International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

AI-Driven Diagnosis of Oral Malignancies through Machine Learning and Deep Learning

[1] Dr. Nanditha B R, [1] Gagan H D, [1] Sugnyan K, [1] Rakshitha A S, [1] Lekhana D S
Information Science and Engineering, Malnad College of Engineering, Hassan-573202, India

brn@mcehassan.ac.in, appugagan72@gmail.com , asrakshitha4@gmail.com , kerurasugnyan@gmail.com , lekhanadslekhanads@gmail.com.

Abstract: Oral cancer is a serious global health issue. The success of treatment also largely depends on early diagnosis. The differentiation between benign and malignant oral lesions requires specialist expertise, which in itself is a challenge because of the lack of advanced medical facilities. This project hopes to address this gap through an AI-based diagnostic tool developed specifically for improving the early diagnosis and classification of oral lesions. It uses recent advancements in deep learning models like MobileNet, Convolutional Neural Networks, or DenseNet that make analyses of a dataset consisting of oral lesion images classify as benign or malignant. Focusing on subtle visual cues, this system serves as a rapid, non-invasive diagnostic tool that guides healthcare professionals toward making right and timely decisions. Thus, in addition to improving precision in diagnostics, it satisfies a critical demand for scalable automated solutions within areas that may lack access to specialized health care. Ultimately, the system would assist health care providers to provide an early diagnosis that contributes toward better patient outcomes due to timely intervention.

Keywords: Oral cancer, AI-driven diagnostic system, image-based AI models, MobileNet, CNN, DenseNet, lesion classification, benign lesions, malignant lesions, deep learning, non-invasive diagnostics, machine learning.

I. INTRODUCTION

Oral cancer is still a challenge in the global health field, with high mortality rates caused primarily by delayed diagnosis. Survival rates for people diagnosed with oral malignancies are significantly improved through early detection, but many regions lack access to specialized diagnostic resources, especially in underserved areas. Traditional diagnostic techniques are time-consuming and dependent on expert interpretation, which may not be readily available in resource-limited settings. This project aims to utilize artificial intelligence and deep learning technologies in order to provide a scalable, automated solution for the early diagnosis of oral lesions. It seeks to bridge the gap in healthcare access in an effort to improve patient outcomes by allowing for faster and more accurate detection. The accurate differentiation of benign from malignant oral lesions is critical for proper management, but such assessments often require specialized expertise and resources not always available.[1] This project is focused on developing an AI-driven diagnostic system that utilizes machine learning and deep learning models to classify oral lesions with high accuracy. By providing a more accessible method for identifying these lesions, the system aims to improve early diagnosis and increase access to quality care, particularly in areas that face challenges in obtaining expert-level analysis. The goal of this project is to create an AI-powered diagnostic tool capable of distinguishing between benign and malignant oral lesions.

This tool will use deep learning models such as MobileNet, CNN, and DenseNet to process image data for high-accuracy classifications. The goal is to create a fast, non-invasive diagnostic method that can help healthcare professionals deliver precise lesion classifications. Improving early detection of oral malignancies will contribute to timely interventions and better patient outcomes. Also, it attempts to respond to the need for specialized diagnosis through providing an automated and scalable form of screening of cancers in underserved areas.[2] The project to be conducted focuses on an AI-based system to develop a classifier based on images used for differential diagnostics, mainly distinguishing benign and malignant oral lesions. The project will involve developing and testing deep learning models, such as MobileNet, CNN, and DenseNet, in order to check whether these models are useful in the classification of lesions.

The system will be tested on an image dataset to validate its accuracy and reliability as a rapid, non-invasive diagnostic tool. Ultimately, the system aims to be deployable across various healthcare settings, providing an accessible, scalable solution for the early detection of oral cancer, especially in areas lacking access to specialized resources.[3] Future work could expand the dataset and enhance the models to cover a broader range of oral lesions and additional types of malignancies. Oral cancer remains a leading cause of death worldwide, with its early detection being vital to improving survival rates. However, accurately distinguishing between benign and malignant lesions requires specialized medical knowledge that may not always be readily available, particularly in resource-constrained settings. The system will classify lesions as benign or malignant by analyzing a dataset of oral lesion images, thus making diagnoses quicker and more reliable.

The classification process will use the latest models such as MobileNet, CNN, and DenseNet that can identify subtle visual patterns in medical images that the human eye may not easily recognize. These models will be trained to identify salient features of oral lesions, ensuring high accuracy in distinguishing between benign and malignant cases.[4] The proposed solution will provide a fast, non-invasive, and automated diagnostic tool deployable in various healthcare settings. This system is valuable particularly in areas with limited access to special medical expertise and equipment. By supporting healthcare professionals, especially in underserved regions, the system will allow for earlier and more accurate diagnosis of oral cancer, thus leading to better outcomes for patients.

II. RELATED WORK

This article may have provided a detailed evaluation of how AI techniques were applied to the diagnosis of oral cancer. The systematic review will probably have highlighted different models of AI, including deep learning and machine learning algorithms that have been applied for the analysis of medical images and other diagnostic data to find oral cancer. It would evaluate the precision, effectiveness, and promise of these technologies in detecting oral cancers, highlighting their contribution toward enhancing early detection, error reduction in diagnosis, and cost-effectiveness, most importantly in resource-poor settings. The article likely also talks about the challenges and limitations of implementing AI in a clinical setting, such as the need for large and diverse datasets, regulatory approval, and integration with existing medical systems. In conclusion, the review would state that though AI holds promise for enhancing oral cancer diagnosis, it still needs further research and clinical validation to ensure widespread adoption.[5]

Most likely, the article assesses diagnostic accuracy through AI technologies and possibly for OPMD, among other conditions, by applying its knowledge in identifying oral cancers. Through a systematic review and meta-analysis, this paper will aggregate the results from many studies for determining how good AI models- such as deep learning or even some algorithms on machine learning algorithms-work on analyzing medical images on oral cancers and such other related diseases.[6] The article likely concludes that AI offers significant promise in improving diagnostic accuracy, early detection, and clinical decision-making, outperforming traditional diagnostic methods in some cases. However, it might also highlight challenges such as the need for high-quality, annotated datasets, variability in performance across different imaging modalities, and the integration of AI systems into clinical workflows. The findings would suggest that AI has strong potential in enhancing oral cancer diagnosis, but further research, validation, and standardization are necessary before such implementation in clinical practice can be made widespread.[7]

The systematic review will involve assessing the diagnostic accuracy of OCT, a non-invasive imaging technique, which is enhanced with AI algorithms for analyzing images. The review probably concludes that combining AI with OCT has a significant potential to improve early oral cancer detection and diagnosis by making them more accurate, reliable, and timely. The article could also emphasize the benefits of OCT over traditional imaging methods, such as its ability to capture high-resolution, real-time tissue images, while AI can help in identifying subtle features that are indicative of malignancy.[8] However, the review might also point out challenges, including the need for large, annotated datasets, variability of results across different AI models, and potential barriers to clinical adoption. In conclusion, the article brings out a promising future with AI-OCT integration for oral cancer detection, which with further research and development would become a powerful tool for clinicians to improve patient outcomes.[9]

The outcome highlights the promising potential of deep learning models in the early detection of oral cancer. These models, when trained on images captured using smartphones and DSLR cameras, provide an accessible and cost-effective alternative to traditional diagnostic methods. The systematic review indicates that deep learning techniques, especially convolutional neural networks (CNN), have shown high accuracy in identifying oral cancer lesions from visual data. In the meantime, with mobile and DSLR-based image analysis, remote and widespread screening are also possible. This opens up the opportunity for earlier diagnosis in underprivileged or rural areas. Overall, the study brings out how AI is gaining importance in health care, particularly in the case of cancer diagnosis, thereby offering a new avenue to detect early enough to achieve better patient outcomes.[10]

III. DATASET

The dataset contains color images of oral lesions, captured with both mobile cameras and intraoral cameras. These images are to be analyzed and identified as possible oral malignancies. Data collection was done in coordination with health care professionals from various hospitals and colleges of Karnataka, India. There are two main folders: original data and augmented data. The original data folder contains 165 images of benign lesions and 158 images of malignant lesions. These images serve as the foundational dataset for training and evaluating diagnostic models. The augmented data folder includes images generated through augmentation techniques applied to the original dataset. Flipping, rotation, and resizing augmentation methods are used to enhance the variability of the images, making the models trained on this data more robust.

IV. ARCHITECTURE DETAILS

a. CNN Model

Convolutional Neural Networks, or CNNs, are deep learning models especially set up to process visual data like images. They are ideal in image classification, object detection, and image segmentation among other applications because they have the ability to learn from raw data in hierarchical forms. The architecture of CNN allows it to automatically recognize the spatial features such as edges, textures, and many others that are important in doing visual tasks. This process begins with the input image and is then forwarded through multiple layers in the CNN architecture.

The main unit in CNNs is the convolutional layer, where small filters, known as kernels, slide over the image to detect basic features such as edges and textures. These filters learn during the training process to detect specific patterns for a given task. The output of this operation is a feature map that emphasizes

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

significant visual features of the image.[11] To introduce non-linearity and enable the network to capture more complex patterns, an activation function like ReLU (Rectified Linear Unit) is applied after convolution. This would allow the network to grasp more intricate relationships in the data beyond simple linear dependencies.

Through the process of learning, CNNs optimize their filters and weights using backpropagation based on the error between predicted and actual labels. Such automatic feature learning from images makes CNNs incredibly powerful and allows them to solve really complex tasks, such as medical image analysis, facial recognition, or autonomous driving.

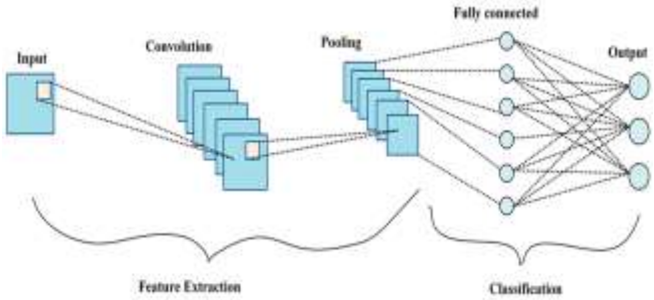

Figure 1 CNN Architecture

b. MobileNet Model

MobileNet is a lightweight, efficient deep learning model designed for mobile and embedded devices where the computational resources are limited. It is based on a kind of convolutional neural network (CNN) architecture that has been optimized for speed and performance without losing too much in terms of accuracy. The main innovation behind MobileNet is depthwise separable convolutions, which split the standard process of a convolution into two much simpler steps. First, a depthwise convolution applies a single filter to each input channel followed by pointwise convolution, which is actually 1x1 convolution. This greatly reduces the number of parameters and computation, which is ideal for resource-constrained environments.

MobileNet's architecture permits it to maintain high accuracy in image classification tasks with being significantly more efficient compared to the traditional CNNs. Its efficiency is achieved with the usage of lightweight convolutions which dramatically reduce the number of operations and the memory required.[12] MobileNet can be adjusted for different levels of computational power by adjusting the width multiplier and resolution multiplier. These parameters allow the model to balance between accuracy and efficiency, making it adaptable for use in a wide range of applications, from real-time object detection to facial recognition on smartphones. The model is trained using the same principles as other CNNs, with filters being learned through backpropagation to detect patterns in input data.

MobileNet is particularly suitable for mobile devices and edge computing, where speed and resource efficiency are key.[13] Its design allows deep learning models to run directly on low processing power devices, reducing the requirement for cloud computing and making predictions and decisions in real-time possible. Therefore, it would be perfect for applications like augmented reality, mobile health care, and on-device image processing.

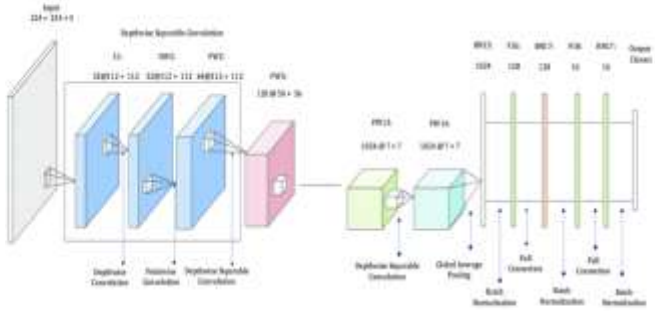

Figure 2 DenseNet Architecture

c. DenseNet

DenseNet or densely connected convolutional networks enhances the traditional convolutional neural network architecture. A new concept, introduced for achieving the depth in dense way by providing connections between various layers, has brought many novelties that exist within this type of connection. Each layer of DenseNet feeds forward into all later layers as well as backwards towards all earlier layers. This dense connectivity serves to alleviate the vanishing gradient problem, improves information flow, and enables faster learning of the network architecture, especially in very deep architectures.[14] In dense blocks, each block will contain more than one layer of convolution. Unlike typical CNN, where a layer receives an input only from the immediately preceding layer, in each layer of a DenseNet block, it receives all the preceding layers within that block as input. This dense connectivity encourages feature reuse, as each layer can directly access the feature maps produced by all preceding layers, leading to more compact and efficient feature representations. Additionally, DenseNet typically uses fewer parameters compared to traditional CNNs due to this feature reuse, making it more efficient for tasks that require deep learning but with fewer resources.

Training a DenseNet is similar to other CNNs; the network learns to recognize patterns and features in the input data through backpropagation. However, the key advantage of DenseNet is that it allows for a deeper and more efficient learning process by reusing features and reducing the number of parameters required.

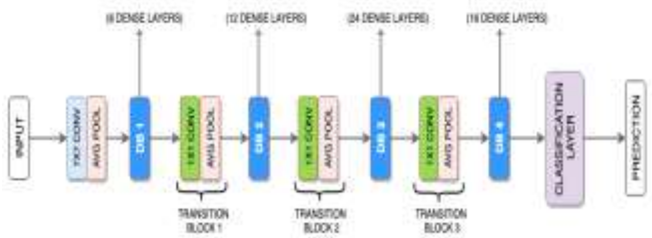

Figure 3 MobileNet Architecture

V. PROPOSED SYSTEM METHODOLOGY

a. Preprocessing Steps

The preprocessing of the oral lesion dataset will be performed by a set of important steps in preparation of images for the training and evaluation of machine learning models, especially CNNs, which ensures that the data is in an optimal form for learning and may therefore enhance the performance and generalization ability of the model.

The first is to resize all the images to the same size, since most neural networks take images as input of the same size. In this dataset, the images can vary in size, so they are resized to a fixed resolution (e.g., 224x224 pixels) so that all the input data is uniform. Resizing helps the input become standardized, and thus makes the model learn from the images of the same size and aspect ratio, and it helps in efficient learning.[15] Next, normalization comes in, where the pixel values are scaled to a value between 0 and 1. This is typically done by dividing each pixel value by 255, because the pixel values range between 0 and 255. Normalization ensures all pixel values have similar magnitudes; this helps the model learn faster during training by stopping large values from dominating the learning process. In addition, to improve model robustness, data augmentation techniques are applied to the dataset including flipping, rotation, and resizing.

These transformations artificially increase the size of the dataset, allowing the model to learn more diverse representations of oral lesions and improving its ability to generalize to unseen data. Finally, for model training, the images are divided into training, validation, and testing sets. Typically, 70-80% of the data is used for training, 10-15% for validation, and the remaining 10-15% for testing. This split ensures that the model is trained on a large portion of the dataset, validated during training to tune hyperparameters, and tested on unseen data to evaluate its performance. Preprocessing also includes the labeling of images as "benign" or "malignant," ensuring that the dataset is correctly annotated for supervised learning.[16] By following these preprocessing steps, the dataset becomes well-suited for training and testing machine learning models designed to detect oral lesions.

b. Model Building

The MobileNet, DenseNet, and CNN architectures process images through a series of layers that progressively extract more complex and abstract features. At first, these models focus on detecting basic visual elements like edges, shapes, and textures. In the first convolutional layers, small filters are applied to the image, capturing low-level features such as edges or contrasts. For oral lesion detection, these early features are essential for distinguishing between different tissue structures. During training, the models learn the most relevant features by adjusting their parameters based on the error between predicted and actual outcomes, using techniques like backpropagation. As the image progresses further within the network, these simple features become chained together to form more complex structures. In traditional CNNs, each subsequent layer can combine the features learned by the previous layers for progressively more complex representations in the model.[17] For instance, when this oral lesion is deeper in, maybe the patterns for such layers are based on how much it's texture-y or shape-y or bounded related that can categorize to whether it is benign or malignant. The separability process in depthwise MobileNet separates this convolution and decomposes the two where they first apply the filters, separately to each of the channel components, and then accumulate all the results to reduce computations, allowing them to have fewer parameters for processing as well as fast runtimes.

The denseNet then goes a mile further using the dense connections because each layer will receive its input from all the prior layers, making it work more in improving the reusage of learned features earlier from the network. This does improve the efficiency of extracting more details as the model focuses. In the oral lesions case, this allows the model to use the information from multiple layers together, so it can identify such subtle features as changes in tissue texture or irregularities in shape that might suggest malignancy. DenseNet's feature reuse leads to a more compact and efficient network, requiring fewer parameters to achieve high performance. Ultimately, these architectures learn to classify images through continuous refinement of their internal parameters. [18] The models enhance their ability to recognize distinguishing characteristics of oral lesions by processing the images and adjusting filters and weights through backpropagation.

With the training on labeled oral images, the network becomes more efficient at determining benign from malignant lesions on the basis of the complicated features that are learned throughout the layers. This creates an efficient diagnostic tool with the potential to classify the oral lesion accurately and further help health professionals in early diagnosis and timely intervention.

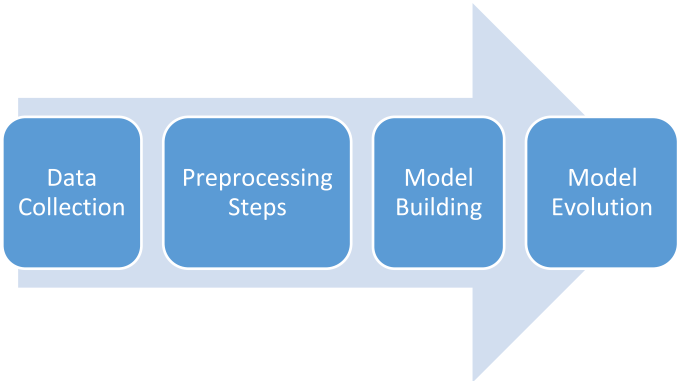

Figure 4 Proposed workflow

VI RESULTS AND DISCUSSION

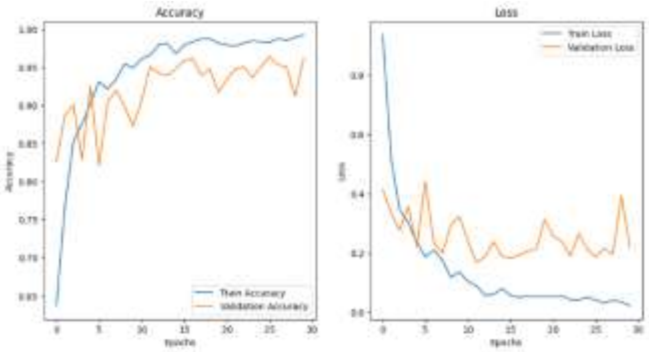

Figure 5 CNN Results

The model shows a strong training performance with accuracy stabilizing around 0.98 and loss near zero, indicating effective learning. However, the fluctuating validation accuracy and loss suggest potential overfitting, as the model performs less consistently on new, unseen data.

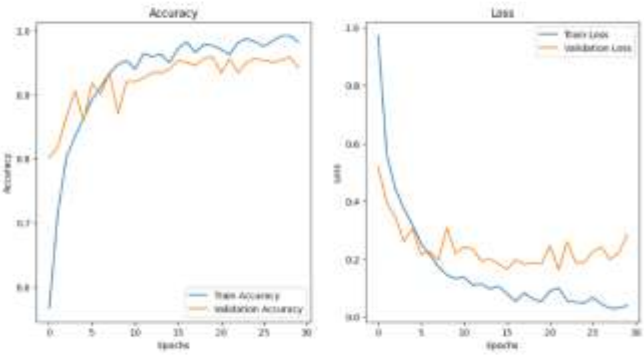

Figure 6 DenseNet Results

The training accuracy shows consistent improvement, stabilizing around 98%, indicating effective learning. However, the validation accuracy fluctuates, and the validation loss slightly increases towards the end, hinting at potential overfitting. This suggests the model performs well on training data but less consistently on unseen data.

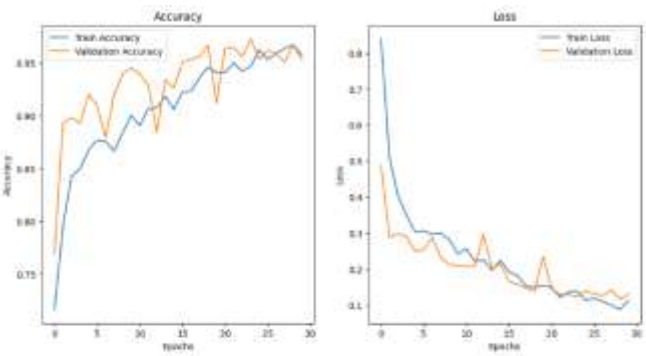

Figure 7 MobileNet Results

The model shows strong performance with both training and validation accuracy increasing and converging around 0.95. Similarly, training and validation loss decrease and converge around 0.1. This indicates effective learning with minimal overfitting, as the metrics are closely aligned.

Table 1 Performance Comparison

Model Name	Test Accuracy	Test Loss
CNN	0.9758	0.1462
MobileNet	0.9714	0.0918
DenseNet	0.9538	0.1727

The CNN model is showing the highest accuracy at 97.58%, which depicts that it performs the best in classifying the test data, though its test loss a bit higher at 0.1462. In terms of accuracy, MobileNet is very close to CNN at 97.14%, but stands apart with the lowest test loss of 0.0918, which depicts it to generalize well with minimal errors. DenseNet also works quite well, but has the least accurate result with 95.38% and also maximum test loss at 0.1727, it may indicate that is comparatively worse than the others. However, CNN and MobileNet top the results; besides, MobileNet performed little better than CNN concerning loss, whereas DenseNet needs to be tamed a bit for further improvement.

VII REFERENCES

[1] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep Learning Applications in Medical Image Analysis," *IEEE Access*, vol. 6, pp. 9375–9379, Dec. 2017, doi: 10.1109/ACCESS.2017.2788044.

[2] P. K. Mall *et al.*, "A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities," *Healthcare Analytics*, vol. 4, p. 100216, Dec. 2023, doi: 10.1016/J.HEALTH.2023.100216.

[3] H. P. Chan, R. K. Samala, L. M. Hadjiiski, and C. Zhou, "Deep Learning in Medical Image Analysis," *Adv Exp Med Biol*, vol. 1213, p. 3, 2020, doi: 10.1007/978-3-030-33128-3_1.

[4] 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.

[5] V. P. Veeraraghavan, G. Minervini, D. Russo, M. Cicciù, and V. Ronsivalle, "Assessing Artificial Intelligence in Oral Cancer Diagnosis: A Systematic Review," *Journal of Craniofacial Surgery*, Nov. 2024, doi: 10.1097/SCS.0000000000010663.

[6] K. Warin and S. Suebnukarn, "Deep learning in oral cancer- a systematic review," *BMC Oral Health*, vol. 24, no. 1, p. 212, Dec. 2024, doi: 10.1186/S12903-024-03993-5.

[7] R. K. Sahoo *et al.*, "Diagnostic performance of artificial intelligence in detecting oral potentially malignant disorders and oral cancer using medical diagnostic imaging: a systematic review and meta-

analysis," *Frontiers in Oral Health*, vol. 5, p. 1494867, 2024, doi: 10.3389/FROH.2024.1494867/FULL.

[8] M. García-Pola, E. Pons-Fuster, C. Suárez-Fernández, J. Seoane-Romero, A. Romero-Méndez, and P. López-Jornet, "Role of Artificial Intelligence in the Early Diagnosis of Oral Cancer. A Scoping Review," *Cancers 2021, Vol. 13, Page 4600*, vol. 13, no. 18, p. 4600, Sep. 2021, doi: 10.3390/CANCERS13184600.

[9] W. Jerjes, H. Stevenson, D. Ramsay, and Z. Hamdoon, "Enhancing Oral Cancer Detection: A Systematic Review of the Diagnostic Accuracy and Future Integration of Optical Coherence Tomography with Artificial Intelligence," *Journal of Clinical Medicine 2024, Vol. 13, Page 5822*, vol. 13, no. 19, p. 5822, Sep. 2024, doi: 10.3390/JCM13195822.

[10] T. Thakuria *et al.*, "Deep learning for early diagnosis of oral cancer via smartphone and DSLR image analysis: a systematic review," *Expert Rev Med Devices*, Nov. 2024, doi: 10.1080/17434440.2024.2434732.

[11] "What is a Convolutional Neural Network (CNN)? | Definition from TechTarget." Accessed: Dec. 19, 2024. [Online]. Available: https://www.techtarget.com/searchenterpriseai/definition/convolutional-neural-network

[12] J. de Chauveron, M. Unger, G. Lescaille, L. Wendling, C. Kurtz, and J. Rochefort, "Artificial intelligence for oral squamous cell carcinoma detection based on oral photographs: A comprehensive literature review," *Cancer Med*, vol. 13, no. 1, p. e6822, Jan. 2024, doi: 10.1002/CAM4.6822.

[13] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Apr. 2017, Accessed: Dec. 19, 2024. [Online]. Available: https://arxiv.org/abs/1704.04861v1

[14] D. U. Kapoor *et al.*, "AI illuminates paths in oral cancer: transformative insights, diagnostic precision, and personalized strategies," *EXCLI J*, vol. 23, p. 1091, Jan. 2024, doi: 10.17179/EXCLI2024-7253.

[15] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, Sep. 2014, Accessed: Dec. 19, 2024. [Online]. Available: https://arxiv.org/abs/1409.1556v6

[16] L. Perez and J. Wang, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning," Dec. 2017, Accessed: Dec. 19, 2024. [Online]. Available: https://arxiv.org/abs/1712.04621v1

[17] M. H. Arzmi *et al.*, "Deep Learning in Cancer Diagnostics," 2023, doi: 10.1007/978-981-19-8937-7.

[18] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 2261–2269, Nov. 2017, doi: 10.1109/CVPR.2017.243.