able to get immediate access to medical facilities, an oral screening mechanism that can be used for mass screening purposes needs to be developed. This mass screening mechanism should be able to diagnose early malignancies and refer the patient to experts for further diagnosis and treatment. For diagnosing oral cancer at an early stage, an automated system needs to be developed that minimizes human intervention. Researchers have developed many systems for oral lesion analysis to classify them as benign or malignant lesions. Various imaging modalities have been built that can capture images of oral lesions in various forms. A dual mode imaging system that combines autofluorescence and white light imaging for detection of oral malignancies has been proposed in [7]. A deep learning model has been used for the classification that results in good performance. A multimodal imaging system that combines fluorescence imaging, white light imaging and microendoscopic images for oral cancer detection has been proposed by the authors of [8]. Various image processing algorithms have been used for the analysis of suspicious regions in the oral images. Authors in [9] have tried to classify oral lesions into precancerous and normal lesions by analyzing the histopathological images obtained using biopsy of the suspected area. They have used classifiers like support vector machine for training them with image texture features. Digital true color images captured using mobile camera have been used in [10] for classifying oral lesions into benign and malignant lesions. They have used first order and second order statistical features and texture features for performing the classification. Autoflourescence images of oral cavity have been analysed using artificial neural networks for detection of oral leukoplakia in [11]. Oral neoplasia have been diagnosed using the fluorescent properties of oral images in [12]. Machine learning has proven to be useful in such automated systems to improve classification accuracy; especially deep learning minimizes the human efforts for large datasets [13]. Deep learning has advantages over conventional machine learning approaches [14] like support vector machine, nearest neighbor classifiers and decision tree classifiers. The automated feature extraction capacity of deep learning minimizes the human efforts in performing various tasks [15]. It generates three different levels of features namely, low, medium and high-level features. Also, the accuracy of deep learning techniques is found to be considerably good for image classification purposes [16]. Many deep learning techniques have been developed for oral cancer detection. Authors in [17] have used two deep learning approaches over annotated oral images for classifying oral lesions. Marc Aubreville et.al, [18] have analysed confocal laser endomicroscopic images to diagnose oral squamous cell carcinoma using deep learning techniques. Gupta et.al, [19] have made use of a deep architecture for diagnosis of dysplasia in microscopic images. Orofacial disease classification using deep learning architectures have also been attempted by the authors of [20]. Most of the existing methods for screening of oral cancer are based on analysis of microscopic images [21], autofluorescence images and hyperspectral images [22]. Since white light images can be easily captured with devices like mobile phone cameras and intraoral cameras [23], an automated system for oral lesion analysis in white light images is the need of the hour.

The rest of the paper is structured as follows: Section II portrays the proposed technique. Section III gives the discussion of results. Section IV provides conclusion of the work.

2 Proposed Method

2.1 Deep learning model

An ensemble deep learning model to classify oral lesions into premalignant and normal lesions from digital true color images has been proposed in the present work. Two deep learning models have been considered - residual network-50 and skip connected VGG. These two models have been ensembled for better classification results. The combination of these networks helps to differentiate the images based on extracted features and classifiers. There are two advantages of this proposed method. Firstly, the combination of these networks almost extracts all useful features. If any of the required features are missed out by one of the networks, the other one will extract the missed information. Second, the addition of skip connection in the transfer learning VGG network makes the feature extraction step more effective. The proposed architecture is elucidated in Fig. 1.

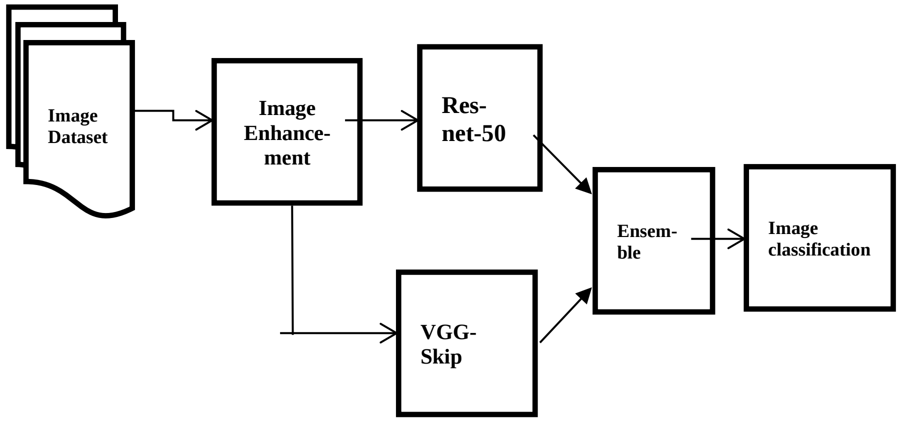

Fig. 1. Proposed Ensemble model

2.2 Collection of datasets and preprocessing of images

Oral lesion images have been collected from different colleges and hospitals in Karnataka, India. The images have been captured using mobile phone cameras and intraoral cameras. A total of 332 oral lesion images have been collected out of which 63 are benign lesions and 269 are precancerous lesions.

Since the images captured are of varying sizes and resolution, they need to be preprocessed for a uniform dimension and good resolution. Preprocessing is done both in spatial domain and frequency domain and the output of these is combined to get better quality images.

Image enhancement is done by discrete wavelet transform followed by adaptive histogram equalization [24]. Initially, wavelet transform is applied on the input image. Then, adaptive histogram equalization is applied for all the subbands separately.

Finally, reconstruction is done to get a better enhanced image. Then the same image is gamma-corrected to obtain better images. A fusion technique is applied to obtain the better-enhanced pixels. Finally, to normalize the image before being fed into the model, min-max normalization is applied according to eqn(1). Image preprocessing steps have been depicted in Fig. 2.

$$z = \frac{x - min(x)}{max(x) - min(x)} \tag{1}$$

Here, x represents input value and z represents normalized value.

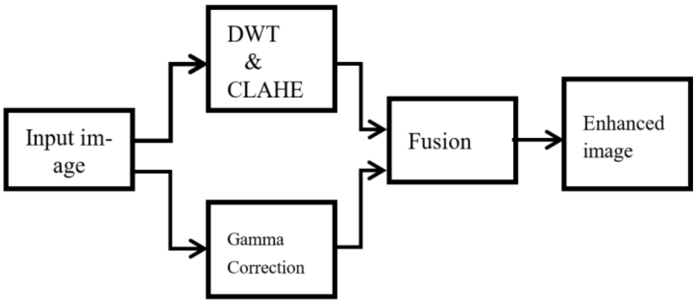

Fig. 2. Image Preprocessing

The results of preprocessing are shown in Fig. 3.

(a) Benign lesion

(b) Enhanced benign lesion

(c) Malignant lesion

(d) Enhanced malignant lesion

Fig. 3.

2.3 ResNet-50 architecture

The residual network is a powerful deep learning network mostly used for image classification [25]. The main advantage of this network is that it consists of a skip connection i.e., the previous layer output is connected to the next layer output to avoid the vanishing gradients. So, if the present layer does not effectively work for feature extraction, at least the previous layer output can go to the next layer. This helps to improve the feature extraction process. Fig. 4 depicts the architecture of ResNet-50. The skip connection for the architecture is shown in Fig. 5.

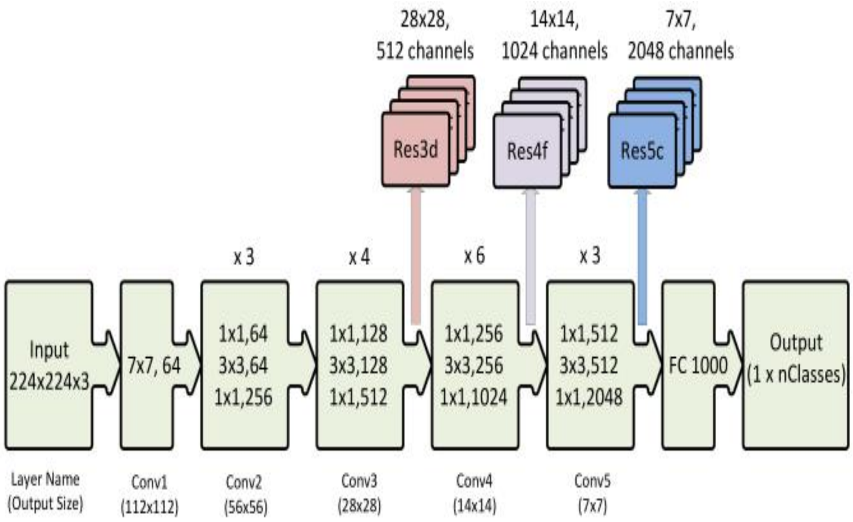

Fig. 4. Architecture of ResNet-50

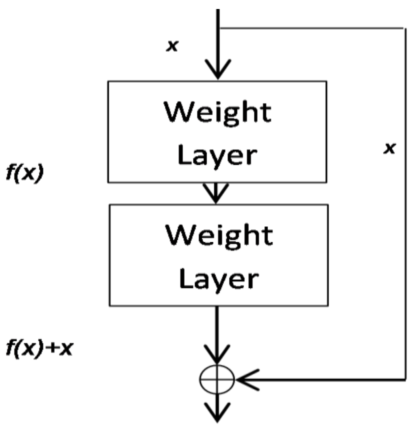

Fig. 5. Skip connection of ResNet-50

The architecture begins with a convolution layer that contains 64 filters of size 7x7, and stride 2. Next, comes the max-pooling layer with stride 2. Following this is a layer that contains 64 filters followed by 256 filters respectively. This is repeated for 3 times. The next layer contains 128, 128, 512 filters. This is repeated four times. Then, 256, 256, 1024 filters repeated six times. Finally, 512, 512, 2048 layers repeated 3 times. So, totally 50 convolution layers are present and the final layer is the classification layer. To get better classification results, ResNet-50 has to be fine-tuned. Here, the final fully connected layer is replaced by two layers for two different classes, and the input layer is tuned to accept images of size 300X300.

2.4 VGG-16 architecture

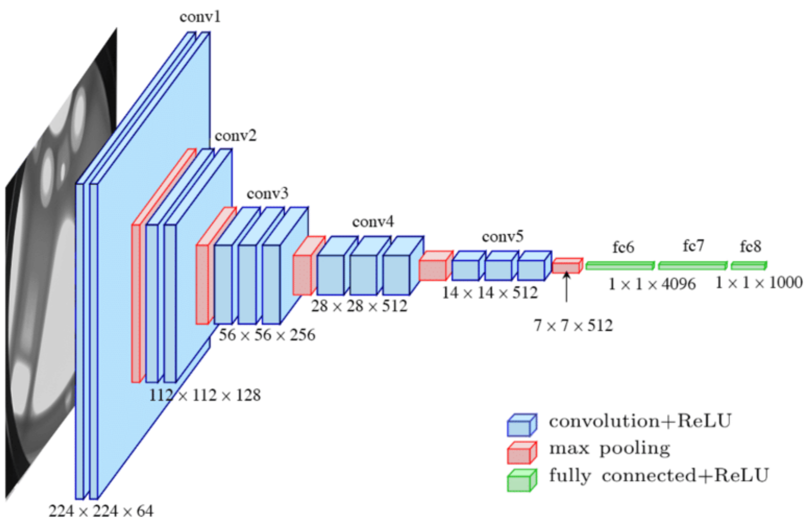

Fig. 6. Architecture of VGG-16

VGG is a complex powerful network. It is the combination of 16 convolution layers. The number of filters in each convolution layer changes from 64 to 512. The architecture of VGG-16 is shown in Fig. 6. Because of its complex architecture, there is a possibility that VGG may overlearn the features, which leads to getting null values or zeros. To avoid this, the VGG-16 model has been fine-tuned by applying the skip connections for the last two sets of convolution layers. This reduces the risk of getting null features. The skip-connected VGG is depicted in Fig. 7. Skip-connected VGG is designed based on existing VGG architecture and short connections. Here the previous convolution layer output is connected to the next layer output, so that there is a possibility of neglecting the null feature map outcome. VGG-16 has totally 16 convolution layers. First four convolution layers use 64, 64 and 128,128 filters respectively. Here, the number of filters is less compared to consecutive convolution layers. The fifth, sixth and seventh convolution layers have 256 filters respectively. As the number of filters increases, there is a chance of vanishing gradients problem. To avoid this, short connections have been performed at the fifth, sixth, seventh convolution layers respectively. Similarly, the remaining convolution layers have 512 filters. So, here also short connections have been made to avoid the zero feature maps during the time of training for better classification. After the construction of these two models, the output from each of these has been combined. Results show that ensembling of these two networks outperforms individual networks.

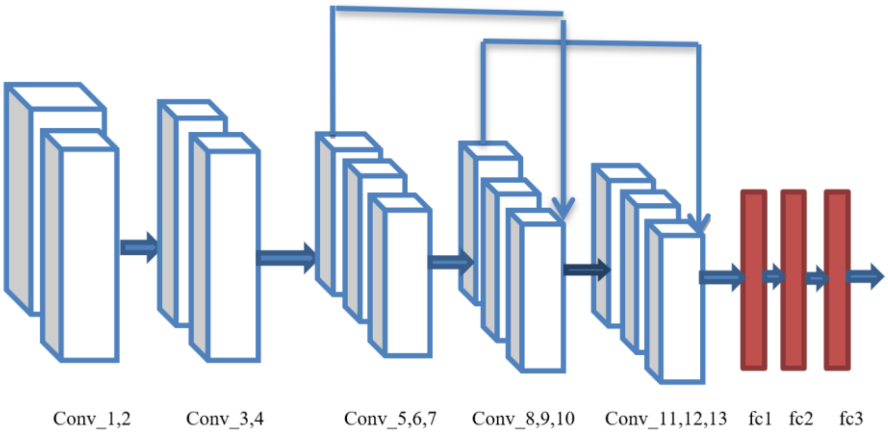

Fig. 7. Skip-connected VGG

3 Results and Discussion

The proposed ensemble method performs well compared to other conventional methods. The accuracy, sensitivity and specificity measures are considered for performance analysis.

True- Positive (TP) = Sum of positively identified samples

False-Positive (FP) = Sum of negative identified samples

True-Negative (TN) = Sum of positively identified samples identified as negative

False-Negative (FN = Sum of negative identified samples identified as positive

Accuracy = $(TP + TN)/(TP + TN + FP + FN)$

Specificity = $TN/(TN + FP)$

The augmented dataset is given as input for all networks. The experiment is carried out using MATLAB 2019 in an NVIDIA GEFORCE GPU. 80% of data is used for training and 20% is used for testing.

Table 1. Hybrid Preprocessing Performance

Model	Accuracy%	Sensitivity%	Specificity%
VGG-16	87.5	89.42	86.35
ResNet-50	93	95.52	91.62
Skip-VGG	89.5	91.41	87.71
Ensemble	**96.2**	**98.14**	**94.23**

Table 1 presents the performance of the system when hybrid preprocessing technique is used. The ensemble method performs well compared to other individual methods. It gives an accuracy of 96.2 %.

Table 2. Spatial Domain Preprocessing Performance

Model	Accuracy%	Sensitivity%	Specificity%
VGG-16	86.2	88.9	84.5
ResNet-50	89.52	91.2	88.34
Skip-VGG	88.43	90.3	86.24
Ensemble	**94.3**	**96.3**	**92.3**

Table 2 presents the performance of the system when only spatial domain preprocessing technique is used. The ensemble method performs well compared to other individual methods. It gives an accuracy of 94.3 %.

Table 3. Frequency Domain Preprocessing Performance

Model	Accuracy%	Sensitivity%	Specificity%
VGG-16	86.52	87.19	85.42
ResNet-50	92.62	93.71	91.52
Skip-VGG	91.51	92.34	90.82
Ensemble	**95.1**	**96.8**	**94.8**

Table 3 presents the performance of the system with frequency domain preprocessing. Results show that the ensemble method performs better compared to other individual models. It gives an accuracy of 95.1 %.

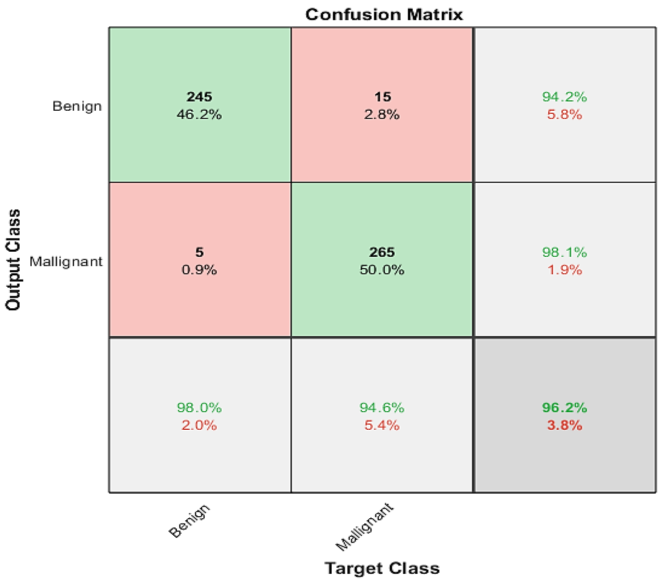

Fig. 8. Confusion matrix for proposed ensemble method

Fig. 8 shows the confusion matrix for the proposed ensemble model. The accuracy for the benign class is 98% and the accuracy for the malignant class is 94.6%. The overall accuracy of the proposed model is 96.2%.

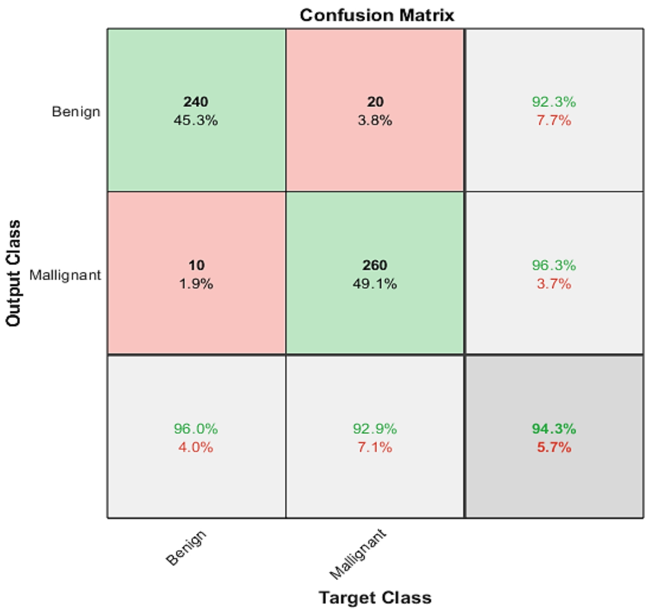

Fig. 9. Confusion matrix for spatial- ensemble method

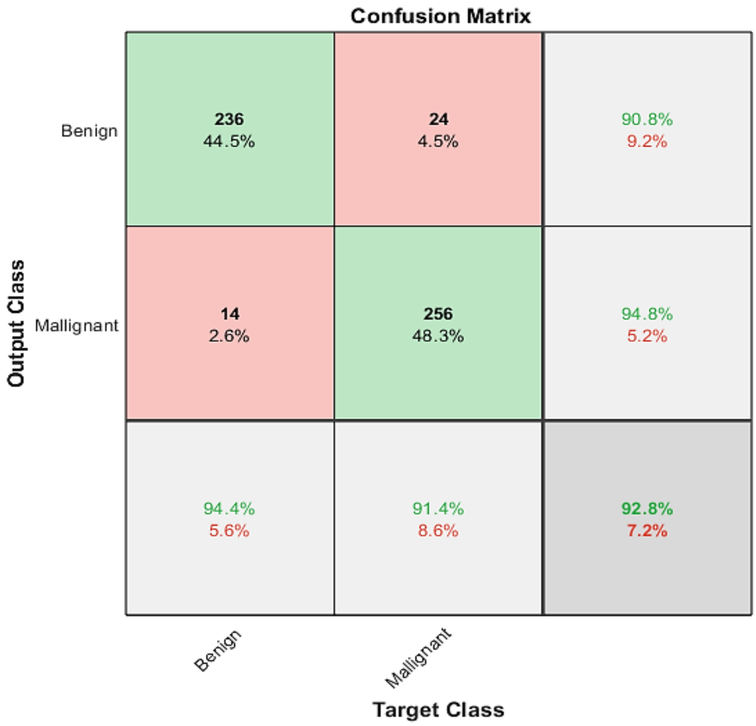

Fig. 10. Confusion matrix for frequency- ensemble method

Fig.9 and Fig.10 depict the performances of different models in terms of confusion matrix for spatial enhancement and frequency enhancement methods of preprocessing. The ensemble model shows better results compared to other models.

The receiver operating characteristics for two models, Alexnet and VGG with original data and augmented data are elucidated in fig. 11 and fig. 12. The RoC curves depict that the models perform better with augmented data samples rather than limited number of images in the original dataset. Fig 13. shows the RoC curves for different traditional classification techniques like neural networks and support vector machines, also for deep learning models VGG and ResNet. The proposed ensemble model achieves better performance compared to others.

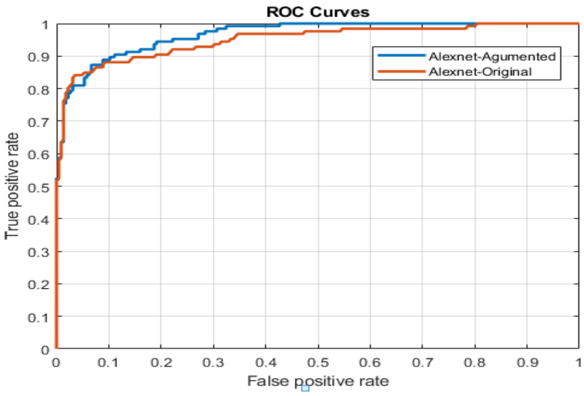

Fig. 11. Performance analysis of Alexnet

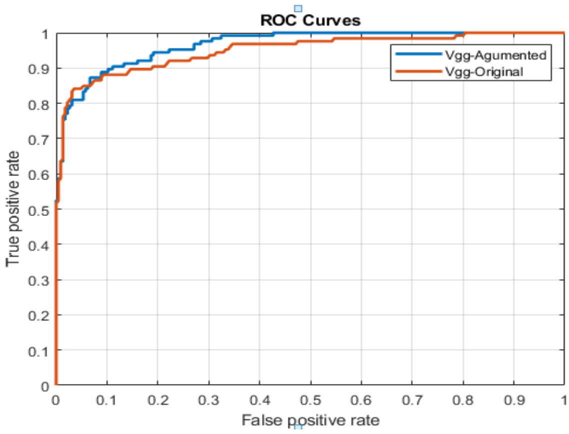

Fig. 12. Performance analysis of VGG-16

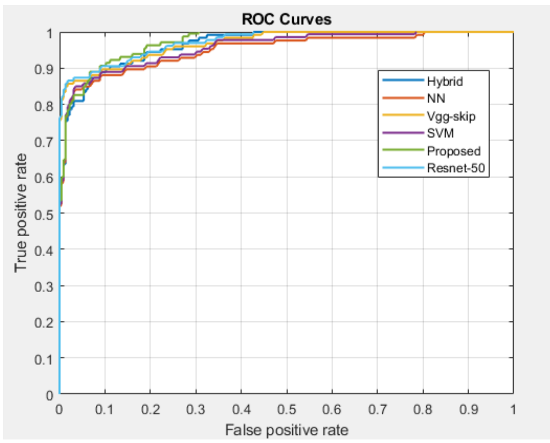

Fig. 13. Performance analysis of different models

4 Conclusion

The present work aims at developing an effective and economical means of oral cancer screening that can assist clinicians in making decisions about oral malignancies. The current work evaluates the efficacy of an ensemble model in the classification of oral malignancies using an augmented dataset of oral images. The proposed ensemble deep learning model helps to classify oral lesions as benign or malignant lesions from digital color images. The two deep network models considered are residual network - 50 and skip connected VGG-16. The ensemble model gives better results compared to the performance of individual models. So, this system can serve as an adjunct to oral screening techniques in health care systems and hospitals without the need of any special capturing devices.

5 References

[1] World Health Organization, "Oral cancer", https://www.who.int/cancer/prevention/ diagnosis -screening/oral-cancer/en.

[2] S. Petti and C. Scully, "Oral cancer knowledge and awareness: primary and secondary effects of an information leaflet", Oral Oncol. **43**(4), 408–415 (2007). https://doi.org/10.1016/j.oraloncology.2006.04.010

[3] American Society of Clinical Oncology, "Oral and Oropharyngeal Cancer: Statistics", https://www.cancer.net/cancer-types/oral-and-oropharyngeal-cancer/statistics.

[4] M. K. Mallath, D. G. Taylor, R. A. Badwe, G. K. Rath, V. Shanta, C. S. Pramesh, R. Digumarti, P. Sebastian, B. B. Borthakur, A. Kalwar, S. Kapoor, S. Kumar, J. L. Gill, M. A. Kuriakose, H. Malhotra, S. C. Sharma, S. Shukla, L. Viswanath, R. T. Chacko, J. L. Pautu, K. S. Reddy, K. S. Sharma, A. D. Purushotham, and R. Sullivan, "The growing burden of cancer in India: epidemiology and social context", The Lancet, Oncol. **15**(6), e205– e212 (2014). https://doi.org/10.1016/s1470-2045(14)70115-9

[5] Le Campion, C. M. B. Ribeiro, R. R. Luiz, F. F. da Silva Júnior, H. C. S. Barros, K. C. B. dos Santos, S. J. Ferreira, L. S. Gonçalves, and S. M.S. Ferreira, "Low Survival Rates of Oral and Oropharyngeal Squamous Cell Carcinoma", Hindawi, International Journal of Dentistry, Volume 2017, https://doi.org/10.1155/2017/5815493.

[6] National Institute of Cancer Prevention and Research, "Statistics," http://cancerindia.org.in/statistics/

[7] Bofan Song, Sumsum Sunny, Ross D. Uthoff, Sanjana Patrick, Amritha Suresh, Trupti Kolur, G. Keerthi, Afarin Anbarani, Petra Wilder-Smith, Moni Abraham Kuriakose, Praveen Birur, Jeffrey J. Rodriguez, Rongguang Liang, "Automatic classification of dual-modalilty, smartphone-based oral dysplasia and malignancy images using deep learning", Biomedical Optics Express, 2018. https://doi.org/10.1364/boe.9.005318

[8] Eric C. Yang, Imran S. Vohra, Hawraa Badaoui, Richard A. Schwarz, Katelin D. Cherry, Timothy Quang, Justin Jacob, Alex Lang, Nancy Bass, Jessica Rodriguez, Michelle D. Williams, Nadarajah Vigneswaran, "Development of an integrated multimodal optical imaging system with real-time image analysis for the evaluation of oral premalignant lesions", Journal of Biomedical Optics, 2019. https://doi.org/10.1117/1.jbo.24.2.025003

[9] M. M. R. Krishnan et al., "Automated oral cancer identification using histopathological images: a hybrid feature extraction paradigm," Micron, vol. 43, no. 2–3, pp. 352–364, 2012. https://doi.org/10.1016/j.micron.2011.09.016

[10] B. Thomas, V. Kumar, and S. Saini, "Texture analysis-based segmentation and classification of oral cancer lesions in color images using ANN," in 2013 IEEE International Conference on Signal Processing, Computing and Control (ISPCC), 2013, pp. 1–5. https://doi.org/10.1109/ispcc.2013.6663401

[11] H. J. Van Staveren, R. L. P. Van Veen, O. C. Speelman, M. J. H. Witjes, W. M. Star, and J. L. N. Roodenburg, "Classification of clinical autofluorescence spectra of oral leukoplakia using an artificial neural network: a pilot study," Oral Oncol. **36**(3), 286–293 (2000). https://doi.org/10.1016/s1368-8375(00)00004-x

[12] D. Roblyer, C. Kurachi, V. Stepanek, M. D. Williams, A. K. El-Naggar, J. J. Lee, A. M. Gillenwater, and R. Richards-Kortum, "Objective detection and delineation of oral neoplasia using autofluorescence imaging," Cancer Prev. Res. (Phila.) **2**(5), 423–431 (2009). https://doi.org/10.1158/1940-6207.capr-08-0229

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv Prepr. arXiv1409.1556, 2014.

[14] Dargan, S., Kumar, M., Ayyagari, M.R. et al. "A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning. Arch" Computat Methods Eng 27, 1071–1092 (2020). https://doi.org/10.1007/s11831-019-09344-w

[15] Mario Coccia, "Deep learning technology for improving cancer care in society: New directions in cancer imaging driven by artificial intelligence", Technology in Society, 2019. https://doi.org/10.1016/j.techsoc.2019.101198

[16] J. Folmsbee, X. Liu, M. Brandwein-Weber, and S. Doyle, "Active deep learning: Improved training efficiency of convolutional neural networks for tissue classification in oral cavity cancer," in 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018, pp. 770–773. https://doi.org/10.1109/isbi.2018.8363686

[17] Roshan Alex Welikala, Paolo Remagnino, Jian Han Lim, Chee Seng Chan, Senthilmani Rajendran, Thomas George Kallarakkal, Rosnah Binti Zain, Ruwan Duminda Jayasinghe, "Automated Detection and Classification of Oral Lesions Using Deep Learning for Early Detection of Oral Cancer" IEEE Access, vol. 8, pp. 132677-132693, 2020. https://doi.org/10.1109/access.2020.3010180

[18] Marc Aubreville, Christian Knipfer, Nicolai Oetter, Christian Jaremenko, Erik Rodner, "Automatic classification of cancerous tissue in laser endomicroscopy images of the oral cavity using deep learning," Scientific Reports, vol. 7, no. 1, p. 11979, 2017. https://doi.org/10.1038/s41598-017-12320-8

[19] R. K. Gupta, M. Kaur, J. Manhas, and others, "Tissue Level Based Deep Learning Framework for Early Detection of Dysplasia in Oral Squamous Epithelium," J. Multimed. Inf. Syst., vol. 6, no. 2, pp. 81–86, 2019. https://doi.org/10.33851/jmis.2019.6.2.81

[20] R. Anantharaman, V. Anantharaman, and Y. Lee, "Oro vision: Deep learning for classifying orofacial diseases," 2017 IEEE International Conference on Healthcare Informatics (ICHI), 2017, pp. 39–45. https://doi.org/10.1109/ichi.2017.69

[21] Navarun Das, Elima Hussain, Lipi B.Mahanta, "Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network", Neural Networks, Volume 128, August 2020, Pages 47-60. https://doi.org/10.1016/j.neunet.2020.05.003

[22] Jeyaraj, P.R., Samuel Nadar, E.R. Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm. J Cancer Res Clin Oncol 145, 829–837 (2019). https://doi.org/10.1007/s00432-018-02834-7

[23] Mohammed Shamim, Sadatullah Syed, Mohammad Shiblee, Mohammed Usman, Syed Ali, "Automated detection of oral pre-cancerous tongue lesions using deep learning for early

diagnosis of oral cavity cancer", Computer Journal, Oxford University Press, 2019, DOI: 10.13140/RG.2.2.28808.16643. https://doi.org/10.1093/comjnl/bxaa136

[24] Huang Lidong, Zhao Wei, Wang Jun, Sun ZebinCombination of contrast limited adaptive histogram equalisation and discrete wavelet transform for image enhancement, IET Image Processing, 2015. https://doi.org/10.1049/iet-ipr.2015.0150

[25] F. Wang et al., "Residual attention network for image classification," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3156–3164.

6 Authors

Nanditha B.R., is currently serving as Assistant Professor, Department of Information Science and Engineering at Malnad College of Engineering, Hassan, Karnataka, India. She has a teaching experience of 17 years. She is currently pursuing her PhD in image processing and computer vision. She has 04 research articles published in International Journals / conferences. She is a member of professional societies like CSI and ISTE.

Dr. Geetha Kiran A., is currently serving as Professor & Head, Department of Computer Science and Engineering at Malnad College of Engineering, Hassan, Karnataka, India. She has about 23 years of experience in teaching. She has her Ph.D. in the field of Image Processing from Mysore University. She is an active researcher and academician. Her areas of research include image processing, computer vision, machine learning and applications of Python programming.

Her research work entitled "Fruit Crop Yield Estimation Using Machine Vision Techniques" has received "TEQIP Competitive Research Grant". She has also been awarded with "Certificate of Appreciation" from Deputy Superintendent of Police, Hassan, Karnataka State Police, towards development of web-based application for monitoring quarantine persons of COVID-19 during April, 2020.

She has been presented with "**Most Influential Educational Leadership Award**" for her excellence & leadership in education during July, 2020 by Golden AIM awards. She has won best paper award for the paper titled "Instigating Research Culture at Institute level" in 2nd International Conference on Sustainable Growth through Universal Practices in Science Technology & Management, 2018.

She has over 25 research articles published in International Journals/conferences. She is a member of various professional societies like IEEE, CSI, ISTE and also Fellow member Institute of Engineers, India, Mysuru Chapter.

Article submitted 2020-10-14. Resubmitted 2020-11-22. Final acceptance 2020-11-23. Final version published as submitted by the authors.