**Article Information**

**Article Title**

Exploring the Efficacy of Various CNN Architectures in Diagnosing Oral Cancer from Squamous Cell Carcinoma

**Authors**

Prerna Kulkarni

Nidhi Sarwe

Abhishek Pingale

Yash Sarolkar

Dr. Rutuja Kadam (\*)

Dr. Gitanjali Shinde

Dr. Rupali Gangarde

**Affiliations**

Prerna Kulkarni – Department of CSE (AIML), Vishwakarma Institute of Information Technology, Kondhwa (Budruk) Pune – 411048, Maharashtra (India)

Nidhi Sarwe – Department of CSE (AIML), Vishwakarma Institute of Information Technology, Kondhwa (Budruk) Pune – 411048, Maharashtra (India)

Abhishek Pingale - Department of CSE (AIML), Vishwakarma Institute of Information Technology, Kondhwa (Budruk) Pune – 411048, Maharashtra (India)

Yash Sarolkar - Department of CSE (AIML), Vishwakarma Institute of Information Technology, Kondhwa (Budruk) Pune – 411048, Maharashtra (India)

Dr. Rutuja Kadam – Assistant Professor, Vishwakarma Institute of Information Technology, Kondhwa (Budruk) Pune – 411048, Maharashtra (India), Department of CSE (AIML)

Dr. Gitanjali Shinde – Head of Department, CSE (AIML), Vishwakarma Institute of Information Technology, Kondhwa (Budruk) Pune – 411048, Maharashtra (India)

Dr. Rupali Gangarde – Assistant Professor, Department of CSE, Symbiosis Institute of Technology, Pune

**Corresponding author’s email address and twitter handle**

prerna.22211460@viit.ac.in

nidhi.22211105@viit.ac.in

abhishek.22210280@viit.ac.in

yash.22211527@viit.ac.in

rutujapat@gmail.com

[gitanjali.shinde@viit.ac.in](mailto:gitanjali.shinde@viit.ac.in)

[rupali.gangarde@sitpune.edu.in](mailto:rupali.gangarde@sitpune.edu.in)

**Keywords**

Oral cancer Histopathologic OSCC; Squamous Cell Carcinoma; Convolutional Neural Networks; Deep learning; Histopathological Images

**Abstract**

Oral cancer can be caused by the mutation of the cells in the lips or in the mouth.[1]  Oral cavity squamous cell carcinoma (OCSCC) diagnosis remains challenging, often detected at advanced stages. Computer-aided diagnosis approaches are resorted to with a view to solving these problems. In this regard, a deep learning-based approach has been presented which makes use of deep learning models such as VGG16, ResNet50, LeNet-5, MobileNetV2, and Inception V3 by availing the dataset of NEOR and OCSCC samples for feature extraction. Virtual slide images were divided into tiles and labelled as normal or squamous cell cancer. Through the analysis of performance parameters like accuracy, F1-score, AUC, precision, and recall, the prerequisites for obtaining strong CNN performances were determined. CNN approaches are investigated for OCSCC and oral dysplasia classification. The best classification accuracy of 95.41% was obtained when using MobileNetV2.

**Graphical** **Abstract**



Fig 1: Structured Workflow for Binary Classification of Normal and OSCC Images

**Specifications Table**

|  |  |
| --- | --- |
| Subject area | Medicine and Dentistry |
| More specific subject area | OralSquamousCellCarcinoma |
| Name of your method | Convolutional Neural Networks |
| Name and reference of original method | * VGG16: [2] * ResNet50: [3] * LeNet5: [4] * InceptionV3: [5] * MobileNetV2: [6] |
| Resource availability | [Code (Github Link)](https://github.com/Prernak12/Oral-Cancer-Comparative-Study), [Dataset](https://data.mendeley.com/datasets/ftmp4cvtmb/1) |

**Background**

Oral cancer is a prevalent illness, frequently identified only in its advanced stages, which poses a significant challenge to effective treatment. More than 90% of instances of oral cancer are of the most prevalent kind, oral squamous cell carcinoma (OSCC), which arises from the mucosal lining of the mouth. Consequently, OSCC has been identified as a subtype of head and neck squamous cell carcinoma, the seventh most common cancer worldwide. It appears that head and neck cancer is a more complicated illness that needs highly specialized advice from oral and medical oncologists, radiologists, pathologists, and surgeons.[7]

The development of oral cancer frequently results from exposure to carcinogens, mostly alcohol and tobacco, on the upper aerodigestive tract mucosal coverings. A sequence of events within the mucosa may result in premalignant and malignant lesions as a result of this exposure. It is interesting, nonetheless, that some people with oral cancer are not known to have used tobacco products or alcohol, nor do they display any other known risk factors.

The oral areas are home to neoplasms with a variety of cellular origins, such as salivary gland tumors, nasopharyngeal carcinoma, lymphomas, mucosal melanoma, and sarcomas. There are a number of rare histological variations of squamous cell carcinoma that affect prognosis and treatment options. Spindle cell or sarcomatoid squamous malignancies are less common in the oral cavity but more frequently seen on the lip and larynx.

Indeed, early detection is intrinsic to successful treatment and improvement in the outcomes of OSCC patients. Many forms of cancer are treatable or curable if treated early, and so many clinical protocols have been developed for the detection and staging of cancers [8]

So far, one considers a biopsy as the gold standard test to confirm the diagnosis of cancer.[9] Over the years, various research works have concentrated on combining AI with the improvement of medical diagnostics.[10] With the high consumption rate in diagnostic imaging guiding them, the research community has been well-positioned to investigate multiple applications in medical image analysis with substantive building blocks for further improvements in diagnostic accuracy and efficiency. Deep learning techniques, especially CNNs, have become very effective and powerful tools in analyzing medical images that diagnose oral cancer in squamous cells as compared to Machine Learning Algorithms.[11], [12], [13] Based on the large dataset for histopathological images, CNNs can learn complicated patterns and characteristics connected with cancerous cells with high accuracy and efficiency in identifying malignancies present in oral tissue samples.[14] This study focuses on the development and evaluation of algorithms for squamous cell images using CNNs with the aim to achieve much more accurate and improved results. The present study intends to propose a hybrid deep learning approach based on the power of CNNs for feature extraction so that the quality of oral cancer detection becomes more accurate and consistent.

**Method details**

Deep learning is a subset of machine learning that normally involves learning depictions at distinct levels of hierarchy to allow construction of complex concepts.[15] When referring to techniques that enable computers to act like people, the term "artificial intelligence" is used generally. Regular pattern recognition involves identifying salient characteristics by careful inspection, which is followed by feeding the features into a simple neural network to perform sorting. Deep learning, on the other hand, exploits the system's ability to recognize significant features on its own to solve problems. Similar to a network of neurons, deep learning gathers input from the user and processes it via multiple layers to provide a response. One such deep learning algorithm is Convolutional Neural Networks. DCNNs are useful for learning the patterns from the images and are able to classify using these patterns, thereby eliminating the need for manual feature extraction. [17], [18], [19] The present research is a comparative study of 5 Convolutional Neural Networks namely VGG-16, ResNet-50, LeNet-5, MobileNetV2 and Inception V3.[16]

The classification model predicts oral cancer in squamous cells using deep learning techniques such as VGG-16, ResNet-50, LeNet-5, MobileNetV2and Inception V3. The model is trained and tested on a labelled dataset that classifies the images into:

1. OSCC Histopathologic Image

2. Normal Cavity Histopathologic Image

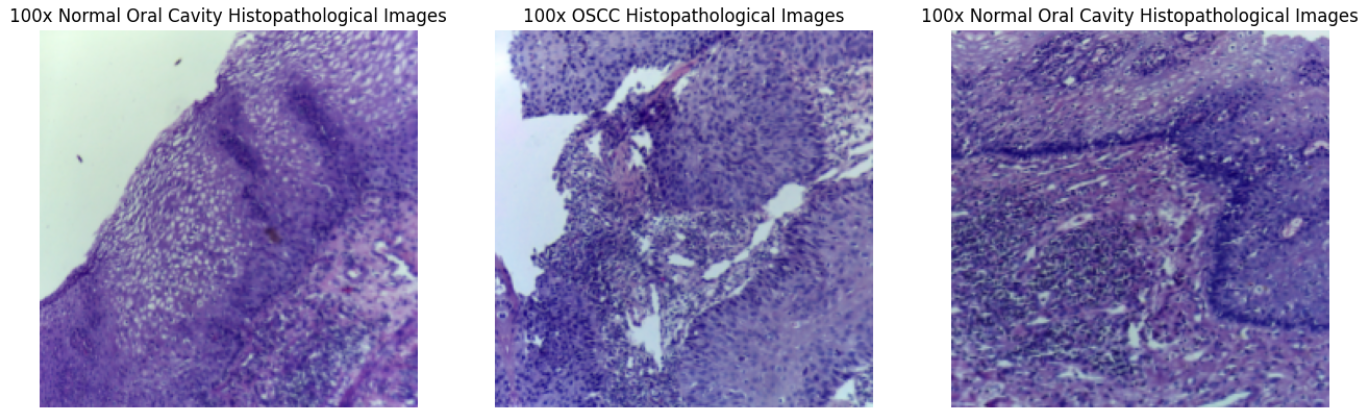
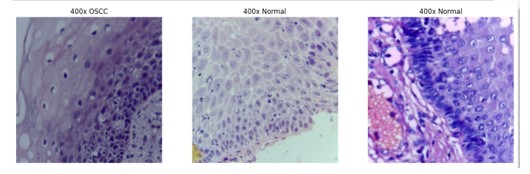


Fig 2: Different Classes of Histopathological Images from the Original Dataset.

After classifying the images into OSCC Histopathologic Images and normal Cavity Histopathologic Images, Image processing is done to extract relevant information from this visual data. The Fastai library based on PyTorch would provide a really great set of tools for creating and implementing fine-tuned image processing pipelines. These might include a number of processes that could involve augmentation, normalization, scaling, and modification. Image augmentation helps avoid overfitting by integrating techniques such as rotation, flipping, scaling, and color changes in the training dataset.

A total of 6931 images were generated by augmenting the original dataset of 1224 images which was sourced from [20] with the help of the ImageDataGenerator class of keras. The images were rescaled by about 1/255, rotated randomly by about 0–40 degrees, translated randomly by 20% of height or width in either direction, randomly zoomed by 20%, randomly flipped, and the newly created pixels were filled by the nearest pixel value.

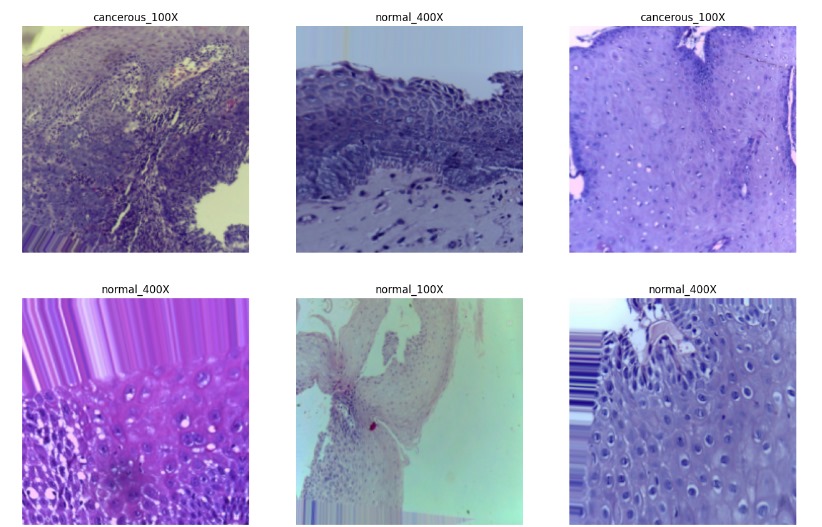


Fig 3: Images Generated after Augmentation of Original Dataset

|  |  |  |  |
| --- | --- | --- | --- |
| Resolution | Class | Original number of Images | Number of Images after Augmentation |
| 100X Magnification | Normal | 89 | 1780 |
| OSCC | 439 | 1615 |
| 400X Magnification | Normal | 201 | 1815 |
| OSCC | 495 | 1721 |

Table 1. Image Details from dataset in terms of Type and Quantity.

Expanding the dataset helps deep learning models perform better and prevent overfitting by ensuring good model generalization.[21] Augmenting the dataset helps to expand size of the dataset without the necessity of gathering fresh data, as deep learning necessitates a significant amount of data.

Consequently, this augmentation resulted in the division of the dataset into 80% for training and 20% for testing. In this work, four CNN architectures—VGG-16, ResNet-50, LeNet-5, MobileNet V2, and Inception V3—for the diagnosis of oral cancer will be implemented and evaluated using the Keras framework. The top layers of pre-trained models for VGG-16, ResNet-50, and Inception V3 were improved by using custom-made fully connected layers designed for the specific classification purpose. These models were further refined after initial training by freezing the base layers.

**VGG16 Deep Learning Algorithm**

'VGG' refers for the University of Oxford's Visual Geometry Group, and the number '16' in VGG16 denotes the network's 16 weighted layers. The deep convolutional neural network called VGG16 is employed in image classification. To improve prediction accuracy, the 16 layers of artificial neurons in the network analyse each image independently. VGG16 incorporates maxpool layers with a 2x2 filter and a stride of 2, as well as convolution layers with a 3x3 filter and a stride of 1, that are in place of a large number of hyperparameters. In the subsequent stage, pre-trained weights are inserted into the VGG16 model to fine-tune it for data training.

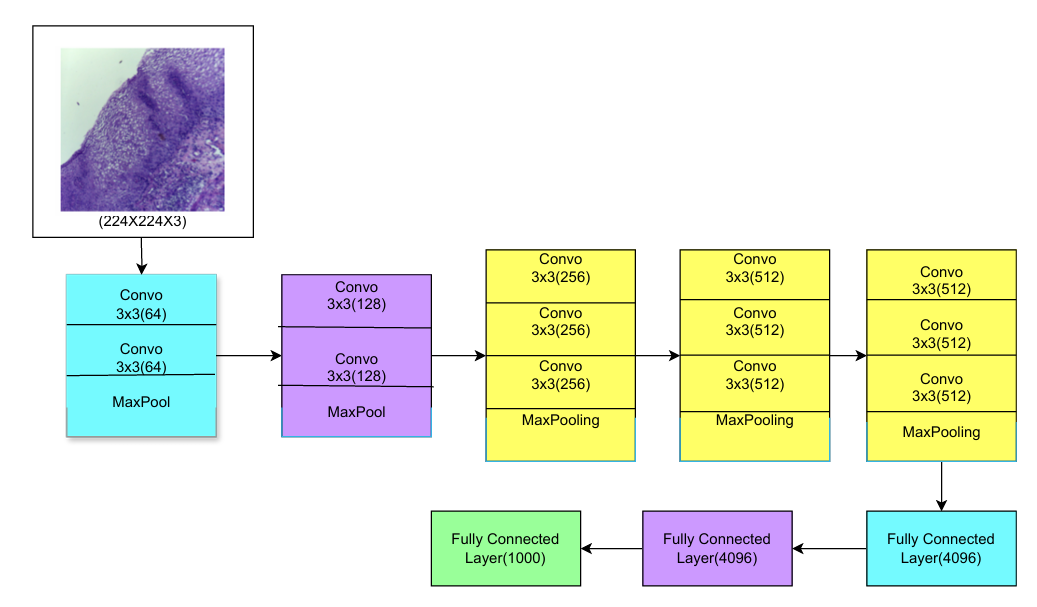


Fig. 4: Architecture of VGG-16

The model's performance is assessed for each class, and errors are identified using performance metrics like as accuracy, recall, AUC, f1-score, and confusion matrix.[22]

**ResNet-50 Deep Learning Algorithm**

ResNet-50, or Residual Networks, is a deep neural network used for many computer vision applications For eg; object detection and image segmentation. The topology, with residual connections, is very deep. This provides considerable mastering and subsequently turns the network into a very strong tool for classes of image classification tasks in medical imaging. The findings derived from the present study have emphasized the capacity of the model to support medical diagnosis and research, as pointed out hereinbelow by its capability to distinguish between images with and without cancerous lesions.[21]

The ResNet-50 model was imported with pre-training on weights on ImageNet and adapted for a binary classification assignment. Only the top layers were removed so that the network behaves like a feature extractor and allows only the last few layers to be trainable. Further, on top of the base model, a GlobalAveragePooling2D layer was added, followed by Dense layer with sigmoid activation for binary classification. The network was trained using the Adam optimizer, using a learning rate of 0.01 and a loss function of binary cross-entropy. The training lasted ten epochs, and the picture data creators provided both the training and validation datasets. Callbacks such as ReduceLROnPlateau and EarlyStopping were used to avoid overfitting and to dynamically adjust the learning rate in a more efficient way during training.

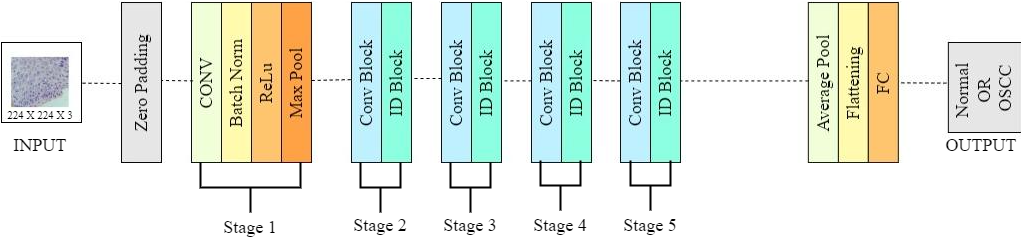


Fig 5: Architecture of ResNet-50

**LeNet-5 Deep Learning Algorithm**

LeNet-5, developed in 1998 by Yann LeCun and colleagues, was one of the earliest convolutional neural network (CNN) architectures. Layers one to three are fully connected layers at the end, followed by two convolutional layers and subsampling (pooling) layers. The first layers in the architecture to apply various filters to extract feature maps are the convolutional layers. After that, these maps are down-sampled by average pooling to reduce dimensionality while keeping important details. The final fully linked layers divide the features into various groupings.

The MNIST dataset is a popular benchmark in the fields of computer vision and machine learning. It comprises of a vast collection of handwritten digit images that are often used to train and test convolutional neural networks (CNNs).

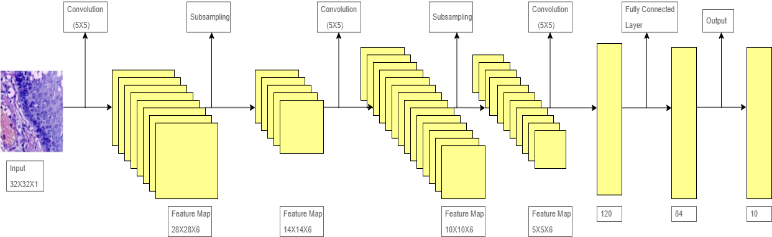


Fig 6: Architecture of LeNet-5

**MobileNetV2 Deep Learning Algorithm**

MobileNetV2 is an efficient, light-weight, yet powerful convolutional neural network model that has been designed to be deployed for mobile and embedded vision applications. Google's MobileNetV2 extends the success of its predecessor, MobileNetV1, by incorporating several state-of-the-art improvements contributing to its better performance and effectiveness.

Optimal performance in embedded and mobile applications is achieved with the MobileNetV2 architecture because of a well-chosen balance between efficiency and accuracy. First of all, the first convolutional layer down-samples an RGB image of the fixed size of 224 x 224 pixels and adds 32 channels after processing by the input layer. The basic building block of this system's architecture is the Inverted Residual Block. It includes depthwise convolution, projection layer-1x1 linear convolution, and expansion layer-1x1 convolution with ReLU6. These blocks are further facilitated by the shortcut connections that enhance the flow of a gradient especially when the dimensions of input and output are same. After these blocks, a 1x1 Conv2D layer extends channels to 1280, followed by a completely linked layer for classification and global average pooling.

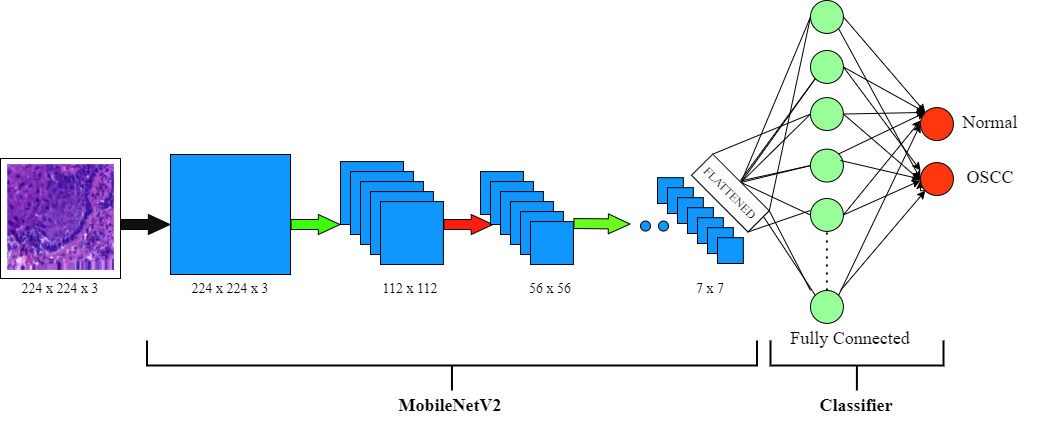


Fig 7: Architecture of MobileNetV2

**Inception V3 Deep Learning Algorithm**

To extract features from images, the InceptionV3 architecture employs a number of convolutional, pooling, and inception modules. By employing filters of various sizes, inception modules—blocks of layers—enable the network to learn a range of features at various scales and resolutions.

Necessary libraries are imported for image processing (PIL) and for defining data configuration and transformations. Data configuration for the model is resolved and creates transformation functions based on the configuration. File path for an image is defined and loads the image using PIL, converting it to RGB format. Transformations to the image are applied to prepare it for input into the model.

For each image, it assumes the variable probabilities holds the probabilities for each image. It then prints the top 5 predicted categories along with their probabilities.

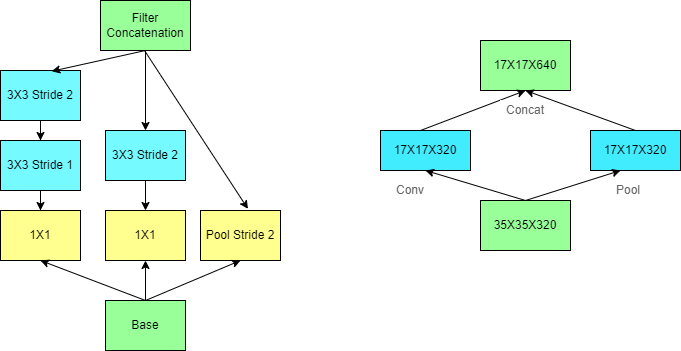


Fig 8: Architecture of Inception V3

**Method validation**

The pre-trained models VGG16 and ResNet50 beat LeNet5 and InceptionV3, suggesting that transfer learning is effective for OSCC image categorization. VGG16 and ResNet50 demonstrated superior precision, AUC, recall, and F1-scores, indicating their suitability for practical applications in medical image processing.

**Performance of VGG-16:**

A well-known convolutional neural network architecture, the VGG16 model, classified the dataset with an accuracy of 76.74%. This excellent performance shows that, in spite of any potential complications in the dataset, the model can learn and generalize features from the photos. The dense layers and narrow receptive fields of the model's design provide accurate feature extraction, which enhances its robustness.

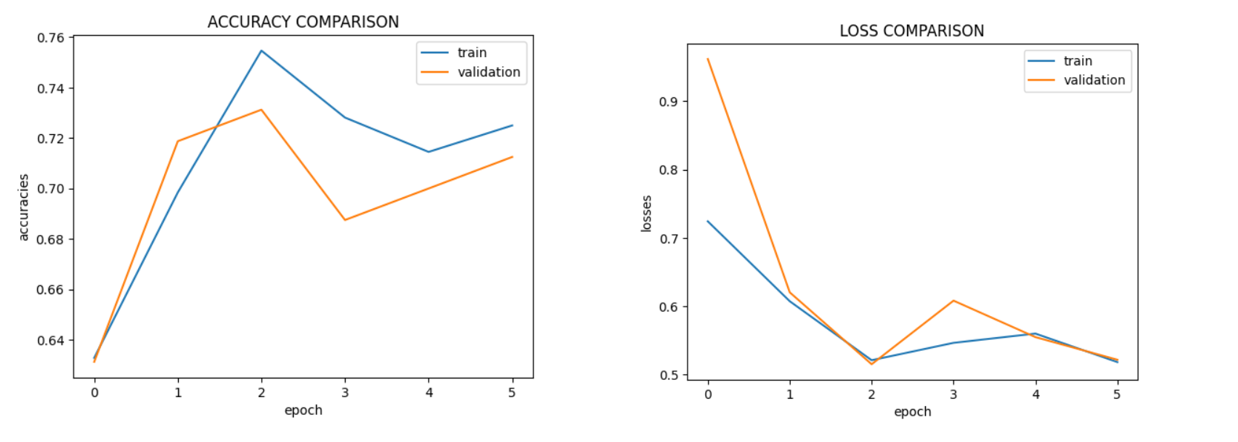


Fig 9: VGG-16 Model Performance Metrics

**Confusion Matrix:**

The generated confusion matrix indicates normal cells as 0 and Oral Squamous Carcinoma Cells as 1, thereby doing binary classification and generating a 2X2 matrix for VGG16 model.

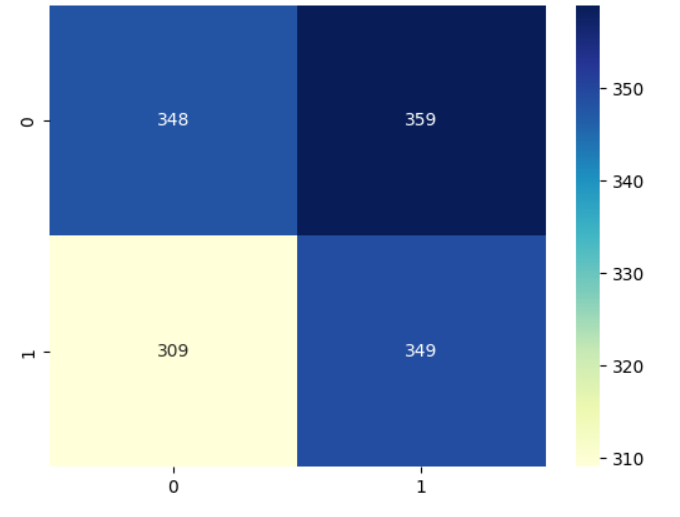


Fig 10: VGG16 Model Confusion Matrix

**Performance of ResNet-50:**

Classifying photos into normal and malignant categories was accomplished with an accuracy of 71% by the ResNet-50 model, which is renowned for its deep residual learning architecture. This outcome demonstrates the model's reliable operation and potent feature extraction skills. ResNet-50's capacity to differentiate between complicated image classes in the dataset is demonstrated by its success.

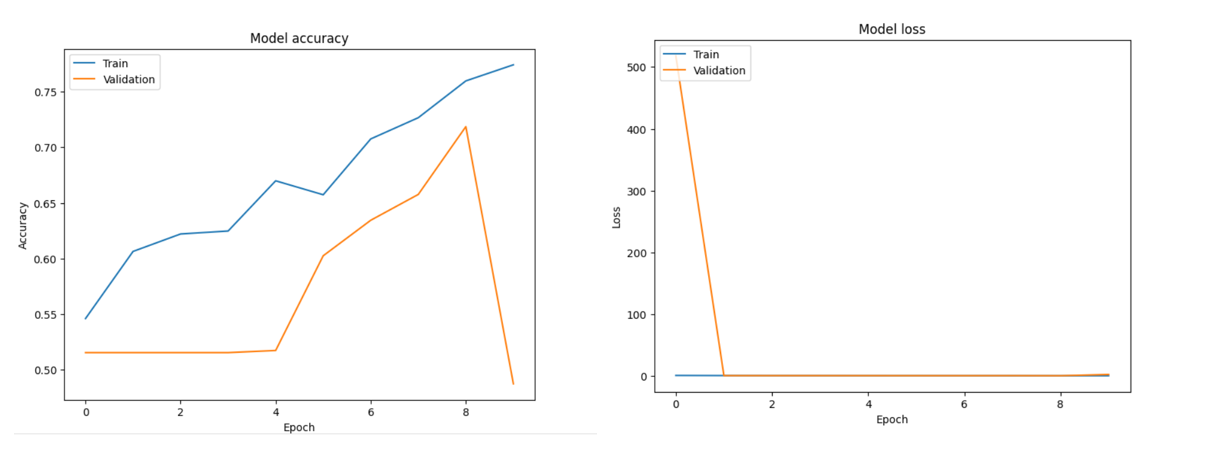
**

Fig 11: ResNet-50 Model Performance Metrics

**Confusion Matrix:**

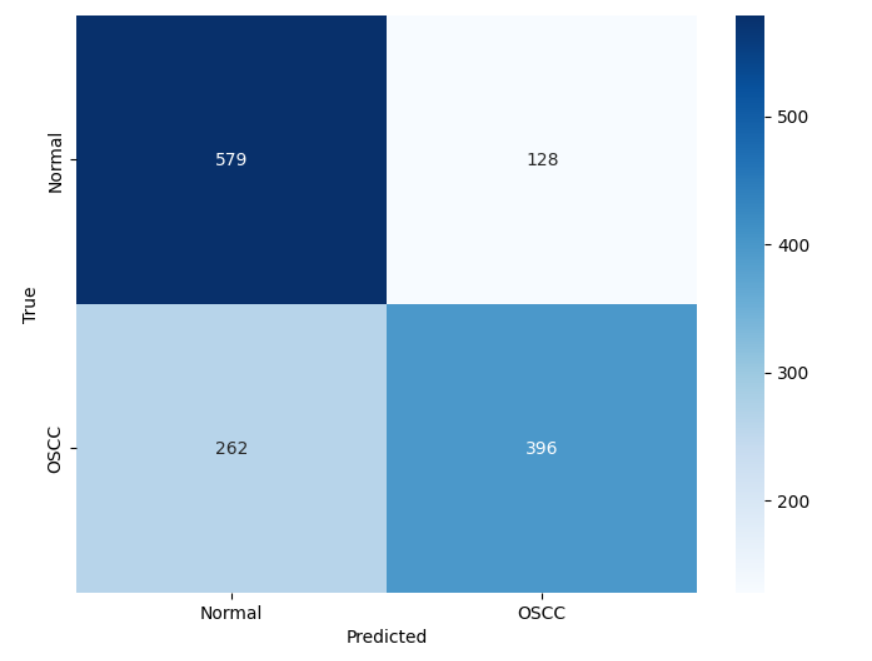


Fig 12: ResNet-50 Model Confusion Matrix

**Performance of LeNet-5:**

The comparison investigation revealed that the LeNet-5 model is capable of achieving 76.54% accuracy in picture classification tasks. However, when compared to other models such as VGG16 and ResNet-50, LeNet-5 demonstrated lower accuracy and feature extraction efficiency.

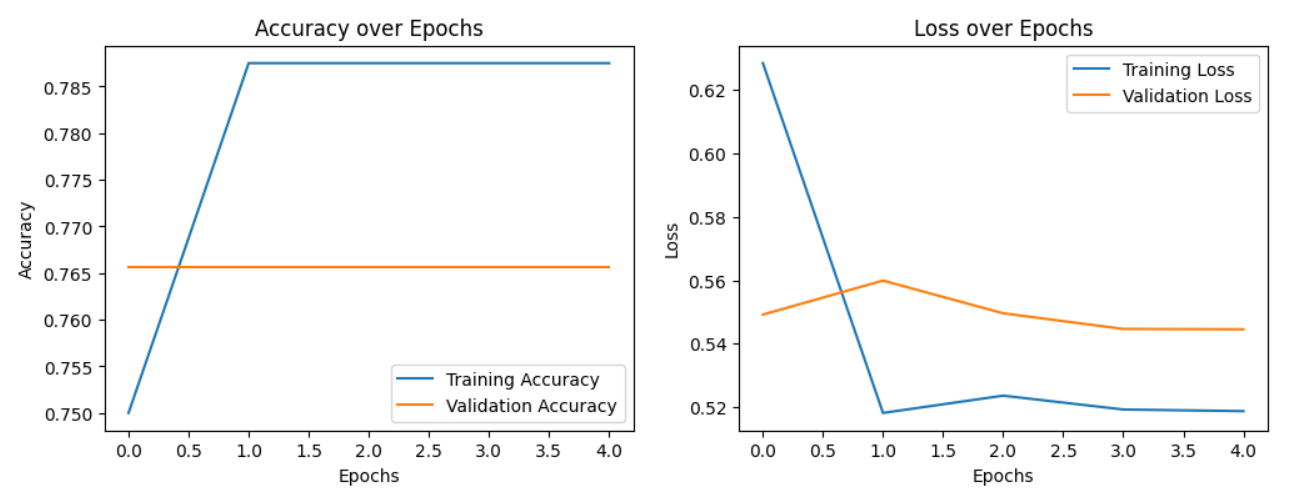


Fig 13: LeNet-5 Model Performance Metrics

**Performance of MobileNetV2:**

MobileNetV2 achieves an accuracy of 95.41%, exhibiting robust performance metrics. In order to acknowledge the variations in both training and validation loss over the epochs, emphasize that the model has shown good generalization capabilities, as revealed by the small difference between training accuracy and validation accuracy. The alignment of training accuracy and validation accuracy indicates that the model is not overfitting, despite these occasional oscillations. This great degree of accuracy is a result of the architecture's skill at striking a balance between computing efficiency and model complexity. This performance is enhanced by the inclusion of inverted residual blocks, which minimize the number of parameters while optimizing both depth wise separable convolutions and linear bottlenecks.



Fig 14: MobileNetV2 Model Performance Metrics

**Confusion Matrix:**

The generated confusion matrix indicates normal cells as 0 and Oral Squamous Carcinoma Cells as 1, thereby doing binary classification and generating a 2X2 matrix for MobileNetV2 model.

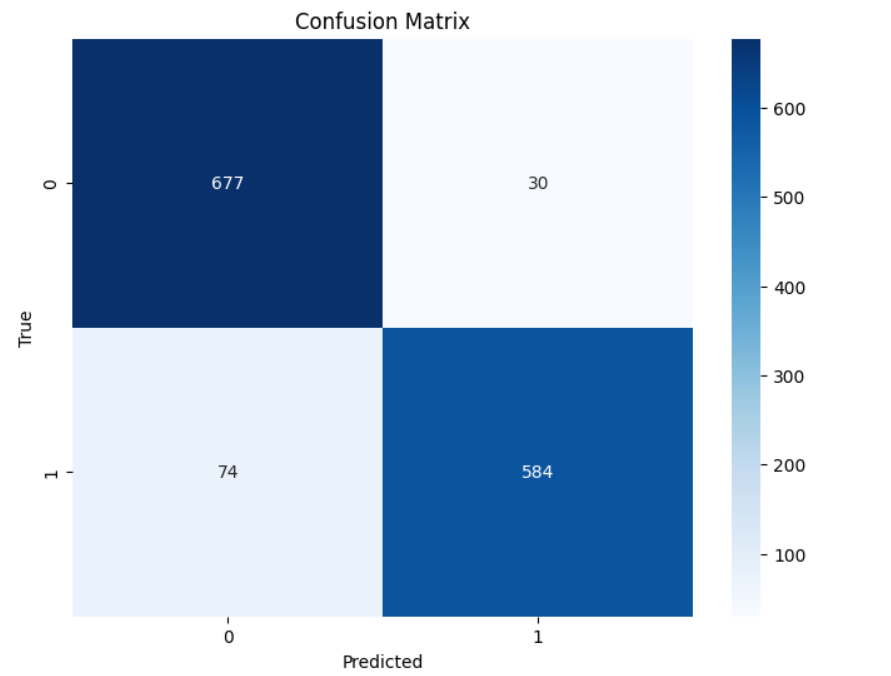


Fig 15: MobileNetV2 Model Confusion Matrix

**Performance of Inception V3:**

The Inception V3 model, which is well-known for its utilization of computational power and multi-scale feature extraction, successfully classified images into normal and malignant groups with an accuracy of 51.86%. This further reflects the performance of the model in handling diverse visual aspects with its intricate design.

The accuracy of Inception V3 model, when applied to the supplemented dataset, is found to be very less. The collection of photographs includes 100X and 400X magnification with normal tissue and oral squamous cell cancer. After supplementation, there were 1615 photos at 100X magnification for OSCC, 1815 normal photos at 400X magnification, and 1780 normal shots at 100X magnification. This augmentation might not have been enough to remove the bias in training because of the huge class imbalance between the normal and OSCC images in the original dataset. There could also have been anomalies or recurring patterns introduced by these augmentation methods that wouldn't generalize well for new data. The low performance could have been exacerbated by the Inception V3 model, too.

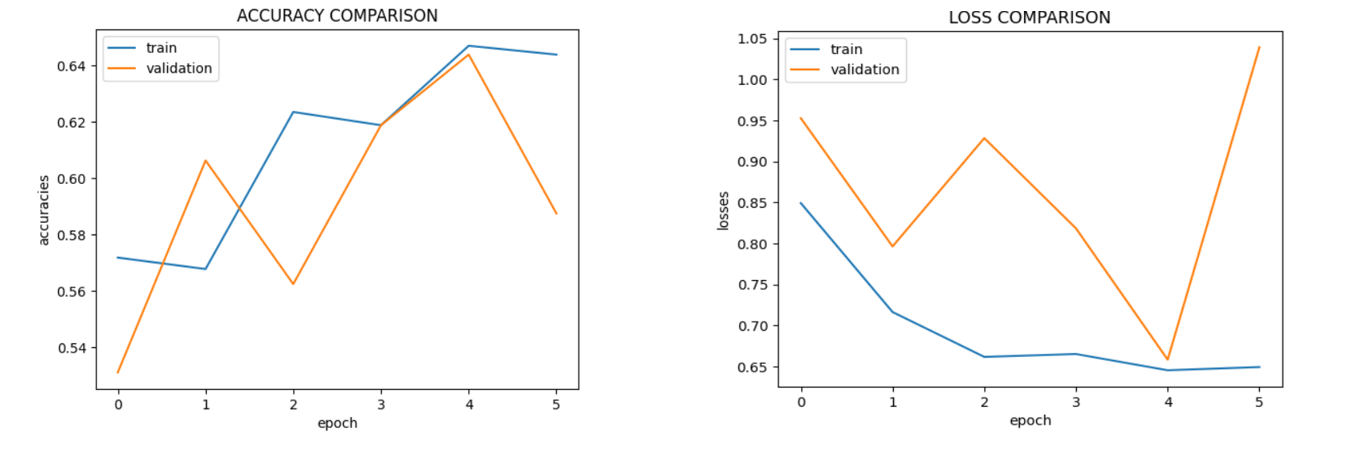


Fig 16: Inception V3 Model Performance Metrics

**Confusion Matrix:**

The resulting confusion matrix performs a binary classification, producing a 2X2 matrix for the Inception V3 model, in which the normal cells are labeled as 0 and oral squamous carcinoma cells are labeled as 1.

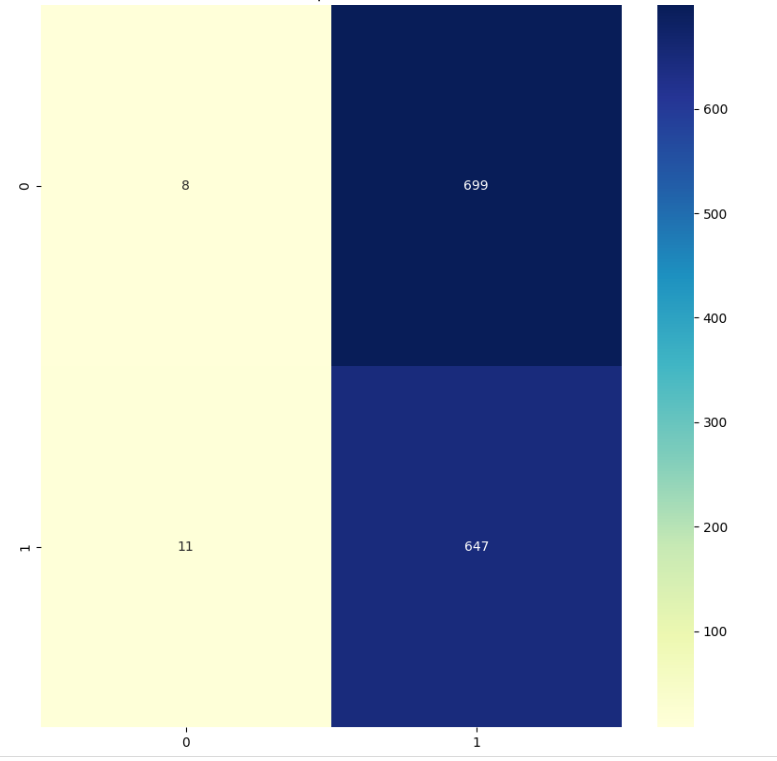


Fig 17: Inception V3 Model Confusion Matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Measures | Vgg-16 | ResNet50 | LeNet-5 | MobileNetV2 | Inception V3 |
| Accuracy (%) | 76.54 | 71.00 | 76.56 | 95.41 | 51.86 |
| Precision (%) | 74.90 | 76.12 | 87.00 | 95.03 | 51.29 |
| Recall (%) | 78.99 | 60.43 | 77.00 | 89.06 | 99.21 |
| F1-Score (%) | 76.89 | 67.28 | 81.47 | 92.08 | 67.62 |

Table 2: Model Comparison Study using Evaluation Metrics

**Limitations**

Considering the model complexities, this work had a few limitations with respect to its major dataset of 6391 images. Basically, increasing the volume of data enhances the model complexity, but the challenge is maintaining optimal performance. The more complicated these models get to be able to handle this enormous dataset, the risk of overfitting or inefficiency in computations impairs the overall accuracy and generalization capability of the algorithms.

Being that the normal class is the most underrepresented among the classes, class balancing techniques range from adjusting the weights of classes to using oversampling and under-sampling techniques in order to reduce imbalance. By adding more variety and volume to the training set for the normal class, data augmentation may enhance the model's recognition performance. Furthermore, regularization methods like dropout or L2 regularization, in conjunction with model tuning—which includes modifying parameters or investigating different architectures—could lessen overfitting and enhance validation performance. Optimizing the rate of learning could potentially enhance the convergence and overall performance of the model.

**Ethics statements**

**None**

**CRediT author statement**

Prerna Kulkarni: Methodology, Software Nidhi Sarwe: Conceptualization, Writing- Original draft preparation, Writing - Review & Editing Abhishek Pingale: Data Curation, Writing - Review & Editing.YashSarolkar*:* Project administration, Supervision. Dr.Rutuja Kadam: Conceptualization, Validation, Supervision  Dr.Gitanjali Shinde: Validation Dr. Rupali Gangarde: Validation

**Acknowledgments**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors*.*

**Declaration of interests**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

**Supplementary material [OPTIONAL]**

None

**References**

|  |  |
| --- | --- |
| [1] Y. B. Bakare and K. M, “HISTOPATHOLOGICAL IMAGE ANALYSIS FOR ORAL CANCER CLASSIFICATION BY SUPPORT VECTOR MACHINE,” *INTERNATIONAL JOURNAL OF ADVANCES IN SIGNAL AND IMAGE SCIENCES*, vol. 7, no. 2, pp. 1–10, Dec. 2021, doi: 10.29284/IJASIS.7.2.2021.1-10.  [2] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Sep. 2014, doi: https://doi.org/10.48550.  [3] T. Hattiya, K. Dittakan, and S. Musikasuwan, “Diabetic Retinopathy Detection using Convolutional Neural Network: A Comparative Study on Different Architectures,” *Engineering Access*, vol. 7, no. 1, pp. 50–60, 2021, doi: 10.14456/mijet.2021.8.  [4] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based Learning Applied to Document Recognition,” *proc. OF THE IEEE*, 1998, [Online]. Available: http://ieeexplore.ieee.org/document/726791/#full-text-section  [5] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” Dec. 2015, [Online]. Available: http://arxiv.org/abs/1512.00567  [6] M. Sandler, A. Howard, M. Zhu, and A. Zhmoginov, “Sandler\_MobileNetV2\_Inverted\_Residuals\_CVPR\_2018\_paper.pdf,” *ArXiv*, pp. 4510–4520, 2018.  [7] S. M. Fati, E. M. Senan, and Y. Javed, “Early Diagnosis of Oral Squamous Cell Carcinoma Based on Histopathological Images Using Deep and Hybrid Learning Approaches,” *Diagnostics*, vol. 12, no. 8, p. 1899, Aug. 2022, doi: 10.3390/diagnostics12081899.  [8] R. Mohan *et al.*, “OralNet: Fused Optimal Deep Features Framework for Oral Squamous Cell Carcinoma Detection,” *Biomolecules*, vol. 13, no. 7, p. 1090, Jul. 2023, doi: 10.3390/biom13071090.  [9] T. Y. Rahman, L. B. Mahanta, A. K. Das, and J. D. Sarma, “Histopathological imaging database for oral cancer analysis,” *Data Brief*, vol. 29, p. 105114, Apr. 2020, doi: 10.1016/j.dib.2020.105114.  [10] K. Warin, W. Limprasert, S. Suebnukarn, S. Jinaporntham, and P. Jantana, “Automatic classification and detection of oral cancer in photographic images using deep learning algorithms,” *Journal of Oral Pathology & Medicine*, vol. 50, no. 9, pp. 911–918, Oct. 2021, doi: 10.1111/jop.13227.  [11] A. Mohamed, G. E. Dahl, and G. Hinton, “Acoustic Modeling Using Deep Belief Networks,” *IEEE Trans Audio Speech Lang Process*, vol. 20, no. 1, pp. 14–22, Jan. 2012, doi: 10.1109/TASL.2011.2109382.  [12] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, “Extracting and composing robust features with denoising autoencoders,” in *Proceedings of the 25th international conference on Machine learning - ICML ’08*, New York, New York, USA: ACM Press, 2008, pp. 1096–1103. doi: 10.1145/1390156.1390294.  [13] S. Ramya, R. I. Minu, and K. T. Magesh, “An Automated Approach for Identification of Oral Squamous Cell Carcinoma based on 2D-ICNN,” in *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, IEEE, Aug. 2023, pp. 155–161. doi: 10.1109/ICAISS58487.2023.10250540.  [14] C. Dutta, P. Sandhya, K. Vidhya, R. Rajalakshmi, D. Ramya, and K. Madhubabu, “Effectiveness of deep learning in early‐stage oral cancer detections and classification using histogram of oriented gradients,” *Expert Syst*, vol. 41, no. 6, Jun. 2024, doi: 10.1111/exsy.13439.  [15] S. Panigrahi and T. Swarnkar, “Machine Learning Techniques used for the Histopathological Image Analysis of Oral Cancer-A Review,” *Open Bioinforma J*, vol. 13, no. 1, pp. 106–118, Nov. 2020, doi: 10.2174/1875036202013010106.  [16] K. Warin, W. Limprasert, S. Suebnukarn, S. Jinaporntham, and P. Jantana, “Performance of deep convolutional neural network for classification and detection of oral potentially malignant disorders in photographic images,” *Int J Oral Maxillofac Surg*, vol. 51, no. 5, pp. 699–704, May 2022, doi: 10.1016/j.ijom.2021.09.001.  [17] H.-C. Shin *et al.*, “Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning,” *IEEE Trans Med Imaging*, vol. 35, no. 5, pp. 1285–1298, May 2016, doi: 10.1109/TMI.2016.2528162.  [18] H. R. Roth *et al.*, “Improving Computer-Aided Detection Using Convolutional Neural Networks and Random View Aggregation,” *IEEE Trans Med Imaging*, vol. 35, no. 5, pp. 1170–1181, May 2016, doi: 10.1109/TMI.2015.2482920.  [19] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Med Image Anal*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.  [20] M. Das, R. Dash, and S. K. Mishra, “Automatic Detection of Oral Squamous Cell Carcinoma from Histopathological Images of Oral Mucosa Using Deep Convolutional Neural Network,” *Int J Environ Res Public Health*, vol. 20, no. 3, p. 2131, Jan. 2023, doi: 10.3390/ijerph20032131.  [21] I. U. Haq, M. Ahmed, M. Assam, Y. Y. Ghadi, and A. Algarni, “Unveiling the Future of Oral Squamous Cell Carcinoma Diagnosis: An Innovative Hybrid AI Approach for Accurate Histopathological Image Analysis,” *IEEE Access*, vol. 11, pp. 118281–118290, 2023, doi: 10.1109/ACCESS.2023.3326152.  [22] M. Ahmad *et al.*, “Multi-Method Analysis of Histopathological Image for Early Diagnosis of Oral Squamous Cell Carcinoma Using Deep Learning and Hybrid Techniques,” *Cancers (Basel)*, vol. 15, no. 21, p. 5247, Oct. 2023, doi: 10.3390/cancers15215247. |  |