2024 Second International Conference on Advances in Information Technology (ICAIT-2024)

Oral Cancer detection using Histopathology Images

M.Meyyappan
Computing Technologies
SRM Institute of Science and Technology
kattankulathur, Chennai, India
mm0177@srmist.edu.in

Aniket Verma
Computing Technologies
SRM Institute of Science and Technology)
kattankulathur, Chennai, India
av1597@srmist.edu.in

Aarush Kaushik
Computing Technologies
SRM Institute of Science and Technology)
kattankulathur, Chennai, India
ak3033@srmist.edu.in

Sathyapriya L
Computing Technologies
SRM Institute of Science and Technology
kattankulathur, Chennai, India
sathyapl@srmist.edu.in

Suresh Anand M
Computing Technologies
SRM Institute of Science and Technology
kattankulathur, Chennai, India
suresham1@srmist.edu.in

Selvanayaki Kolandapalayam Shanmugam
Department of Mathematics and Computer Science
Ashland University OH, USA
skolanda@ashland.edu

2024 Second International Conference on Advances in Information Technology (ICAIT) | 979-8-3503-8386-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICAIT61638.2024.10690424

***Abstract*— To increase the chances of survival for the millions of individuals impacted by oral cancer (OC), early detection of the disease is essential. More than 177,384 people died from OC in 2018, and people in the low- to middle-class demographic are primarily affected. The histological examination is the antiquated method of detecting mouth cancer. We are utilizing Deep Learning algorithms to identify OC and automate the process. This paper suggests using histopathology photos to identify OC by building CNN-based models using a transfer learning (TL) technique, then assembling the models to get the best accuracy possible. DenseNet-201, Resnet-50, and EfficientNet-B3 will be used to train the TL models using histopathology pictures. When the three models are combined, the final model has an accuracy of 93.16%, which enables the model to be used in real time and is useful for categorizing images of OC.**

***Keywords*—OC, Deep Learning, CNN, Transfer Learning.**

I. INTRODUCTION

An important worldwide health concern is OC, for which improved early detection techniques are urgently needed. Improving patient outcomes and lowering the death rate linked to OC depend critically on early detection of the disease. Conventional diagnostic approaches, primarily histopathological examination, remain pivotal in identifying OC. However, the inherent limitations of these methods, including their time-consuming nature and susceptibility to human error, underscore the pressing need for innovative and efficient diagnostic techniques.

CNNs have shown great promise as an automated medical image processing tool in recent years, with the potential to completely transform the diagnostic process. These DL algorithms demonstrate remarkable capabilities in recognizing patterns and features within biological images, including those derived from oral histopathology slides. Of particular concern is the escalating incidence of oral squamous cell carcinoma (OSCC) in Asian nations, attributable in part to prevalent lifestyle practices such as alcohol consumption and tobacco use. The rising burden of

OSCC underscores the urgency of implementing effective diagnostic strategies to mitigate its impact. The field of cancer research has witnessed significant growth, resulting in the accumulation of extensive datasets comprising clinical and histopathological information. These datasets provide insightful information about the molecular and cellular features of cancer, which aids in the creation of cutting-edge diagnostic instruments and treatment plans.

The World Health Organization (WHO) classification system categorizes OSCC into distinct subtypes based on histological features, emphasizing the importance of precise diagnostic criteria for guiding treatment decisions. Accurate classification of OC subtypes is critical for implementing tailored therapeutic approaches and optimizing patient outcomes. Against this backdrop, our research endeavors to harness the potential of CNN models to differentiate between malignant and non-malignant oral histopathology images accurately. Our goal is to improve the efficiency of OC identification and increase diagnostic accuracy by utilizing DL algorithms and TL approaches.

To overcome the challenges presented by short annotated datasets, we employ TL techniques with previously trained CNN models, such as DenseNet-201, ResNet-50, and Efficient-B3. Our diagnostic algorithms perform better and the training process can be sped up by taking advantage of the knowledge stored in these pre-trained models. With its significant effects on patient prognosis and treatment outcomes, the importance of early cancer identification cannot be emphasized. Early diagnosis enables timely intervention, facilitating the implementation of curative treatments and potentially improving long-term survival rates among patients with OC.

OC, encompassing a spectrum of malignancies affecting the oral cavity and adjacent structures, represents a formidable public health challenge. Among the various subtypes of OC, OSCC predominates, accounting for a substantial proportion of cases globally. OC is mostly caused by behavioral risk factors, such as drinking alcohol, chewing betel nuts, smoking cigarettes, and having an unhealthful diet. Addressing these modifiable risk factors through public health interventions remains crucial in reducing the burden of OC incidence.

The influence of late-stage diagnosis on OC prognosis underscores the critical importance of early detection initiatives. Timely identification of suspicious lesions through routine screening programs and diagnostic tests can facilitate early intervention, potentially improving treatment outcomes and enhancing patient survival rates.

Current developments in DL and artificial intelligence hold tremendous promise for enhancing the efficiency and accuracy of OC diagnostics. By leveraging SOTA CNN models and image analysis algorithms, we can augment the capabilities of healthcare providers in diagnosing OC with greater precision and confidence. To put it briefly, our research aims to create reliable CNN models that can reliably categories images of oral histopathology by utilizing the capabilities of DL and TL approaches. By our work, we hope to enhance the results for patients and increase access to healthcare globally by expanding the field of OC diagnostics.

The numerous shortcomings of the current body of study are outlined in section two. Section three covers the technical aspects of putting our suggested strategy into practice. The research's implementation outcome is covered in Section 4. The piece concludes with closing thoughts and suggestions for future improvement in the last part.

II. LITERATURE SURVEY

According to a study, dermatologists and deep neural networks (DNNs) can both identify skin cancer with similar accuracy. This highlights how deep learning (DL) methods are available to improve the accuracy and efficacy of skin cancer detection in medical imaging and diagnosis. Experts found that the DNN's diagnostic ability was comparable to physicians after training on a big dataset of photos of skin lesions [1].

This study addresses the challenge of limited training data in OC diagnosis by proposing a few-shot learning framework. Using a prototypical network with two feature extractors for prototypical and query features, the approach reduces the need for extensive data. A customized loss function further improves performance. Experiments on a histopathological image dataset demonstrate that this method outperforms traditional models in OC detection [2].

POC devices have the potential to improve survival and quality of life, and a quick diagnosis is important for lowering the death rate from OC. Traditional diagnostic methods like biopsies have limitations and are not ideal for resource-limited settings. This review highlights various POC platforms for detecting OC biomarkers, including immunosensors and imaging methods. It also discusses current commercial devices and the challenges of implementing these POC techniques in clinical settings. [3].

Gum disease and OC are two examples of oral and dental disorders that pose a global health risk. In particular, early detection is crucial for cancer of the oral cavity. Combining model predictions causes complexity problems for traditional CNN-based methods. Our study suggests a feature-level fusion method based on EfficientNet models and a self-attention block. This method lowers complexity and increases accuracy. On the MOD dataset, our method's accuracy for detecting OC was 98.83% [4].

In a cloud setting, this study offers a 2-tier m-health care system for OC diagnosis. Preparing the data, segmenting the images using the Region Growing Technique, and extracting features are all steps in the process. ILDA, or Enhanced Linear Discriminant Analysis, chooses important characteristics for classification. An ensemble classifier (EC) is employed in two stages: SVM and MLP for initial classification, followed by an optimized CNN for the final decision. The CNN is fine-tuned using the hybrid Aquila Exploration Updated with Local Movement (AEULM) model, improving detection accuracy. The new Wildebeest Herd Optimization (WHO) combines the Aquila Optimizer (AO) with the hybrid method, and a comparative assessment validates the model's effectiveness [5].

Despite advances in diagnosis, oral cancer (OC) remains a prevalent malignancy that affects different parts of the oral cavity and has a considerable morbidity and fatality rate. Aggressive therapy may not be necessary if early discovery is made. Using a dataset of oral photos from SMS hospital, this study used ML algorithms, such as SVM, XGBoost, and RF, to classify stages of OC. To feed into the models, characteristics such as colour, texture, and shape were retrieved. With an accuracy of 87.1%, Random Forest beat SVM and XGBoost among the models [6].

The publication offers a comprehensive manual for using DL approaches to analyse digital pathology images, together with useful examples showing how to use them in actual situations. It tackles the difficulties associated with digital pathology, including the requirement for objective and trustworthy diagnostic instruments as well as accurate and effective handling of sizable, high-resolution picture collections. The use of DL techniques, such as CNNs and autoencoders, to tasks including feature extraction, picture segmentation, and classification is examined in this study. It also provides a thorough overview of the creation and improvement of DL models, including insights into maximising model performance and lowering the danger of overfitting [7].

This paper presents a two-stage tool for OSCC diagnosis through nucleus detection and segmentation in oral tissue histology images. First, 81×81 patches were downsampled to 21×21 and used to train a CNN, achieving 82.03% precision and 88.87% recall in nucleus detection. In the second stage, detected nuclei were segmented using the Chan-Vese model, resulting in a 97.56% precision, 89.38% Jaccard index, 94.22% Dice coefficient, and 91.58% recall. This approach is the first known method for combined nucleus detection and segmentation in OSCC diagnosis [8].

III. PROPOSED APPROACH

Various ML and DL Approaches have been implemented and extensively used to perform OC detection with the use of histopathology images. Our proposed model provides the implementation TL models with ensembling to enhance the classification and accuracy.

A. *Pre-analysis Processing Of Images*

Preprocessing has been performed on the training and well as validation images where rescaling of the images has been done along with rotation of the images up to 40 degrees and zooming of images by up to 20% with horizontal flips. These data augmentations have been done in order to improve the diversity in the dataset.

B. *TL for Classification*

For classifications of the oral histopathological images into cancerous or non-cancerous we have chosen the approach of developing a TL model [9]. TL is a technique in which a DL model is not trained on a separate but related task after it has already been trained on a certain task. TL models are assigned certain weights. The weights which has been used for our model is ImageNet. When there is dataset imbalance TL models perform quite well. TL models have a upper hand because it has the knowledge it has gained from

979-8-3503-8386-7/24/$31.00 ©2024 IEEE

the first task. This helps in preventing the overfitting of the model. Our approach uses three different TL models mainly EfficientNet-B3, Resnet-50 and DenseNet-201.

The B3 version of EfficientNet, which was built based on EfficientNet-B0, has a higher network size than the B0 version. Another kind of TL model is called Resnet-50, which is made up of 48 convolutional layers, max and average pooling layer each one. DenseNet-201 architecture involves the implementation of 201 convolutional layers. With the provided dataset we have trained these TL models individually and compared their accuracy.

C. Proposed Architecture

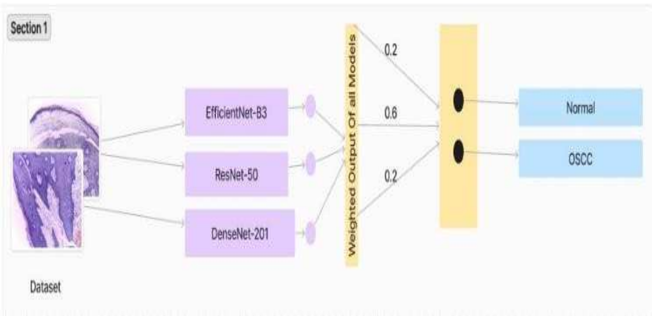


Figure 1: Proposed-Ensemble-model Architecture.

Figure 1 show ensembled model architecture utilizes an ensemble of pre-trained CNNs, EfficientNet-B3, ResNet-50, and DenseNet-201, for OC classification [10]. The ensemble model involves the use of weighted ensembling where in for Resnet-50, EfficientNet-B3 and DenseNet-201 the set weights are 0.6, 0.2 and 0.2 respectively. ResNet-50 has been set with higher weight as it has shown better classification compared to the other two TL models. Each model's output is weighted and combined to generate the final weighted ensemble model leveraging the strengths of each model for improved accuracy and reduced variance in classifying normal and OSCC cancer compared to the models without the ensemble feature. The model has been made such that the input image sizes are 224*224 with RGB channeled images so as to get consistent results.

IV. RESULTS

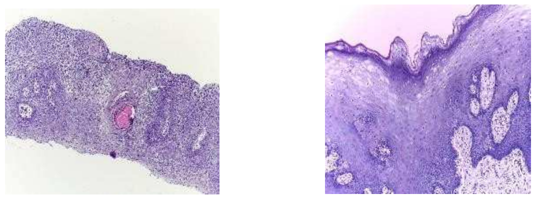
A. Experimental Setup

The hardware requirements entail utilizing an Intel i5 12th generation processor coupled with 16GB of RAM, ensuring sufficient computational power for data processing and model training. Additionally, a hard disk size of 512 GB provides ample storage capacity for datasets, model checkpoints, and software installations.

On the software front, the operating system employed is Windows 11, offering a stable environment for development and execution. The Integrated Development Environment (IDE) of choice is Visual Studio Code (VS Code), renowned for its versatility and robust features conducive to coding and debugging. Python version 3.11.1 serves as the programming language backbone, with essential ML libraries such as TensorFlow, Keras, NumPy, and Pandas facilitating data preprocessing and model development. Moreover, the Matplotlib library is utilized for visualization tasks, enabling insightful analysis and interpretation of results.

B. Dataset description

The dataset was obtained from the Kaggle database. There are 5072 pictures total in the collection, consisting of 2578 OSCC pictures and 2494 Normal pictures. It is partitioned into 3 classes: test, validation, and train. The percentages of each class are 15%, 15%, and 70%, respectively. The train dataset consists of 3550 images and the validation and test dataset consists of 761 and 761 images respectively. Figure 2 (a) represents OSCC image and 2 (b) represents Normal image.



(a) (b)

Figure 2: Histopathology Images.

C. Performance Metrics

Using the ensemble model and the three selected TL models, we have compared our results. To assess how well each model performs, we have compared accuracy, loss, recall, f1-score, and confusion matrix. The accuracy of the EfficientNet-B3 model [11] on the train and validation datasets can be inferred from Figure 3 and in table 1. Dropout and kernel regularizations are used to create the model at a learning rate of 0.001. On training datasets, it can attain an accuracy of 54.4%, while on validation datasets, it can reach 55.3%.

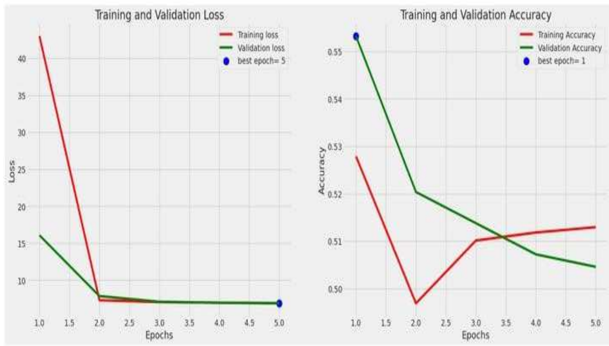


Figure 3: Displays Accuracy graph on left and loss graph on right for EfficientNet-B3.

Table 1: Performance Metrics of EfficientNet-B3.

	Precision	Recall	f1-score	support
Normal	0.53	0.81	0.64	370
OSCC	0.64	0.33	0.44	391

For Resnet-50 model we have used Adamax as optimizer and rate growth has been set at 0.001 and the model is able to achieve and accuracy of 98.2% and validation accuracy of 91.4% which can be inferred from Figure 4 and in table 2. For DenseNet-201 model [12] we have used Relu has activation function and Adamax as optimizer and rate growth has been set to 0.001 through which the model is able to

979-8-3503-8386-7/24/$31.00 ©2024 IEEE

provide an training accuracy of 57.88% on training dataset and 57.55 on validation dataset [13] which can be inferred from Figure 5 and in table 3.

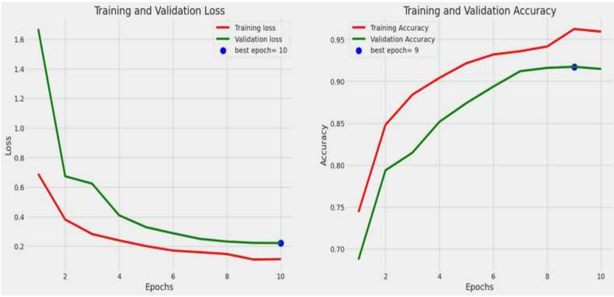


Figure 4: Displays Accuracy graph on left and loss graph on right for Resnet-50.

Table 2: Performance Metrics of Resnet-50.

	Precision	Recall	f1-score	support
Normal	0.88	0.94	0.91	370
OSCC	0.94	0.88	0.91	391

Table 3: Performance Metrics of Densenet-201.

	Precision	Recall	f1-score	support
Normal	0.94	0.18	0.30	370
OSCC	0.56	0.99	0.72	391

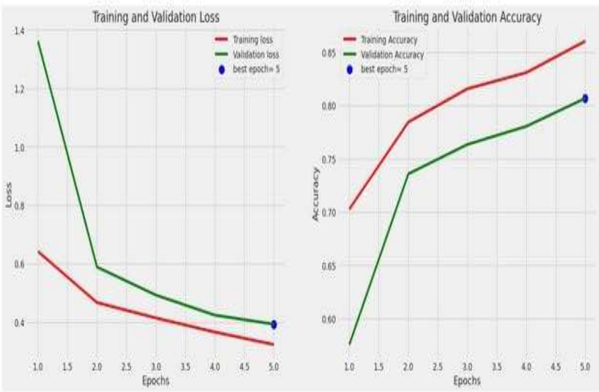


Figure 5: Displays Accuracy graph on left and loss graph on right for Densenet-201.

We have performed ensembling of all the three models with weighted ensembling as the approach to prioritize higher weights for the models which are performing well. Resnet-50 has been set with a weight of 0.6 and EfficientNet-B3 [14] and Densenet-201 has been set with weights of 0.2 and 0.2 respectively. From Figure 6 and in table 4 it is quite evident that the ensembling model has performed quite well as the we have achieved an training accuracy of around 99.09% on the training dataset and 93.16% on the validation dataset which is a huge improvement compared to all the individual model performances [15]. Figure 7 shows that the model is able to classify most of the images perfectly into Normal or OSCC images.

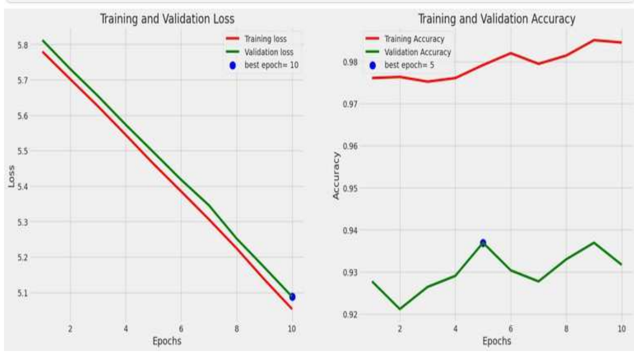


Figure 6: Displays Accuracy graph on left and loss graph on right for Ensemble Model.

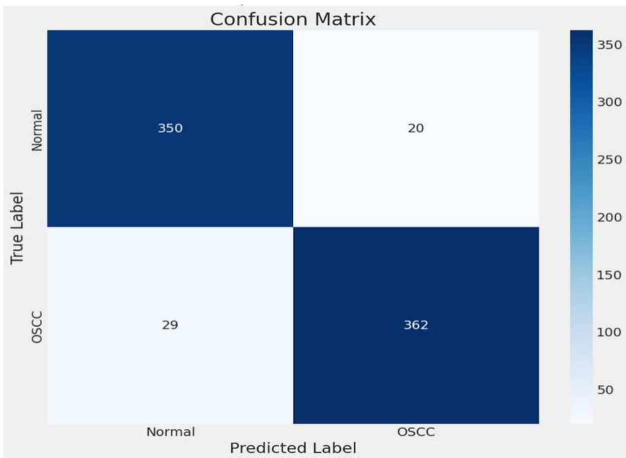


Figure 7: Confusion Matrix of Ensembled Model.

Table 4: Performance Metrics of Ensembled Model.

	Precision	Recall	f1-score	support
Normal	0.92	0.95	0.93	370
OSCC	0.95	0.93	0.94	391

V. Conclusion

Resolving the problem of differentiating between benign and malignant histology images is the main objective of the research. While TL models do a good job of correctly identifying the photos, our work suggests using weighted ensemble learning to increase the model's accuracy to 93.16%—a greater percentage than what individual TL models could achieve at 90%. These findings support the notion that DL is a highly effective method or instrument for doctors.

References

[1] [1] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with Deep Neural Networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017.

[2] Zijun Guo, Sha Ao, Bo Ao, "Few-shot learning based oral cancer diagnosis using a dual feature extractor prototypical network",

979-8-3503-8386-7/24/$31.00 ©2024 IEEE

Journal of Biomedical Informatics, Volume 150, 2024, 104584, ISSN 1532-0464, doi : 10.1016/j.jbi.2024.104584.

[3] Jasmeen Kaur, Rohit Srivastava, Vivek Borse, "Recent advances in point-of-care diagnostics for oral cancer", Biosensors and Bioelectronics, Volume 178, 2021, 112995, ISSN 0956-5663, doi : 10.1016/j.bios.2021.112995.

[4] Saif Ur Rehman Khan, Sohaib Asif, "Oral cancer detection using feature-level fusion and novel self-attention mechanisms", Biomedical Signal Processing and Control, Volume 95, Part A, 2024, 106437, ISSN 1746-8094, doi: 10.1016/j.bspc.2024.106437.

[5] Savita Shetty, Annapurna P. Patil, "Oral cancer detection model in distributed cloud environment via optimized ensemble technique", Biomedical Signal Processing and Control, Volume 81, 2023, 104311, ISSN 1746-8094, doi : 10.1016/j.bspc.2022.104311.

[6] Pinky Agarwal, Naman Gupta, Yashita Bharadwaj, Anju Yadav, Pratishtha Mathur, "Oral Cancer Stage Classification Using Machine Learning", Procedia Computer Science, Volume 235, 2024, Pages 3174-3180, ISSN 1877-0509, doi : 10.1016/j.procs.2024.04.300.

[7] A. Janowczyk and A. Madabhushi, "Deep Learning for Digital Pathology Image Analysis: A comprehensive tutorial with selected use cases," Journal of Pathology Informatics, vol. 7, no. 1, p. 29, 2016.

[8] Dev Kumar Das, Subhranil Koley, Surajit Bose, Asok Kumar Maiti, Bhaskar Mitra, Gopeswar Mukherjee, Pranab Kumar Dutta, "Computer aided tool for automatic detection and delineation of nucleus from oral histopathology images for OSCC screening", Applied Soft Computing, Volume 83, 2019, 105642, ISSN 1568-4946, doi: 10.1016/j.asoc.2019.105642.

[9] W. Zhu, L. Xie, J. Han, and X. Guo, "The Application of Deep Learning in Cancer Prognosis Prediction," Cancers, vol. 12, no. 3, p. 603, Mar. 2020, doi: 10.3390/cancers12030603.

[10] Santisudha Panigrahi, Jayshankar Das, Tripti Swarnkar, "Capsule network based analysis of histopathological images of oral squamous cell carcinoma", Journal of King Saud University - Computer and Information Sciences, Volume 34, Issue 7, 2022, Pages 4546-4553, ISSN 1319-1578, doi : 10.1016/j.jksuci.2020.11.003.

[11] Yingci Liu, Elizabeth Bilodeau, Brian Pollack, Kayhan Batmanghelich, "Automated detection of premalignant oral lesions on whole slide images using convolutional neural networks", Oral Oncology, Volume 134, 2022, 106109, ISSN 1368-8375, doi: 10.1016/j.oraloncology.2022.106109.

[12] M. S. Abd Rahim, F. Yakub, M. Omar, R. Abd Ghani, I. Dhamanti and S. Sivakumar, "Prediction of Influenza A Cases in Tropical Climate Country using Deep Learning Model," 2023 IEEE 2nd National Biomedical Engineering Conference (NBEC), Melaka, Malaysia, 2023, pp. 188-193, doi: 10.1109/NBEC58134.2023.10352612.

[13] Soubraylu Sivakumar, S.S. Sridhar, Ratnavel Rajalakshmi, M. Pushpalatha, S. Shanmugan, G. Niranjana, "Intelligent and assisted medicine dispensing machine for elderly visual impaired people with deep neural network fingerprint authentication system". Internet of Things. 23, ISSN 2542-6605, (2023). Doi: 10.1016/j.iot.2023.100821.

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord "Format" pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.

[14] M. Aubreville et al., "Automatic Classification of Cancerous Tissue in Laser Endomicroscopy Images of the Oral Cavity using Deep Learning," Scientific Reports, vol. 7, no. 1, p. 11979, Sep. 2017, doi: https://doi.org/10.1038/s41598-017-12320-8.

[15] Sivakumar, S., Haritha, D., Rajalakshmi, R., Shanmugan, S., Nagaraj, J. (2023). Artificial Intelligence Based Transfer Learning Approach in Identifying and Detecting Covid-19 Virus from CT-Scan Images. In: Kumar, M., Gill, S.S., Samriya, J.K., Uhlig, S. (eds) 6G Enabled Fog Computing in IoT. Springer, Cham. https://doi.org/10.1007/978-3-031-30101-8_9.