2025 5th Asian Conference on Innovation in Technology (ASIANCON)
Pune, India. Aug 22-23, 2025

Evaluating Deep Learning Models for Histopathologic Oral Cancer Detection

Arunabha Tarafdar
SCSET, Bennett University
Greater Noida, India
arunabhatarafdar@gmail.com

Atrija Haldar
SCSET, Bennett University
Greater Noida, India
atrija04@gmail.com

Susmita Das
SCSET, Bennett University
Greater Noida, India
susmitad900@gmail.com

Sounak Sadhukhan
SCSET, Bennett University
Greater Noida, India
sounaks.cse@gmail.com

2025 5th Asian Conference on Innovation in Technology (ASIANCON) | 979-8-3503-5699-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ASIANCON66527.2025.11281308

***Abstract*—Early detection of oral cancer by histopathological analysis, involves examining tissue samples under a microscope and enchancing treatment results. However, manual inspection is time-consuming and often differs between pathologists. In this work, we have investigated the potential of deep learning, namely convolutional neural networks (CNNs), to help us and researchers automate the identification of malignant patterns in oral tissue histopathology images. Implementing floating point operations per second (FLOPs) as a parameter, we have compared the accuracy and computational requirements of a number of popular CNN architectures, including MobileNetV2, DenseNet, Xception, Inception, VGG19, and ResNet. Due to its lightweight design, MobileNetV2 has been the most efficient of all and achieved the maximum accuracy of 87%. With an accuracy of 83%, DenseNet performed commendably, due to its robust gradient flow and feature reuse capabilities. VGG19 and ResNet were comparatively behind the former, while Xception and Inception exhibited marginally inferior performance than envisioned. Our observations have also shown that, in certain specialized tasks, such as oral cancer diagnosis, well-optimized models may outperform more sophisticated ones. According to our results and inferences, effective CNNs like MobileNetV2 have great potential for application in clinical contexts and in the medical field, where accuracy and speed are crucial.**

***Index Terms*—Oral Cancer, Histopathological Image Classification, Convolutional Neural Networks, Deep Learning Architectures, MobileNetV2, DenseNet, Computational Complexity, Medical Image Analysis, Automated Diagnosis, Machine Learning in Pathology.**

I. Introduction

Histopathology images are an important diagnostic tool in oncology, particularly in the detection and treatment of cancer. Cancer detection requires the examination of tissue samples under a microscope to identify abnormal cell structures and behaviors. This procedure is of great importance as it reveals significant information regarding the morphology of tissues, allowing medical professionals to properly diagnose cancer. The significance of histopathologic examination is that it can guide treatment selection and predict patient outcomes [4]. Pathologists can ascertain the aggressiveness and type of cancer by examining tissue samples, which guides the selection of treatment modalities. Additionally, histopathologic examination identifies suspected biomarkers for cancer and advances the development of tailored therapy. Recent research has proved the value of histopathology image study in enhancing cancer diagnosis and treatment. Several studies established that artificial intelligence (AI)-based practices can process tissue samples more effectively and with greater precision than human pathologists, that set the stage for the success of more effective tools for the diagnosis of cancer. As the current trend indicates, the analysis of histopathology images is increasingly using AI and machine learning (ML) to enhance diagnostic accuracy and efficacy. The use of deep learning (DL) techniques, specifically convolutional neural networks (CNNs), has been promising in the examination of tissue samples. The recent advent of digital pathology platforms has made tissue samples more efficient and collaborative to share and analyze. Several databases are being used to study histopathology images, including the Cancer Genome Atlas (TCGA) and the National Cancer Institute's (NCI) Clinical Proteomic Tumor Analysis Consortium (CPTAC) [14], etc. These databases provide a wealth of information on cancer genomics, proteomics, and histopathology, facilitating the development of more effective cancer diagnostic and therapeutic strategies.

Some articles provide a vivid idea about the recent studies in this field. The diagnosis of oral squamous cell carcinoma (OSCC) has been enhanced through deep learning (DL) frameworks, reaching accuracy levels between 92% and 96%. [13]. A study utilized the VGGNet architecture, a paradigm for image classification due to its depth and simplicity [11]. Additionally, the DenseNet design enhances feature reuse and gradient flow, leading to efficient training and superior performance on medical image analysis. These architectures demonstrate potential for accurate OSCC diagnosis.

Hence, it is understandable that histopathology image analysis is a pivotal step in cancer detection and treatment, providing the required information on tissue structure to single out malignant cells. However, manual expounding of histopathology image analysis could be tedious with a lot of inaccuracies, highlighting the need for automated cancer detection methods. Our research is committed to evaluate the efficiency of various DL models for the detection of histopathologic oral cancer by conducting a comparative analysis. [1]. Six architectures, namely VGG19, DenseNet,MobileNetV2,ResNet50, InceptionV3, and Xception—widely recognized for their efficiency across various computer vision tasks, especially in the domain of medical image analysis, are the cynosure of attention in this experiment [12]. In order to determine the most efficient model in identifying oral cancer from histopathology images, we have assessed the accuracy of many models in this study. Beyond our research, this approach has got the potential of revolutionizing automated diagnostic systems and improve the speed and accuracy of cancer detection in clinics. We have employed a sizable and detailed dataset with 5192 images that includes 2 classes- Normal and OSCC, with 4946 images for training, 120 for validation and 126 for testing. The main objective of each models' rigorous training and testing was to determine the efficiency in distinguishing between benign and cancerous tissue in an image.

The paper is assembled as follows: *Section II* demonstrates the methodology with details about the CNN models and dataset specifics, *Section III* reviews the results and analyse the metrics presented by the models and lastly *Section IV* concludes the work along with future scope of this research.

II. METHODOLOGY

In our work, we have implemented various established CNN architectures. Framework of the models have been discussed in details.

A. MobileNet

MobileNet is particularly perfect for resource-limited applications, as it extracts features effectively at the same time as decreasing computation power through separable convolutions(depthwise) [9]. By decoupling circulate-channel filtering from spatial filtering, MobileNet effectively establishes a balance between version size and latency, turning in the highest quality overall performance in the limits. The architecture of MobileNet is depicted in Fig. 1.

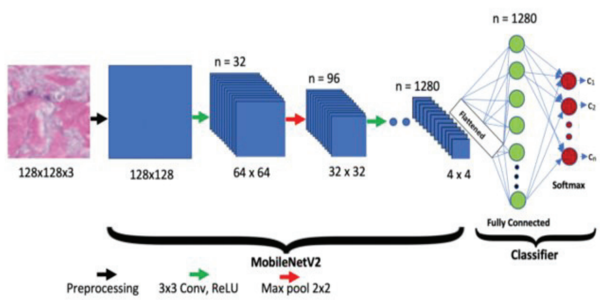


Fig. 1: MobileNet

B. VGG19

VGG19 is a 19-layer deep convolutional neural network architecture consisting of 16 convolutional layers and 3 fully connected (FC) layers [5]. It is a longer variant of the previous VGG19 model, aimed at further improving performance by including more convolutional layers with the same architectural concepts as depicted in Fig. 2. The architecture employs 3x3 convolutional filters, that are well suited to extract fine spatial details from images. 2x2 max pooling layers are employed to shrink the spatial dimensions but preserve the important features. These small receptive fields as well as repeating layer structure yield a uniform and deep network that are both strong and beautiful [10]. VGG19, trained on the ImageNet dataset, obtained a top-5 classification accuracy of 92.9%, which at the time of its release made it one of the most accurate models for many image classification tasks. The deep structure of the model enables it to learn intricate representations and nuanced features in image data, which is especially helpful for applications like object recognition, medical imaging, and histopathological analysis.

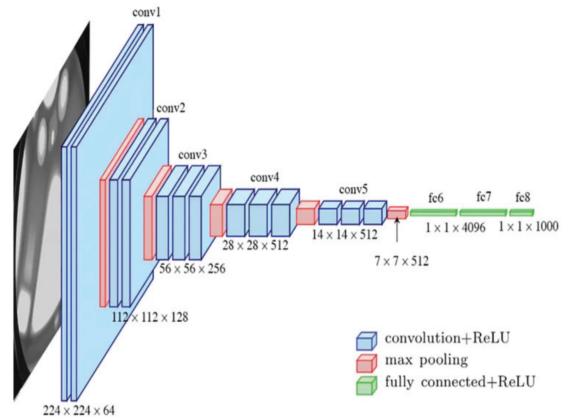


Fig. 2: VGG19 Architecture

C. ResNet-50

One of the deeper CNN structures is ResNet-50. Its 50 layers include convolutional and FC layers as shown in Fig. 3. Residual connections are utilized by ResNet-50 to help train the model and provide higher accuracy. It achieved 92.9% top-5 accuracy and 76.1% top-1 accuracy on the ImageNet dataset. Due to its high accuracy and performance, ResNet-50 is commonly implemented for tasks, such as object recognition, image classification, and segmentation.

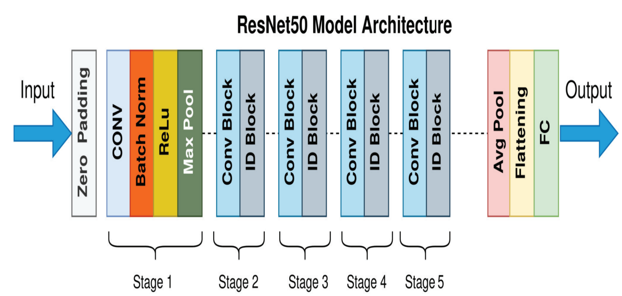


Fig. 3: ResNet 50 Architecture

D. Xception

Xception is another deep CNN architecture as depicted in Fig.4. Depthwise separable convolutions are used to reduce the computational cost. It achieved 79.0% top-1 accuracy and 94.5% top-5 accuracy on the ImageNet dataset. It is well known for its exceptional accuracy and efficiency in picture

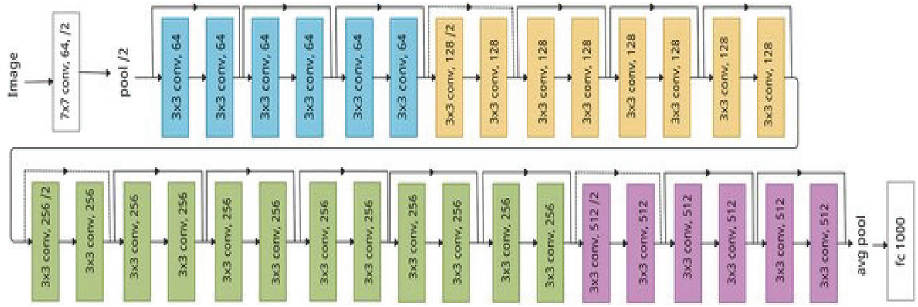


Fig. 4: XceptionNet Architecture

categorization tasks. Xception is widely used in computer vision applications.

E. DenseNet-201

A 201-layer deep CNN architecture is called DenseNet-201 [7]. Dense connections improve feature reuse and reduce parameters. DenseNet-201 achieved 77.3% top-1 accuracy

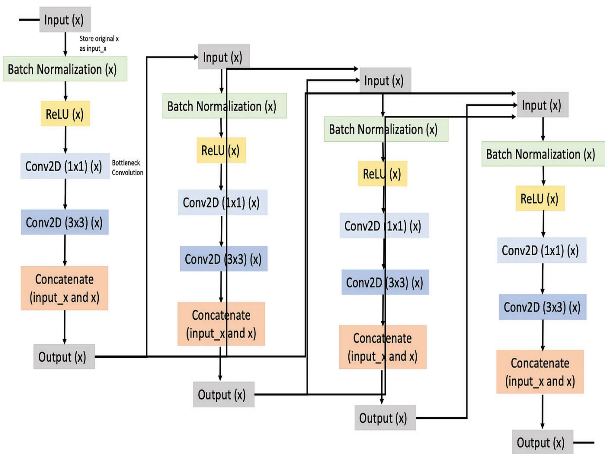


Fig. 5: Densenet 201 Architecture

and 93.6% top-5 accuracy using the ImageNet dataset. It is well known for its exceptional accuracy and efficiency in picture categorization tasks. DenseNet-201 as shown in Fig.5, is frequently used in computer vision applications.

F. Inception v3

Inception v3 is a deep CNN architecture with 22 layers [3]. It uses inception modules to improve feature extraction and reduce parameters. Inception v3 achieved 80.2% top-1 accuracy and 95.1% top-5 accuracy on ImageNet. It is known for its high accuracy and efficiency in image classification tasks. Inception v3 is widely used in computer vision applications.

G. Dataset

Selecting an appropriate dataset is critical in ML applications, particularly for scientific image analysis. The Histopathological Imaging Database for Oral Cancer Analysis is a curated collection of high-resolution microscopic images designed to support research in the early detection

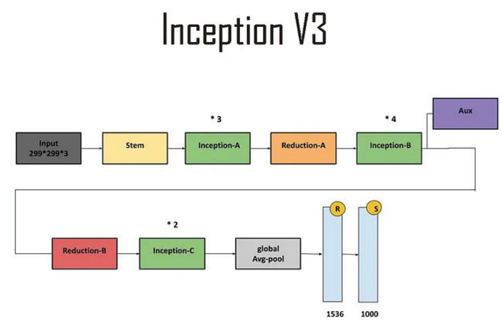
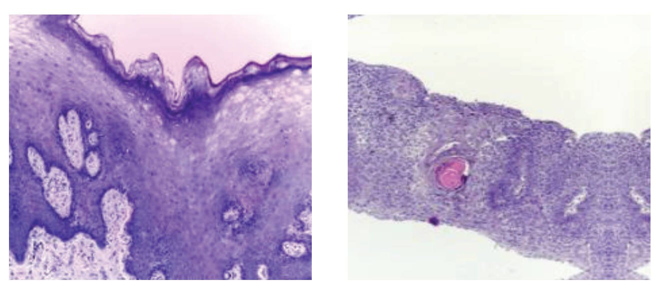


Fig. 6: InceptionNet Architecture

and classification of OSCC [8]. Developed to aid DL and ML applications in digital pathology, the dataset comprises 1224 hematoxylin and eosin (H&E) stained images obtained from biopsy samples collected from 230 patients at Ayursundra Healthcare Pvt. Ltd. and Dr. B. Borooah Cancer Institute (BBCI) in Guwahati, India. The dataset includes images at two magnification levels—100x and 400x—captured using a Leica ICC50 HD microscope. Specifically, it contains 290 images of normal oral epithelium and 934 images of OSCC, offering a balanced distribution for binary classification tasks. The availability of different magnifications makes it ideal for multi-scale analysis, allowing researchers to explore tissue-level patterns and cellular structures with greater flexibility.



(a) Normal Cell (b) Oral squamous Cell

Fig. 7: Distinction Between Different Cells

The images are thoroughly annotated by medical experts to ensure clinical accuracy, making the dataset reliable for both

TABLE I: Performance Comparison between various Models

Model	Time Complexity (FLOPs)	Layers	Accuracy on Medical Images	Year of Release	One-Line Comment
VGG19	∼19.6B FLOPs	19	Moderate	2014	Slightly deeper than VGG16
ResNet-50	∼3.8B FLOPs	50	High	2015	Robust residual learning
Inception	∼6.0B FLOPs	48	Very High	2016	Dense connectivity excels in feature reuse
Xception	∼8.0B FLOPs	71	High	2017	Depthwise separable
DenseNet-201	∼20.0B FLOPs	201	Very High	2017	Dense connectivity excels in feature reuse
MobileNet v2	∼0.3B FLOPs	53	Moderate	2018	Lightweight and efficient

academic research and practical model development. As CNN models are used for study of accuracy, we have augmented the images to increase the image count. Creation of additional images has been performed in order to increase the images used for training and obtain better results. The augmentation operation like rotation, width shift, brightness and contrast adjustments and 90-degree turns have been performed for augmentation. The dataset now has upgraded to 4946 images for training, 120 for validation and 126 for testing. Few sample of the dataset are provided in Fig. 7.

III. Results and Analysis

This article summarizes the investigative study that analyzes the performance of a number of CNN models architectures for the purpose of classifying cancerous patterns from images obtained through histopathological examinations. The main focus of this research is to evaluate and benchmark the effectiveness of these models regarding imaging processes in medicine, particularly those concerned with detecting cancer. This study aims at assessing the classification accuracy of each architecture along with the models, complexities, including computational, layer depth, release year, and overall histopathological appropriateness. These factors are highly essential in model selection for actual healthcare situations because they require high levels of accuracy and operational efficiency.

As a reference, the study includes Table I, capturing the essential details pertaining to the performance of each architecture in relation to their diagnosed accuracy and the sophistication of the model used. This comparison of CNN models sharpened the focus for critical users in need of a model tailored for cancer detection via histopathological imagery, including advanced medical imaging procedures, DL applications, and other pertinent domains.

TABLE II: Metric Comparison

MODEL	Accuracy	Precision	Recall
MobileNetV2	**87%**	**0.86**	**0.79**
DenseNet	83%	0.77	0.75
Xecption	82%	0.76	0.74
Inception	79%	0.72	0.62
VGG19	77%	0.71	0.77
ResNet	75%	0.38	0.5

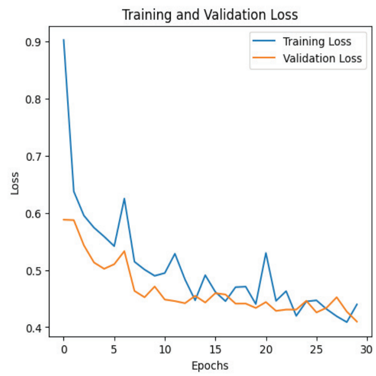
While evaluating the six well-known CNN models, MobileNetV2 provided best results. It produced 87% accuracy in detecting cancer patterns in tissue images. Even though MobileNetV2 is the simplest in terms of FLOPs, this model performed better than other models that need more computing power. It uses special convolutions to compute faster and with accuracy, that makes it great for medical imaging where quick results are needed. As expected, DenseNet, turned out to be effective in obtaining the second-best accuracy out of all the models we tested. Even though DenseNet needs the most computing power, because it has so many layers and numbers to crunch how well it works makes up for how much it needs to run. The model makes good use of reusing features and moving information, which helps in distinguishing tiny details in pictures of tissue samples. Their accuracy scores as shown in Table II, strengthens their position as benchmark CNN models for histopathological cancer detection, offering a strong foundation for further research and practical implementations in the field of medical image analysis.

Models such as Xception and Inception also delivered commendable results, with accuracies of 82% and 79%, respectively. Their architecture designs—focusing on multi-scale feature extraction through depthwise separable convolutions (Xception) and inception modules (Inception)—make them well-suited for heterogeneous medical images where visual cues exist at varying scales.

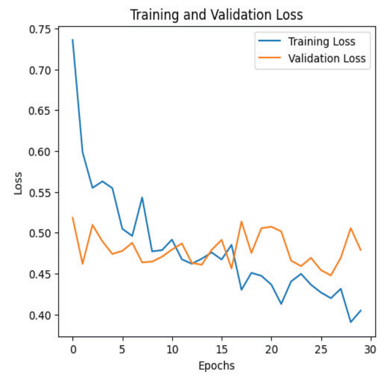
In contrast, VGG19 and ResNet, despite their historical success in various computer vision benchmarks, lagged behind in this specific application, achieving 77% and 75% accuracy, respectively. These relatively lower scores may stem from their deeper, more traditional architectures, which could lead to overfitting or insufficient adaptability to the subtle and complex features present in histopathological data. Observing the average precision and recall value we find almost the same pattern repeated as we see in accuracy. MobilenetV2 emerges victorious with the highest precision and recall values followed by Densenet and the rest.

The training and validation loss curves for the models being evaluated are shown in Fig. 8. In all graphs, MobileNetV2's performance is particularly noteworthy, displaying a consistently steady and smooth loss progression as shown in Fig. 8b. MobileNetV2 maintains a balanced trajectory where training and validation losses converge efficiently, in contrast to other models that might show notable fluctuations or overfitting tendencies.

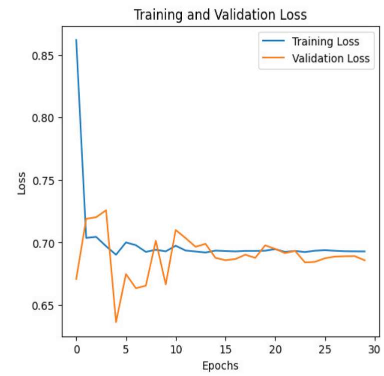
MobileNetV2 exhibits a dependable learning pattern as the number of epochs rises; its validation loss stabilizes without exhibiting unpredictable behavior, suggesting that the model is neither underfitting nor overfitting. This stability at later training stages indicates that MobileNetV2 is successfully identifying the underlying patterns in the data and is generalizing well to validation samples that have not yet been



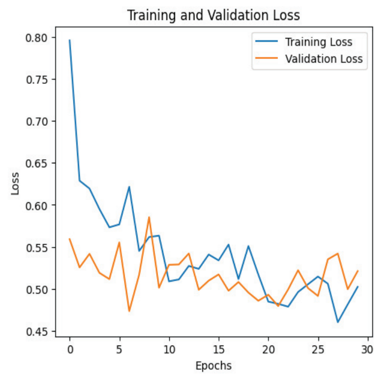
(a) DenseNet



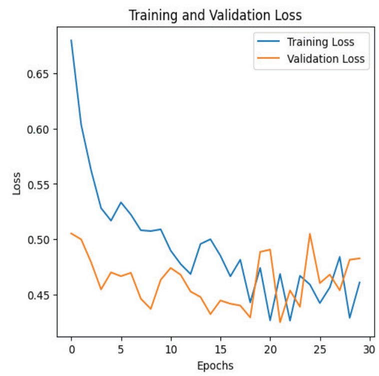
(b) MobileNet



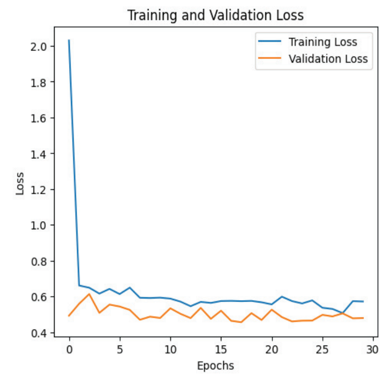
(c) ResNet



(d) Inception



(e) Xception



(f) VGG19

Fig. 8: Training and Validation Loss for a)DenseNet, b)MobileNet, c) ResNet, d)InceptionNet, e)XceptionNet and f)VGG19

seen. Overall, the graphs affirm that MobileNetV2 is a well-suited architecture for our task, yielding dependable and robust performance throughout training. Overall, the results highlight a crucial insight: deeper or more complex models do not always guarantee superior performance, especially in domains where the dataset is domain-specific and rich in fine-grained visual detail. Lightweight architectures like MobileNetV2 offer a compelling balance between computational efficiency and diagnostic accuracy, making them particularly well-suited for deployment in clinical and resource-constrained settings.

IV. Conclusion

The primary goal of this study was to investigate the effectiveness of Convolutional Neural Networks (CNNs) in classifying histopathological images, especially when it comes to cancer detection. Six well-known and benchmark CNN architectures—MobileNetV2, DenseNet, VGG19, Inception, Xception, and ResNet—all of which have been widely used for both medical and non-medical image classification tasks, were used in our comparative analysis to accomplish this.

For researchers commencing studies in the field of histopathological image analysis, this article is a fundamental resource. We offer a transparent performance baseline by using well-known benchmark models, emphasizing the accuracy, computational complexity (FLOPs), and year of launch of each model. This paper provides an improved understanding of the performance of state-of-the-art models for medical images and paves the way for future models. Among the models evaluated, MobileNetV2 was the most efficient, achieving the best accuracy with a small network. DenseNet also achieved good score for its price: expensive to compute but with excellent classification output. These results emphasize that both light and deep networks have their related benefits based on the application reference [2].

In future scope, customized models can be developed, which are adapted to hyperparameter attitude, architectural optimization and especially histopathological data [6]. Additional assessment measurements such as accuracy, recall, F1 score and AUC will also increase the strength of model verification. Training models on large data sets and computer text and clothing methods are likely to give rise to more accurate and reliable clinical equipment.

References

[1] Cancan Chen, Shan Zheng, Lei Guo, Xuebing Yang, Yan Song, Zhuo Li, Yanwu Zhu, Xiaoqi Liu, Qingzhuang Li, Huijuan Zhang, et al. Identification of misdiagnosis by deep neural networks on a histopathologic review of breast cancer lymph node metastases. *Scientific reports*, 12(1):13482, 2022.

[2] GAI Devindi, DMDR Dissanayake, SN Liyanage, FBAH Francis, MBD Pavithya, NS Piyarathne, PVKS Hettiarachchi, RMSGK Rasnayaka, Ruwan Duminda Jayasinghe, Roshan G Ragel, et al. Multimodal deep convolutional neural network pipeline for ai-assisted early detection of oral cancer. *IEEE Access*, 2024.

[3] Na Dong, Li Zhao, Chun-Ho Wu, and Jian-Fang Chang. Inception v3 based cervical cell classification combined with artificially extracted features. *Applied Soft Computing*, 93:106311, 2020.

[4] Suliman Mohamed Fati, Ebrahim Mohammed Senan, and Yasir Javed. Early diagnosis of oral squamous cell carcinoma based on histopathological images using deep and hybrid learning approaches. *Diagnostics*, 12(8):1899, 2022.

[5] Kanwarpartap Singh Gill, Vatsala Anand, and Rupesh Gupta. An efficient vgg19 framework for malaria detection in blood cell images. In *2023 3rd Asian Conference on Innovation in Technology (ASIANCON)*, pages 1–4. IEEE, 2023.

[6] Zewen Li, Fan Liu, Wenjie Yang, Shouheng Peng, and Jun Zhou. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, 33(12):6999–7019, 2021.

[7] Pablo Ormeño-Arriagada, Eduardo Navarro, Carla Taramasco, Gustavo Gatica, and Juan Pablo Vásconez. Deep learning techniques for oral cancer detection: Enhancing clinical diagnosis by resnet and densenet performance. In *International Conference on Applied Informatics*, pages 59–72. Springer, 2024.

[8] Tabassum Yesmin Rahman, Lipi B Mahanta, Anup K Das, and Jagannath D Sarma. Histopathological imaging database for oral cancer analysis. *Data in brief*, 29:105114, 2020.

[9] Archana Saini, Kalpna Guleria, and Shagun Sharma. A pre-trained mobilenetv2 model for oral cancer classification. In *2023 1st DMIHER International Conference on Artificial Intelligence in Education and Industry 4.0 (IDICAIEI)*, volume 1, pages 1–6. IEEE, 2023.

[10] Disha Sharma, Vidya Kudva, Vathsala Patil, Adarsh Kudva, and Rahul S Bhat. A convolutional neural network based deep learning algorithm for identification of oral precancerous and cancerous lesion and differentiation from normal mucosa: a retrospective study. *Engineered Science*, 18(15):278–287, 2022.

[11] Mohd Umair Ali Siddique, Sonu Moni Rabha, Janoo Periwal, Nupur Choudhury, and Rupesh Mandal. Early detection of oral cancer using image processing and computational techniques. In *NIELIT's International Conference on Communication, Electronics and Digital Technologies*, pages 37–54. Springer, 2024.

[12] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[13] Yunhan Tan, Zhihan Wang, Mengtong Xu, Bowen Li, Zhao Huang, Siyuan Qin, Edouard C Nice, Jing Tang, and Canhua Huang. Oral squamous cell carcinomas: state of the field and emerging directions. *International journal of oral science*, 15(1):44, 2023.

[14] John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013.