EXPERIMENTAL ONCOLOGY

https://doi.org/10.15407/exp-oncology.2024.02.119

K. DHANYA, D. VENKATA VARA PRASAD *, LOKESWARI Y. VENKATARAMANA
Department of Computer Science Engineering,
Sri Sivasubramaniya Nadar College of Engineering, Rajiv Gandhi Salai, Kalavakkam, India
* Correspondence: Email: dvvprasad@ssn.edu.in

DETECTION OF ORAL SQUAMOUS CELL CARCINOMA USING PRE-TRAINED DEEP LEARNING MODELS

Background. Oral squamous cell carcinoma (OSCC), the 13th most common type of cancer, claimed 364,339 lives in 2020. Researchers have established a strong correlation between early detection and better prognosis for this type of cancer. Tissue biopsy, the most common diagnostic method used by doctors, is both expensive and time-consuming. The recent growth in using transfer learning methodologies to aid in medical diagnosis, along with the improved 5-year survival rate from early diagnosis serve as motivation for this study. The **aim** of the study was to evaluate an innovative approach using transfer learning of pre-trained classification models and convolutional neural networks (CNN) for the binary classification of OSCC from histopathological images. **Materials and Methods**. The dataset used for the experiments consisted of 5192 histopathological images in total. The following pre-trained deep learning models were used for feature extraction: ResNet-50, VGG16, and InceptionV3 along with a tuned CNN for classification. **Results**. The proposed methodologies were evaluated against the current state of the art. A high sensitivity and its importance in the medical field were highlighted. All three models were used in experiments with different hyperparameters and tested on a set of 126 histopathological images. The highest-performance developed model achieved an accuracy of 0.90, a sensitivity of 0.97, and an AUC of 0.94. The visualization of the results was done using ROC curves and confusion matrices. The study further interprets the results obtained and concludes with suggestions for future research. **Conclusion**. The study successfully demonstrated the potential of using transfer learning-based methodologies in the medical field. The interpretation of the results suggests their practical viability and offers directions for future research aimed at improving diagnostic precision and serving as a reliable tool to physicians in the early diagnosis of cancer.

Keywords: oral squamous cell carcinoma, transfer learning, pre-trained models, convolutional neural networks, histopathological images, classification.

Oral cancer, which subsumes malignancy of the lips, cheeks, tongue, hard and soft palate, and the base of the mouth extended to the oropharynx, ranks 13th worldwide as the most common type of cancer in 2020 [1]. In the same year, there were staggering 744,994 recorded cases and 364,339 deaths. Papua New Guinea tops the list for the highest overall rate with an ASR (age-standardized rate) value of 25.7 compared to the worldwide average of 8.0, while Bangladesh had the highest over-

Citation: Dhanya K, Venkata Vara Prasad D, Venkataramana Lokeswari Y. Detection of oral squamous cell carcinoma using pre-trained deep learning models. *Exp Oncol.* 2024; 46(2): 119-128. https://doi.org/10.15407/exp-oncology.2024.02.119

all mortality rate with 11.9 ASR as opposed to 3.9 globally [2]. Oral squamous cell carcinoma (OSCC) is the most commonly reported oral malignancy, representing up to 90% of all malignant neoplasms of the oral cavity [3]. According to the National Cancer Institute data, 5-year survival rates of OSCC patients drop significantly, by more than half, when the cancer metastasizes before detection [4]. Early detection thus becomes the key to reducing the mortality rate and fatalities. Unfortunately, this type of cancer does not present with clearly distinguishable symptoms [5] in the preliminary stages, which may lead to misdiagnosis or a delay in diagnosis [6].

Conventionally, medical experts and oncologists rely on tissue biopsies — surgical procedures to extract soft tissues of the oral cavity or lymph nodes — to detect oral cancer. The obtained sample, frozen and stained with hematoxylin and eosin, is used for pathological analysis [7]. Recently, there has been a notable emergence in the use of brush biopsy and cytology [8] to diagnose malignancies in the oral cavity. However, these invasive procedures are both time-consuming and expensive. To mitigate this, histopathological images can be used in developing deep learning models trained to identify cancerous cells with accuracies comparable to those of oncology experts [9—12]. Given the widespread success of these models in aiding medical professionals in performing diagnoses and predicting prognosis, this study aims to build a deep learning model using transfer learning of pre-trained models that classify oral histopathological images as OSCC or normal cells.

In this paper, the possibility of using the pre-trained models ResNet-50, VGG16, and InceptionV3 to extract features combined with classification using tuned neural networks to identify cell abnormalities was analyzed, which, to our knowledge, is the first of its kind to achieve an AUC of 0.96, sensitivity of 0.99, and an accuracy of 0.90.

Deep learning, a subset of machine learning, has revolutionized medical image analysis, particularly in the realm of disease detection and classification. In the context of OSCC, researchers have explored various deep learning techniques to improve diagnostic accuracy and patient outcomes.

Fu et al. [13] proposed a novel deep learning model based on cascaded neural networks for detecting OSCC from photographic images. Unlike traditional approaches, which often rely on manual inspection by oncology experts, their model offered a fully automated solution. By achieving an accuracy of 92.3% and sensitivity of 91%, the model demonstrated comparable performance to human experts, showcasing the potential of deep learning in medical imaging tasks.

Das et al. [14] contributed to the field by introducing a multi-class grading method for OSCC using convolutional neural networks (CNNs). In their study, they compared the performance of their CNN-based model with that of pre-trained models such as ResNet-50, VGG-16, and AlexNet. Their CNN model outperformed the pre-trained models, achieving an impressive accuracy of 97.5%. This work represents a significant advancement in the automated grading of OSCC, enabling more accurate and consistent diagnoses.

The utilization of pre-trained image classification models in healthcare has gained traction due to their ability to leverage knowledge learned from large datasets. Palaskar et al. [15] explored the application of transfer learning in oral cancer detection from microscopic images. They evaluated the performance of popular pre-trained models such as ResNet-50, InceptionV3, and MobileNet in combination with data augmentation and fine-tuning techniques. Their study highlighted the effectiveness of transfer learning in adapting pre-trained models to the medical imaging tasks, achieving an accuracy of 94.35% when combined with Over-Sampling.

Deif et al. [16] introduced an innovative approach by employing the Binary Particle Swarm Optimization (BPSO) for feature selection in OSCC detection. They combined features extracted by the InceptionV3 model with the XGBoost algorithm for classification and achieved an accuracy of 96.3%. Additionally, they explored the use of the Reinhard stain normalization to enhance performance, demonstrating the potential of feature selection techniques in improving diagnostic accuracy.

Moreover, the combination of feature extraction by transfer learning algorithms and subsequent classification emerged as a promising strategy. Maliki et al. [17] applied this approach to classify breast cancer histopathology images from the BreakHis dataset. By integrating features extracted by the DenseNet201 model with XGBoost for clas-

sification, they achieved a high accuracy of 93.6%, highlighting the effectiveness of fusion techniques in medical image analysis.

In summary, the recent advancements in deep learning techniques, pre-trained models, and innovative methodologies have significantly improved OSCC detection and classification.

These studies underscore the potential of deep learning in enhancing diagnostic accuracy, facilitating early detection, and improving patient outcomes in oral cancer management.

Materials and Methods

Materials. This study uses histopathological images of normal and malignant oral cells obtained from a public domain, the Kaggle dataset. The collection [18] contains microscopic images of cancerous and non-cancerous cells from different parts of the oral cavity, tongue, and lips. The dataset contains a total of 5192 labeled images, of which 2698 are cells affected by OSCC, while the remaining 2494 images are of normal oral squamous cells. The images are 224 x 224 pixels in size and belong to the RGB color model. The samples of the histopathological images are displayed in Figs. 1 and 2.

The Kaggle dataset was divided into a training set of 4141 images and a validation set of 1014 images. The remaining 126 images were used to test the developed model. For a comprehensive analysis of model performance, this study used AUC (area under curve), accuracy, sensitivity, precision, and F1-score as metrics to evaluate the proposed methodologies.

Proposed model. Fig. 3 diagrammatically represents the proposed model. The model first preprocesses the input — histopathological OSCC images. Then the features from these images are extracted using transfer learning of pre-trained image classification models. Using the extracted features, the tuned convolutional neural network is used to classify the features as normal or OSCC.

Data preprocessing. To make the input images compatible with the pre-trained models as well as improve accuracy, the following preprocessing techniques were applied. The images in the dataset were 224 x 224 pixels in size in the RGB color channel. They were resized to satisfy the specific requirements of the DL models. Further, the data were augmented by applying:

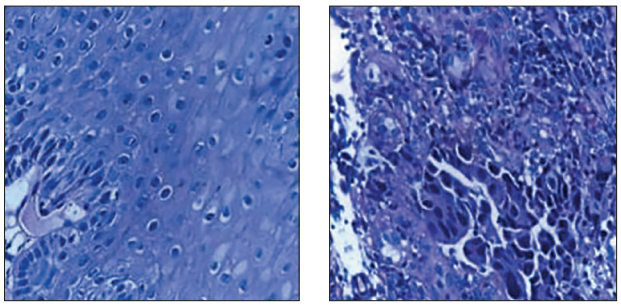

Fig. 1. Normal oral squamous cells

Fig. 2. OSCC-affected oral squamous cells

1. Random zooms: the input image is zoomed in or out by up to 30%.
2. Random rotations: the images are rotated randomly up to 0.3 radians.
3. Random flips: the input is subjected to random horizontal flips.

These techniques help introduce variations in the training set by applying random transformations. This, in turn, prevents overfitting and helps the model generalize better.

Feature extraction using pre-trained image classification models. This work concentrates on using transfer learning of the deep learning models ResNet-50, VGG16, and InceptionV3 to extract features from the augmented input images. The architecture of the 50-layer ResNet-50 model is depicted in Fig. 4.

ResNet-50: a deep convolutional network model, which is part of the Residual Network family and was introduced by Microsoft in 2015. Its popularity in image classification is mainly due to its ability to mitigate the "vanishing gradient" problem. The network is capable of learning intricate features in images, leading to high levels of accuracy in classification tasks. The residual blocks are characteristic of ResNet-50; they enable the network to learn residual mappings instead of entire transformations, thereby causing the gradient to flow smoothly through the blocks.

Features extracted: ResNet-50 learns incremental features through residual connections. The model's weights are adjusted during training and are used to encode the learned features when ResNet-50 is used as a feature-extractor. When the input image is passed through the convolutional layers, operations such as edge detection, object part recognition, and texture extraction are performed. The initial layers of ResNet-50 capture low-

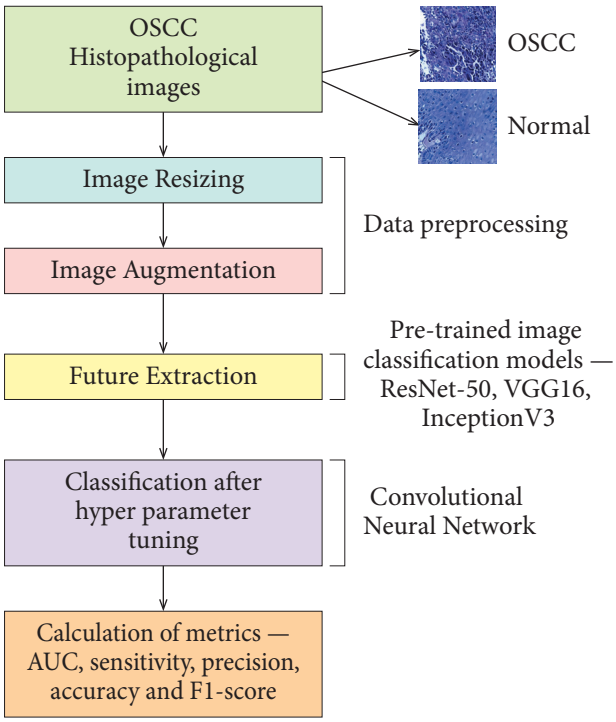

Fig. 3. Proposed architecture

level features, including colors, textures, and edges. Deeper down, the network learns more abstract, high-level features like object parts and shapes. Finally, the global average pooling layer condenses these features into a fixed-size vector, used for classification. This model contains five stages as depicted in Fig. 4. The total features extracted are the sum of the number of filters in each convolutional layer. This implementation of ResNet-50 extracts 26560 features from the given dataset.

VGG16: visual geometry group 16, in short VGG16, is a deep convolutional neural network used predominantly for image classification tasks. It consists of 13 convolutional layers and 3 fully connected layers, with a max-pool layer with 2 x 2 filters after every 2 convolutional layers. ReLu, the activation function used, introduces non-linearity. Despite its commendable performance, the memory-intensive and computationally-expensive nature of this model makes it less preferable. The architecture is depicted in Fig. 5.

Features extracted: the VGG16 model is also trained on the ImageNet dataset, however it has a simpler architecture. This model extracts features by passing the input image through its layers and obtaining the output from the intermediate layers. The layers of VGG16 are arranged hierarchically, and like the ResNet-50 model, it captures low-level features such as edges initially, and later layers capture high-level features. The convolutional filters generate activation maps representing the highlighting areas in the image corresponding to specific features. The pooling layers in between help perform spatial dimension reduction, and the global pooling layer averages the value in each feature map to generate a global summary of the features. VGG16 is particularly suited for limited data, as it learns a wide range of features from the ImageNet dataset. This model extracted 4224 features from the images in the dataset.

Inception V3. It is an extension of the original inception (GoogLeNet) model and has achieved significant results in computer vision tasks. With a ore refined inception module, InceptionV3 consists of 48 layers. The core of InceptionV3 is a combination of different convolutions, making it effective in capturing local and global patterns. This model uses factorization to reduce the number of parameters and also employs global average pooling instead of fully connected layers at the end of the network. This helps it mitigate the issue of overfitting. The auxiliary classifiers help InceptionV3 alleviate the vanishing gradient problem.

Features extracted**:** The layers of InceptionV3 are trained to detect features of different scales and complexities. Unlike ResNet-50 and VGG16, InceptionV3 employs multiple parallel convolutional layers with different kernel sizes. This enables the network to capture fine-grained details as well as a broader context simultaneously. The bottleneck layers help reduce the number of input channels. Similar to the other pre-trained models, InceptionV3 also uses pooling layers to downsample the spatial dimensions of the feature map. A total of 17216 features were extracted by InceptionV3. Fig. 6 denotes the number of features extracted by each pre-trained model.

Classification using convolutional neural network. This study used CNNs to classify the features extracted by the pre-trained models, namely, ResNet-50, VGG16, and InceptionV3. The network is tuned to be better suited for this task by performing 3 rounds of randomized hyper-parameter search, each with 5 trials, and the overall best hyper-parameters are used to train the network.

The output from the pre-trained model is reduced to a fixed-size vector using a global pooling layer. The two dense layers, whose units are deter-

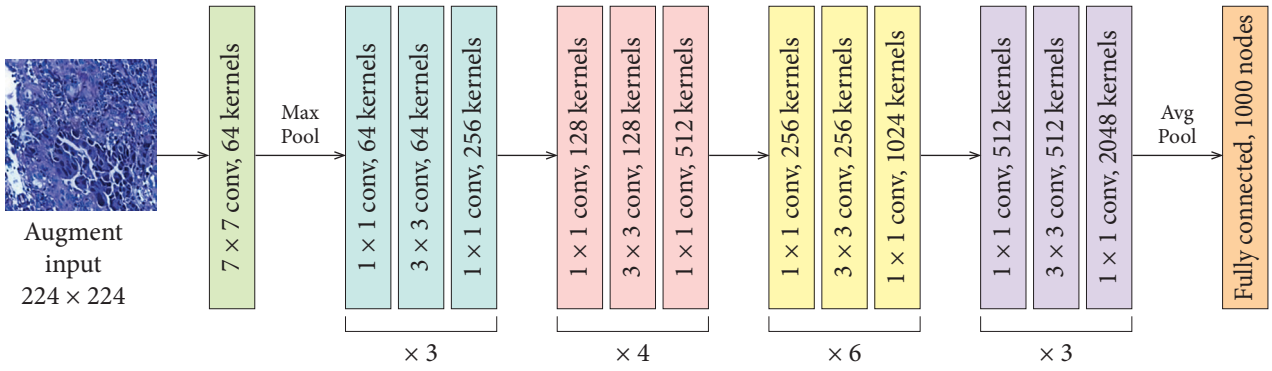

Fig. 4. Architecture of ResNet-50

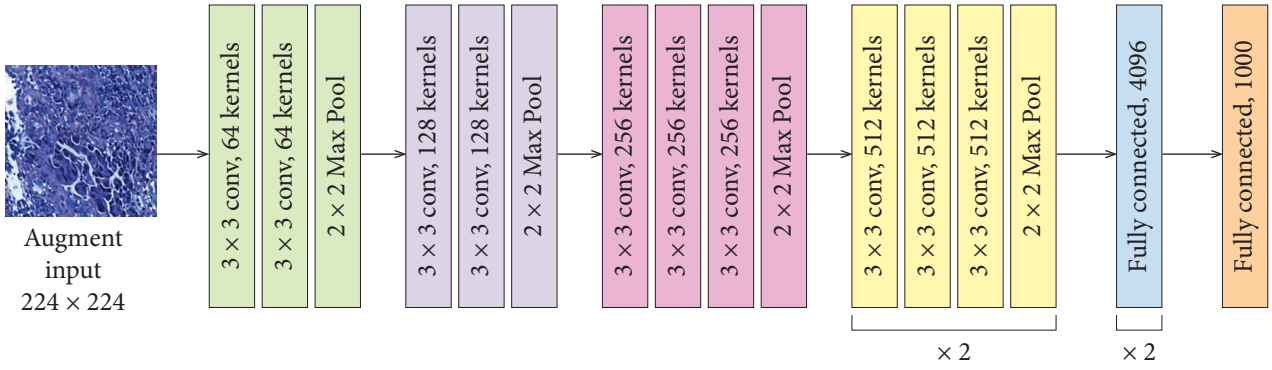

Fig. 5. Architecture of VGG16

mined via the hyper-parameter search, are executed next. ReLu (Rectified Linear Unit) is used as the activation function. Finally, the output —the probability between 0 and 1 that the image belongs to the positive class —is presented by the sigmoid activation function in the last dense layer.

Hyper-parameter tuning. The hyper-parameter search is performed using a RandomSearch tuner from the Keras Tuner library. This tuner performs random sampling of hyper-parameters from the defined search-space and trains the model with the chosen hyper-parameters. The model is then evaluated on the validation set to identify the combination that yields the best performance. The validation accuracy is chosen as an optimization parameter. Two parameters are tuned: the number of units in the dense layer after global pooling and the learning rate for the Adam optimizer. The former determines the model's complexity while the latter controls the step size during gradient descent. The tuner is initialized with a function to build the model, and the objective is set to "validation accuracy". The maximum number of hyper-parameter combinations to try is set to be 5, and the number of training epochs per combination is set

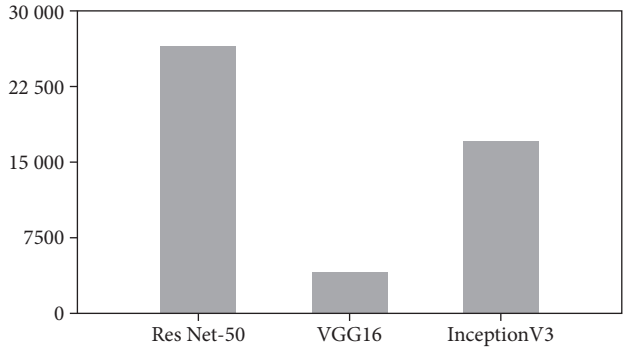

Fig. 6. Features extracted by pre-trained models

to be 10. Each model is trained exactly once for the random search. The search space for the number of units is defined as a range [32,128], and the learning rates for the Adam optimizer are chosen from the values [1e-3, 1e-4, 1e-5]. This search was performed twice, and the best parameters identified were: [learning rate = 0.001, units = 320] and [learning rate = 1e-5, units = 512]. Each model with a different pre-trained model to extract features and optimal hyper-parameters was evaluated on the test set.

Structure of the neural network. The network has an input layer that accepts images of 128 x 128 pixels and RGB color channels. This is followed by the pre-trained model. The top layers of the pre-trained

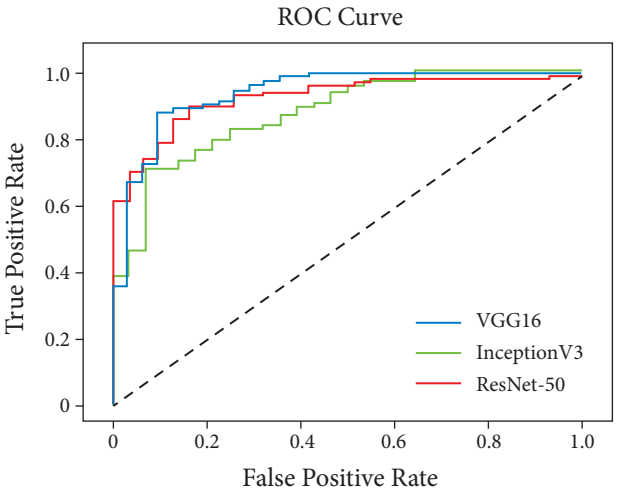

Fig. 7. ROC curves for transfer learning of the deep learning models ResNet-50, VGG16, and InceptionV3

models are set to False to perform classification by the optimized neural network. The neural network then consists of a global average pooling layer — the pre-trained model extracts features from the images and represents them as feature maps. This global average pooling layer helps summarize these maps by computing the average value for each channel across all spatial dimensions. The output is a fixed-size feature vector for each input image. This helps prevent overfitting and enables the model to generalize better. Since the object position is irrelevant after pooling, it makes the model translationally invariant. This layer is followed by Dense layers — one with ReLu activation and one with binary sigmoid activation. The former Dense layer maps the reduced feature vector to higher dimensions, enabling the network to learn intricate patterns in the data. The final Dense layer acts as the output layer and maps the features to a single output unit, a number between 0 and 1 indicating the probability that the image belongs to the positive class. This architecture enables the CNN to accurately classify the features extracted by the pre-trained models as normal or OSCC.

Results and Discussion

The deep learning models (ResNet-50, VGG16, and InceptionV3) were pre-trained on the ImageNet datasets. The convolutional base layers are used to extract features from the histopathological images. These base layers are frozen during training. The classification layers of the models are set to false, and a tuned deep network is used for classification. This study uses the Adam optimizer to compile the deep model and binary cross-entropy loss. The training and validation datasets are in an 80:20 split ratio. This work used a batch size of 32 and 10 epochs. Next, a hyper-parameter tuning using RandomSearchCV is performed to find the ideal hyper-parameters for the convolutional neural network (units for the dense layers and learning rate). The search was performed 3 times, with 5 trials in each run. A new network is created with the top 2 optimal hyper-parameters and evaluated on the test set. A learning rate of 0.001 and 320 units and 1e-05 and 512 units were found to be optimal in the search conducted. All experiments were performed on Google Colab [19] with GPU support. The code is written in Python version 3.10.6, Keras version 2.12.0, and TensorFlow version 2.12.0. The classification is done using CNN. This paper has experimented in depth with the use of ResNet-50, VGG16, and InceptionV3 to extract features and the tuned CNN to classify the extracted features.

The pre-trained deep models (ResNet-50, VGG16, and InceptionV3) were used as feature extractors, and the tuned convolutional neural network was used as a classifier. The models were trained on a set of 4141 histopathological images (2445 normal, 1796 OSCC). Following training, the models were validated on a set of 1014 images. The models were evaluated on the unseen images from a test dataset of 126 images (31 normal, 95 OSCC). The performances of the classifiers were visualized using the ROC (receiver operator characteristic) curve depicted in Fig. 7.

The evaluation metrics of the different models, including AUC, sensitivity, precision, F1-score, and accuracy, on the test set consisting of 126 images were compared. They are tabulated in Table 1. This study considers the sensitivity of the proposed models as an important metric in analyzing performance. The high sensitivity of deep learning models plays a pivotal role in clinical applications such as cancer detection. This ensures that even subtle indications or early manifestations of cancerous cells are not overlooked, enabling timely intervention and treatment. This capability is particularly crucial in cancer detection, where early diagnosis significantly influences treatment outcomes and patient prognosis. Furthermore, the high sensitivity of these models not only improves patient out-

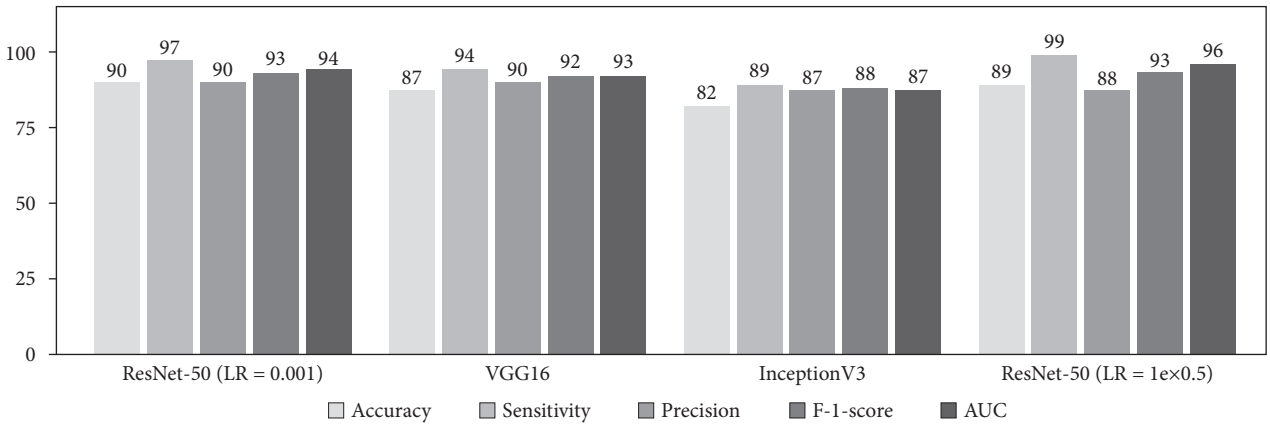

Fig. 8. Comparison of test metrics of proposed models

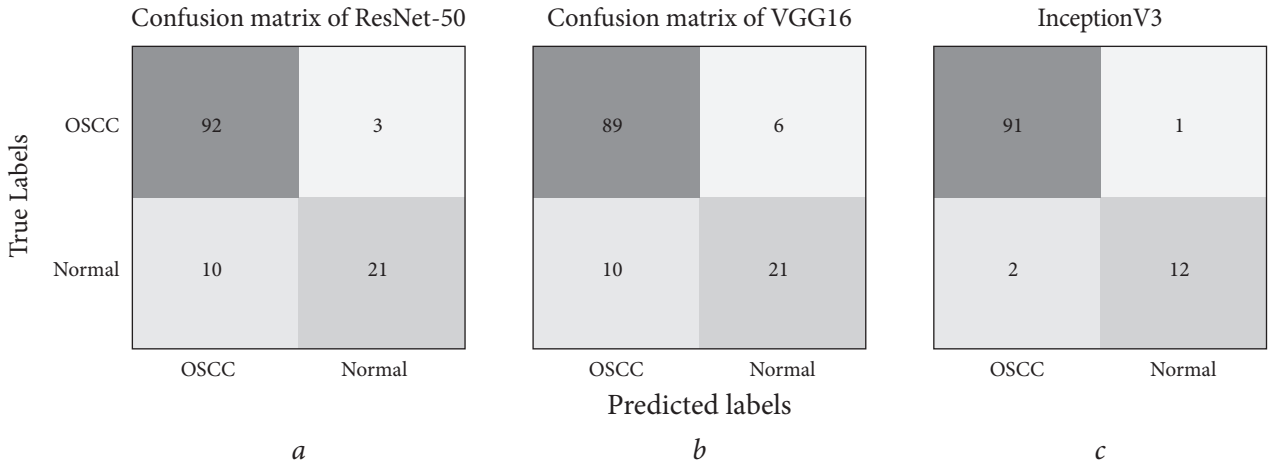

Fig. 9. Confusion matrix of ResNet-50 (*a*), VGG16 (*b*), Inception V3 (*c*)

Table 1. **Test metrics of proposed models**

Model description	Accuracy	Sensitivity	Precision	F1-score	AUC
ResNet-50 + CNN (LR = 0.001)	0.90	0.97	0.90	0.93	0.94
ResNet-50 + CNN (LR = 1e-05)	0.89	0.99	0.88	0.93	0.96
VGG16 + CNN (LR = 0.001)	0.76	0.87	0.83	0.85	0.72
VGG16 + CNN (LR = 1e-05)	0.87	0.94	0.90	0.92	0.93
InceptionV3 + CNN (LR = 1e-05)	0.82	0.89	0.87	0.88	0.87

Table 2. **Test metrics of the proposed model *vs* previous studies**

Study	Method used	Accuracy	Sensitivity
Welikala et al. [22]	ResNet101 and R-CNN	0.81	0.89
Rahman et al. [23]	Transfer Learning of AlexNet	0.90	0.93
Auberville et al. [24]	Deep learning	0.88	0.86
Fu et al. [15]	DCNN	0.92	0.91
Proposed Method	Transfer Learning of ResNet-50	0.90	0.97

comes but also supports healthcare professionals in delivering timely and effective interventions. This can potentially advance the standard of care for OSCC and other malignancies.

Analyzing the performance, this work has observed that using augmented data with ResNet-50 as a feature extractor and a tuned CNN as a classifier exhibited the best overall performance. InceptionV3 underperformed as a feature extractor with an AUC of 0.87 compared to 0.96 of the ResNet-50 model (LR = 1e-05). Moreover, the significant difference between the training and test accuracies of the InceptionV3 model could be a sign of overfitting. Although the VGG16-based model performed decently across various metrics such as sensitivity and precision, it fell short in accuracy (0.76) of prediction. All metrics discussed were obtained from testing the model on unseen data of 126 images (31 normal, 95 OSCC). The best-performance version of each proposed methodology is visualized in Fig. 8. Table 2 displays the test accuracies achieved by the benchmark models compared to our proposed method. As mentioned, the overall best performance was displayed by the ResNet-50 — along with the tuned CNN, it achieved remarkable sensitivities of 97% and 99% against 89% achieved by the transfer learning model proposed by Welikala et al. [20]. The proposed ResNet-50 methodology has a test accuracy of 0.90, the same as that achieved by the Rahman et al. [21] methodology in classifying biopsy images using transfer learning of AlexNet. Auberville et al. [22] used a deep learning model to classify CLE images, achieving an accuracy of 88.3% and a sensitivity of 86.6%, whereas the proposed methodology achieves 90% accuracy and 97% sensitivity.

This study has also analyzed the confusion matrices of the presented models. A comparison of true positives, true negatives, false positives, and false negatives for all the developed models was performed. The ResNet-50-based model and the VGG16-based model both have a true positive value of 21 and a false negative value of 10. However, the false positive value of ResNet-50 (3) is lower compared to that of the VGG16 model (6). InceptionV3 performs comparatively poorly on true positives with a value of only 16. Fig. 9 show the confusion matrices of the models.

Interpreting the reason for misclassifications by the model is crucial for improving the quality of prediction. The heterogeneity of cancerous cells and tissues can lead to variations in appearance, texture, and staining patterns, making it challenging for the model to accurately capture all malignant features. The presence of subtle or atypical cancerous cells, which may not conform to typical morphological characteristics, can also result in misclassification or oversight by the model. Moreover, the histopathological images may contain artifacts, such as staining inconsistencies, tissue folds, or artifacts introduced during slide preparation, which can confound the feature extraction by the pretrained models and the classification by the CNN. Furthermore, the very few datasets publicly available for research purposes may not fully represent the diverse range of cancerous presentations and can limit the model's ability to generalize to unseen cases. The limitations in the resolution or quality of histopathological images can affect the model's ability to discern minute details crucial for accurate cancer detection.

The false positives may arise due to the complexity of histopathological images, where benign squamous cells or tissue structures might resemble cancerous ones, leading the model to misclassify them. The false negatives can occur when cancerous cells exhibit variations in appearance, size, or staining patterns, which the model fails to accurately capture during feature extraction and classification. While still significant, the implications of the false positives in the medical field are less catastrophic compared to the impact of the false negatives. The false negatives can lead to a delayed diagnosis, allowing the cancer to progress to more advanced stages before treatment initiation. This delay can result in poorer treatment outcomes, decreased survival rates, and increased morbidity for patients. The missed diagnoses can lead to missed opportunities for early intervention and treatment, impacting the effectiveness of the therapeutic interventions and potentially necessitating more aggressive treatments with higher associated risks and costs. Therefore, addressing and minimizing the false negatives in OSCC detection are critical for improving patient outcomes, reducing healthcare costs, and ensuring high-quality care delivery.

In conclusion, we have performed comprehensive experiments on using transfer learning of pretrained deep models to aid in OSCC classification and attained strong results. The best performance

was exhibited by the model using ResNet-50 as a feature extractor with hyper-parameters as [LR = 0.001 and units = 320]. A remarkable sensitivity of 97%, an AUC of 94%, and a test accuracy of 90% were achieved. The hope is that this study solidifies the mainstream usage of deep models in medical diagnosis.

Despite the promising results, the study acknowledged several challenges, including the heterogeneity of cancerous cells, limited publicly available datasets, and image artifacts affecting model generalization and performance. Addressing these challenges through further research and tuning of deep learning architectures could lead to improved diagnostic tools for OSCC and other malignancies.

The comprehensive experiments conducted in this study underscored the potential of deep learning models in medical diagnosis, particularly in cancer detection. The achieved results, particularly with the ResNet-50 model, provide a strong foundation for future research and clinical applications. The paper calls for continued efforts in dataset curation, model refinement, and validation to enhance the reliability and utility of deep learning-based diagnostic tools in the medical domain.

REFERENCES

1. World Health Organization. Oral health. Available from: https://www.who.int/news-room/fact-sheets/detail/oral-health. Accessed August 7, 2023.
2. World Cancer Research Fund. Mouth and oral cancer statistics. Available from: https://www.wcrf.org/cancer-trends/mouth-and-oral-cancer-statistics/. Accessed August 7, 2023.
3. Johnson NW, Jayasekara P, Amarasinghe AA. Squamous cell carcinoma and precursor lesions of the oral cavity: epidemiology and aetiology. *Periodontol 2000*. 2011;57(1):19-37. https://doi.org/10.1111/j.1600-0757.2011.00401.x
4. National Cancer Institute. Five-year survival rate. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/five-year-survival-rate. Accessed August 7, 2023.
5. Silverman S Jr. Early diagnosis of oral cancer. *Cancer.* 1988;62(8 Suppl):1796-1799. https://doi.org/10.1002/1097-0142(19881015)62:1+<1796::aid-cncr2820621319>3.0.co;2-e
6. Gigliotti J, Madathil S, Makhoul N. Delays in oral cavity cancer. *Int J Oral Maxillofac Surg*. 2019;48(9):1131-1137. https://doi.org/10.1016/j.ijom.2019.02.015
7. Yang G, Wei L, Thong BKS, et al. A systematic review of oral biopsies, sample types, and detection techniques applied in relation to oral cancer detection. *BioTech*. 2022;11(1):5. https://doi.org/10.3390/biotech11010005
8. Gupta S, Shah JS, Parikh S, et al. Clinical correlative study on early detection of oral cancer and precancerous lesions by modified oral brush biopsy and cytology followed by histopathology. *J Cancer Res Ther*. 2014;10(2):232-238. https://doi.org/10.4103/0973-1482.136539
9. Dabeer S, Khan MM, Islam S. Cancer diagnosis in histopathological image: CNN based approach. *Inform Med Unlocked*. 2019;16:100231. https://doi.org/10.1016/j.imu.2019.100231
10. Iqbal MJ, Javed Z, Sadia H, et al. Clinical applications of artificial intelligence and machine learning in cancer diagnosis: looking into the future. *Cancer Cell Int*. 2021;21:270. https://doi.org/10.1186/s12935-021-01981-1
11. Bakare YB, Kumarasamy M. Histopathological image analysis for oral cancer classification by support vector machine. *IJASIS*. 2021;7(2):1-10. https://doi.org/10.29284/ijasis.7.2.2021.1-10
12. Alabi RO, Almangush A, Elmusrati M, Mäkitie AA. Deep machine learning for oral cancer: from precise diagnosis to precision medicine. *Front Oral Health*. 2022;2:794248. https://doi.org/10.3389/froh.2021.794248
13. Fu Q, Chen Y, Li Z, et al. A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: A retrospective study. *EClinicalMedicine*. 2020;27:100516. https://doi.org/10.1016/j.eclinm.2020.100516
14. Das N, Hussain E, Mahanta LB. Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network. *Neural Networks*. 2020;128:47-60. https://doi.org/10.1016/j.neunet.2020.04.011.
15. Palaskar R, Vyas R, Khedekar V, et al. Transfer learning for oral cancer detection using microscopic images. arXiv preprint arXiv:2011.11610,2020. Available from: https://arxiv.org/pdf/2011.11610
16. Deif MA, Attar H, Amer A, et al. Diagnosis of oral squamous cell carcinoma using deep neural networks and binary particle swarm optimization on histopathological images: an AIoMT approach. *Comput Intel Neurosc*. 2022;2022:6364102. https://doi.org/10.1155/2022/6364102
17. Maleki A, Raahemi M, Nasiri H. Breast cancer diagnosis from histopathology images using deep neural network and XGBoost. *Biomed Signal Process Control*. 2023;86(Part A):105152. https://doi.org/10.1016/j.bspc.2023.105152
18. Ashenafi Fasil Kebede. Oral Cancer Histopathology Images Dataset. Available from: https://www.kaggle.com/datasets/ashenafifasilkebede/dataset. Accessed August 7, 2023.

19. Google Colaboratory FAQ. https://research.google.com/colaboratory/faq.html. Accessed August 7, 2023.
20. Welikala RA, Dissanayake DMNK, Housden J. Automated detection and classification of oral lesions using deep learning for early detection of oral cancer. *IEEE Access*. 2020;8:132677-132693. https://doi.org/10.1109/ACCESS.2020.3010180
21. Rahman AU, Alqahtani A, Aldhafferi N, et al. Histopathologic oral cancer prediction using oral squamous cell carcinoma biopsy empowered with transfer learning. *Sensors*. 2022;22(10):3833. https://doi.org/10.3390/s22103833
22. Aubreville M, Knipfer C, Oetter N, et al. Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning. *Sci Rep*. 2017;7(1):11979. https://doi.org/10.1038/s41598-017-12320-8

Submitted: January 23, 2024

D. Крішнан, Д.В. Прасад, Ю.В. Локешварі
Відділ комп'ютерних наук, Шрі Сівасубраманія
Надарський інженерний коледж, Раджив Ганді Салай, Калаваккам, Індія

ВИЯВЛЕННЯ ПЛОСКОКЛІТИННОЇ КАРЦИНОМИ ПОРОЖНИНИ РОТА ЗА ДОПОМОГОЮ ПОПЕРЕДНЬО НАВЧЕНИХ МОДЕЛЕЙ ГЛИБОКОГО НАВЧАННЯ

Стан питання. Плоскоклітинний рак ротової порожнини (ПКРРП) є 13-м за поширеністю типом злоякісних новоутворень, що став причиною загибелі 364339 осіб у 2020 р. Встановлено існування кореляційного зв'язку між ранньою діагностикою та кращим клінічним прогнозом для цього типу раку. Біопсія тканин, найпоширеніший метод діагностики, який використовують у клініці, є доволі дорогою і трудомісткою. Нещодавнє зростання використання методів трансферного навчання для допомоги в медичній діагностиці, а також покращення 5-річної виживаності при ранній діагностиці слугували мотивацією для цього дослідження. **Метою** дослідження було оцінити інноваційний підхід з використанням трансферного навчання попередньо навчених моделей класифікації та згорткових нейронних мереж (ЗНМ) для бінарної класифікації ПКРРП на гістопатологічних зображеннях. **Матеріали та методи**. Набір даних, використаний для експериментів, складався з 5192 гістопатологічних зображень. Для встановлення ознак ПКРРП на патогістологічних зображеннях використовували такі попередньо навчені глибинні моделі: а) ResNet-50, б) VGG16, в) InceptionV3 разом із налаштованими ЗНМ для класифікації. **Результати**. Встановлено високу чутливість та перспективність запропонованого підходу виявлення ПКРРП у медичній галузі. Всі три моделі були використані в експериментах з різними гіперпараметрами і протестовані на наборі з 126 гістопатологічних зображень. Найефективніша розроблена модель досягла точності 0,90, чутливості 0,97 та AUC 0,94. Візуалізація результатів була виконана за допомогою ROC-кривих та матриць помилок. У роботі також інтерпретовано отримані результати та сформульовано пропозиції щодо подальших досліджень. **Висновки**. Проведене дослідження успішно продемонструвало потенціал використання методології трансферного навчання в медичній галузі. Інтерпретація результатів свідчить про його практичну життєздатність і пропонує напрямки подальших досліджень, спрямованих на підвищення точності діагностики.

Ключові слова: плоскоклітинний рак порожнини рота, трансферне навчання, попередньо навчені моделі, згорткові нейронні мережі, гістопатологічні зображення, класифікація.