COMMUNICATION SCIENCE ET TECHNOLOGIE
Vol23, N°2, pp14, 26
ENPO – Maurice AUDIN
ASJP/ ESSN 2773-3483

Oral Cancer Detection Using Deep Learning: Feature Extraction and Activation Function Analysis

A. Sarra Rouabhi[1*], B. Racha Ikram Hakiki[1] and C. Redouane Tlemsani[2,]
[1] Laboratoire de Codage et Sécurité d'Information (LACOSI) - Département d'Electronique - Faculté de Génie Electrique, Université des Sciences et de la Technologie d'Oran Mohamed Boudiaf, USTO-MB, BP 1505, 31000 EL M'naouer, Oran, Algeria.
[2]Département d'Informatique - Faculté des Mathématique et Informatique, Université des Sciences et de la Technologie d'Oran Mohamed Boudiaf, USTO-MB, BP 1505, 31000 EL M'naouer, Oran, Algeria.

ABSTRACT

The classification of oral diseases, particularly distinguishing between cancerous and non-cancerous lesions, is a critical task for early diagnosis and treatment planning. In this study, we propose a deep learning-based approach that utilizes Inception ResNet V2 for feature extraction and a simple dense neural network for classification. To optimize model performance, we experimented with different activation functions in the dense layers, specifically ReLU, Sigmoid, and Softmax. Our results demonstrate that this configuration significantly improves classification accuracy, achieving a test accuracy of 92.11%. The separation of feature extraction and classification enhances interpretability and allows for better optimization of network parameters. This study highlights the potential of deep learning in improving the accuracy and robustness of AI-assisted oral disease diagnosis.

Keywords: Classification of oral lesions, Feature extraction, Activation Function, Deep Learning.

Author Correspondence, e-mail: sarra.rouabhi@univ-usto.dz

https://www.asjp.cerist.dz/en/PresentationRevue/42

1. INTRODUCTION

The integration of artificial intelligence (AI) and deep learning methods in medical image analysis and classification [1] has revolutionized the field of computer-aided diagnosis. In particular, the recognition and classification of oral lesions as cancerous or non-cancerous represent a major challenge for early and accurate diagnosis [2]. Traditional medical analysis methods rely heavily on clinicians' expertise, which can lead to variability in interpretations and increased diagnostic time. The use of deep learning models allows for the automation and optimization of this classification by automatically extracting discriminative features from medical images [3]. These advances significantly improve the efficiency, accuracy, and reproducibility of diagnosing oral diseases.

State of the Art

Several studies have already explored the application of deep learning methods for classifying oral pathologies, particularly distinguishing between cancerous and non-cancerous lesions:

- Deep Learning for Oral Cancer Detection [4]: In this study, Shamim et al. used a CNN model based on ResNet50 to classify images of oral lesions. The model was trained on a clinical dataset and achieved an accuracy of 85% in differentiating oral cancer cases from benign cases.
- Hybrid CNN-SVM Model for Oral Cancer Classification [5]: In this approach, Fati et al. combined a convolutional neural network for feature extraction and a support vector machine (SVM) for final classification. The results showed an improvement in accuracy compared to purely CNN-based methods, reaching 88% accuracy.
- Attention-Based Deep Learning Model for Oral Cancer Diagnosis [6]: In this model, Song et al. based on an attention mechanism was developed to highlight the most relevant regions in medical images. This approach increased diagnostic accuracy to 90% by focusing on key lesion features.
- Transfer Learning with InceptionV3 for Oral Lesion Classification [7]: This study, Panigrahi et al. explored the use of transfer learning by leveraging a pre-trained InceptionV3 model fine-tuned on a dataset of oral images. The approach achieved an accuracy of 87%, demonstrating the effectiveness of transfer learning in the medical domain.

Importance of Our Contribution

Our work stands out by using Inception ResNet V2 for feature extraction and a dense neural network for classifying oral lesions as cancerous or non-cancerous. Unlike previous studies that directly use CNNs for classification, our approach separates feature extraction from classification, allowing for better interpretability of results and greater flexibility in model parameter tuning. Additionally, by varying the activation functions used in the dense neural network, we analyze their impact on model performance—an approach rarely explored in existing literature. This study thus makes a significant contribution by enhancing the accuracy and robustness of AI-assisted diagnosis for oral diseases.

Our paper is organized into 3 key sections: Section 1, Introduction. In Section 2, our approach is explained in detail, outlining specific steps such as preprocessing, Inception ResNet V2 to extract features and a neural network is employing to classify cancer and non-cancer oral images. Section 3 presents the results of our experiments, analyzing the performance to evaluate the accuracy of the proposal. Finally, Section 4 concludes our paper, summarizing the main results.

2. OUR METHODOLOGY

This section provides a comprehensive description of the operational framework of the Inception ResNet V2 network, integrated with a neural network designed for the classification of oral disease images into two categories: cancerous and non-cancerous. The study utilizes the ORAL CANCER dataset, which is structured into two distinct classes. The architecture combines two Inception ResNet V2 models for feature extraction with a dense, three-layer neural network for classification. To optimize performance, four experimental setups were conducted, each employing different activation functions. The objective was to identify the activation function combination that yields the highest classification accuracy.

2.1. Dataset

ORAL CANCER dataset [8] consists of 950 images, divided into two categories: 500 images of oral cancer and 450 images of non-cancerous oral conditions. Each image is carefully annotated to facilitate straightforward classification. The dataset is designed to support research and development in oral cancer detection using machine learning techniques. By providing a balanced mix of cancerous and non-cancerous samples, researchers can investigate novel methods to improve diagnostic precision. As a valuable resource for the healthcare sector, this dataset promotes advancements in early detection and intervention for oral cancer, contributing to the broader mission of enhancing oral health through innovative technology. Fig. 1 shows an example.

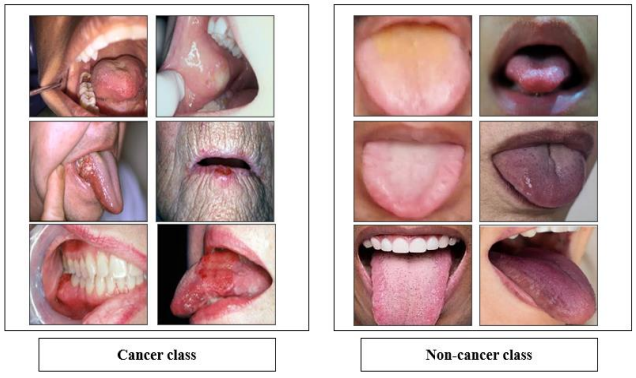


Fig. 1. ORAL CANCER dataset [8].

2.2. Feature Extraction using Inception ResNet V2

Imagine a neural network that combines the best of two groundbreaking architectural paradigms: Inception and Residual Networks. This is precisely what Inception ResNet V2 achieves—a sophisticated deep learning architecture that pushes the boundaries of image recognition [9]

Inception ResNet V2 is renowned for its impressive feature extraction capabilities. The network typically extracts:

Initial Layers: Approximately 64-192 initial features

Intermediate Layers: Expanding to 256-1024 features

Final Layers: Up to 1,536 high-level features

In its standard configuration, the network architecture can extract around 1,536 distinct features, allowing for extremely rich and nuanced representation of visual information [10].

At the heart of this network lies the Inception-ResNet block as in Fig.2, an ingenious design that orchestrates multiple convolutional pathways and residual connections. Picture a neural network where:

Parallel convolutional layers capture features at different scales (1×1, 3×3, 5×5)

A max pooling branch extracts multi-scale spatial information

Outputs are artfully concatenated to create rich, hierarchical feature representations

Residual Connections: Solving the Gradient Flow Puzzle

The genius of ResNet lies in its elegant solution to the vanishing gradient problem. By introducing a simple yet powerful concept [11]:

$$\text{Output} = F(\text{Input}) + \text{Input} \quad (1)$$

This skip connection allows gradients to flow more freely through the network, enabling the training of extraordinarily deep models.

Inception Modules: A Multi-Perspective Feature Extraction

Inception modules are like Swiss Army knives of feature extraction [12]. They simultaneously apply convolutions with different kernel sizes and pooling operations:

$$\text{Output} = \text{Concat}(\text{Conv1}\times1(\text{Input}), \text{Conv3}\times3(\text{Input}), \text{Conv5}\times5(\text{Input}), \text{MaxPool}(\text{Conv3}\times3(\text{Input}))) \quad (2)$$

This approach ensures that the network can capture features from multiple perspectives, enhancing its ability to understand complex visual patterns.

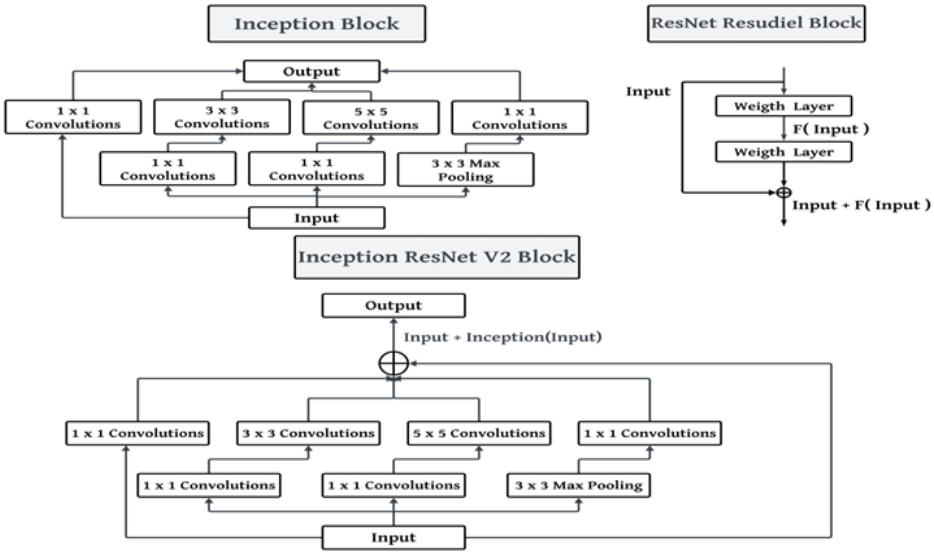


Fig. 2. Inception ResNet Architecture.

2.3. Classification

The extracted features were then processed by a simple neural network with three fully connected layers:

- A first dense layer (256 neurons) with ReLU or Sigmoid activation, followed by Batch Normalization (to stabilize training) and Dropout (0.3) (to prevent overfitting).
- A second dense layer (128 neurons) with Sigmoid activation, followed by Dropout (0.2).
- An output layer with Sigmoid or Softmax activation, depending on the test-ed configuration, to produce class probabilities (cancer/non-cancer).

We experimented with four activation function combinations:

- ReLU - ReLU - Sigmoid
- Sigmoid - ReLU - Softmax
- Sigmoid - Sigmoid - Sigmoid
- ReLU - Sigmoid - Softmax

Each function plays a specific role:

- ReLU (Rectified Linear Unit) [13]: Prevents the vanishing gradient problem and speeds up training by allowing only positive activations.
- Sigmoid [13]: Maps inputs to a (0,1) range, making it suitable for binary classification.
- Softmax [14]: Converts outputs into a probability distribution, ensuring they sum to 1, useful for multi-class classification.
- Batch Normalization: Normalizes activations to improve stability and train-ing speed.
- Dropout [15]: Randomly deactivates neurons during training to reduce over-fitting.

The model was trained using the categorical cross entropy loss function, an initial learning rate of 0.0001, and the Adamax [16] optimizer, ensuring stable convergence.

3. RESULTS AND DISCUSSION

This In this section, we presented the results obtained from our classification model, comparing four different executions while changing the activation functions each time. This comparative analysis was performed to select the optimal activation functions for our objective of classifying oral cancer images.

For our binary classification task of oral cancer detection, we implemented a dataset partitioning strategy of 80%/10%/10% for training, validation, and testing, respectively. This division was applied to our dataset comprising oral cavity images belonging to two distinct classes: cancerous and non-cancerous tissues.

To address the inherent challenge of limited medical imaging data, we employed data augmentation techniques on the training set. This process expanded our original training dataset from 945 to 1,620 images across both classes, enhancing the model's ability to generalize by introducing controlled variations in the training data.

For feature extraction, we utilized the Inception-ResNet-V2 architecture, a deep convolutional neural network pre-trained on large-scale image datasets. This transfer learning approach enabled us to extract 1,536 high-dimensional features from each image, capturing complex visual patterns relevant to oral cancer detection without requiring extensive training from scratch as shown in Fig.3.

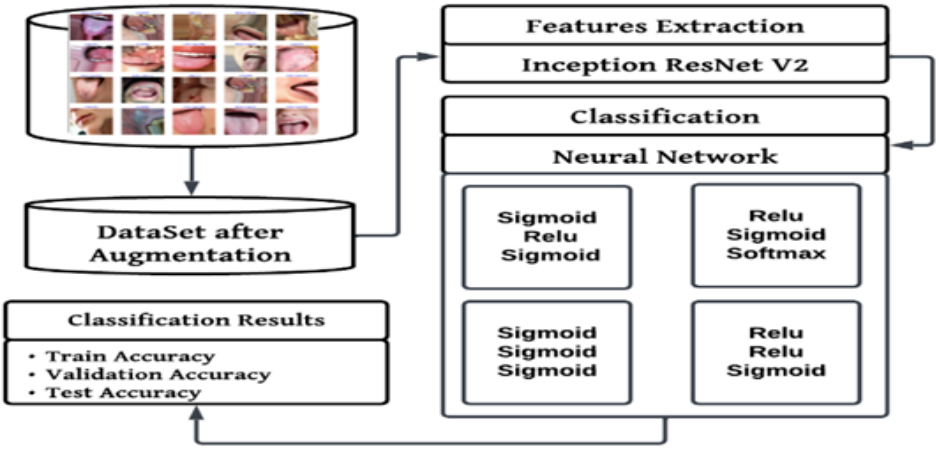


Fig.3. Our Approach.

3.1. Activation Function

Following feature extraction, we conducted a systematic investigation to determine the optimal combination of activation functions for our classification model. The experimental design involved multiple iterations of model training and evaluation, with each experiment employing different activation functions while maintaining consistent hyperparameters across all other aspects of the model architecture.

This methodical comparative analysis was essential for identifying which activation function configuration maximized the classification accuracy for our specific task. By isolating the activation function as the sole variable between experiments, we could directly attribute performance differences to this specific component of the neural network architecture, thereby enabling an evidence-based selection of the optimal configuration for oral cancer classification.

Table 1. The comparative analysis of four different activation function configurations

Activation Function	**Sigmoid / ReLU / Softmax**	**ReLU / Sigmoid / Softmax**	**Sigmoid/ Sigmoid / Sigmoid**	**ReLU / ReLU / Sigmoid**
Train accuracy (%)	99.83	99.87	99.97	99.88
Validation accuracy (%)	94.59	94.59	97.30	94.59
Test accuracy (%)	**92.11**	89.47	84.21	86.84

Table 1 presents the comparative analysis of four different activation function configurations and their impact on model performance across training, validation, and test datasets for oral cancer classification.

The experimental results demonstrate notable variations in model performance based on the activation function combinations employed. The Sigmoid/Sigmoid/Sigmoid configuration achieved the highest validation accuracy of 97.30%, suggesting superior generalization to unseen data during the development phase. However, this configuration yielded the lowest test accuracy (84.21%), indicating potential overfitting despite its strong validation performance.

Conversely, the Sigmoid/ReLU/Softmax configuration, while achieving an impressive training accuracy of 99.83%, demonstrated the most robust generalization capability with the highest test accuracy of 92.11%. This suggests that this particular combination of activation functions provided the optimal balance between fitting the training data and generalizing to

entirely new samples.

The ReLU/Sigmoid/Softmax configuration showed competitive training (99.87%) and validation (94.59%) accuracies but underperformed on the test set (89.47%). Similarly, the ReLU/ReLU/Sigmoid configuration achieved excellent training accuracy (99.88%) but failed to maintain comparable performance on the test dataset (86.84%).

These results highlight the significant impact of activation function selection on model generalization. The superior test performance of the Sigmoid/ReLU/Softmax configuration can be attributed to the complementary properties of these activation functions: Sigmoid constrains the output range while promoting smooth gradients, and ReLU introduces non-linearity while mitigating vanishing gradient problems. This combination appears particularly well-suited for the feature space created by the Inception ResNet V2 architecture for oral cancer image classification.

It is noteworthy that while all configurations achieved remarkably high training accuracies (>99.8%), their generalization capabilities varied substantially, with test accuracies ranging from 84.21% to 92.11%. This underscores the importance of thorough evaluation on independent test data rather than relying solely on training or validation metrics when developing medical image classification systems.

3.2. Training Dynamics and Convergence

The training plots in Fig.4 provide valuable insights into the behaviour of our optimal sigmoid/relu/sigmoid configuration model during the training process. The left graph illustrates the progressive reduction in both training and validation loss across 100 epochs. The convergence pattern shows a consistent decrease, with training loss (red line) and validation loss (green line) tracking closely together throughout most of the training process. This parallel reduction suggests that the model generalized well without significant overfitting. The best epoch (epoch 99) was reached near the end of training, indicating that the model continued to improve until the later stages of training.



Fig.4. Train and Validation Accuracy/Loss.

The right graph in Fig.4 displays accuracy metrics, revealing that training accuracy (red line) rapidly increased within the first few epochs, achieving and maintaining values near 100% after approximately epoch 20. The validation accuracy (green line) demonstrates more variation, fluctuating between 90% - 97% throughout the training process. The best validation performance was achieved at epoch 3, as indicated by the blue marker, suggesting that early epochs captured effective generalization patterns. The fluctuations in validation accuracy, despite stable training accuracy, indicate some degree of sensitivity to the validation subset composition.

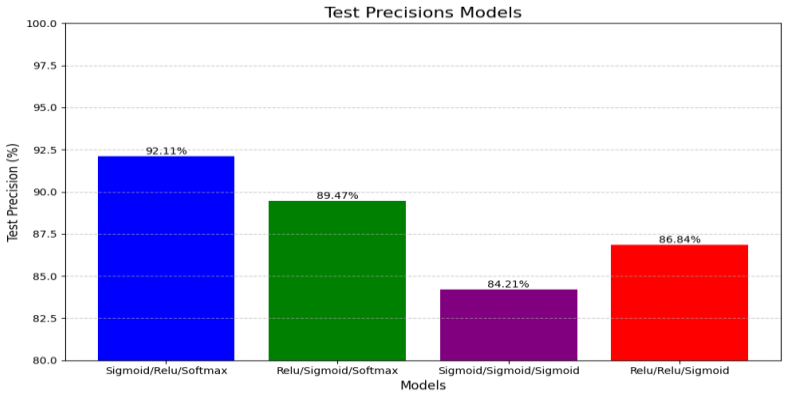


Fig.6. Test accuracy comparison.

The Sigmoid/ReLU/Softmax configuration is justified as the best choice because it achieves the highest test accuracy, 92.11%. Shows good generalization from training to testing. Provides an effective balance of feature normalization, Sigmoid, non-linearity, and gradient flow ReLU, and probability outputs Softmax.

Sigmoid in the input layer: Normalizes inputs between 0 and 1, which helps control feature scaling for the features extracted by Inception ResNet V2, then ReLU in the hidden layer: Provides faster training and helps mitigate vanishing gradient problems that might occur with sigmoid functions, and Softmax in the output layer: Ideal for the binary classification task, as it produces well-calibrated probabilities
This combination appears optimal for our binary classification task on oral cancer detection.

3.3. Confusion Matrix

The confusion matrix in Fig.6 reveals a critical aspect of our model's performance: while it achieves perfect precision (false positive), it demonstrates a slight weakness in recall due to the three cancer cases misclassified as non-cancer. This characteristic is particularly important

in the medical context of oral cancer detection, where false negatives (missing cancer cases) can have serious clinical consequences. However, the absence of false positive indicates that when the model predicts cancer, it does so with high reliability.

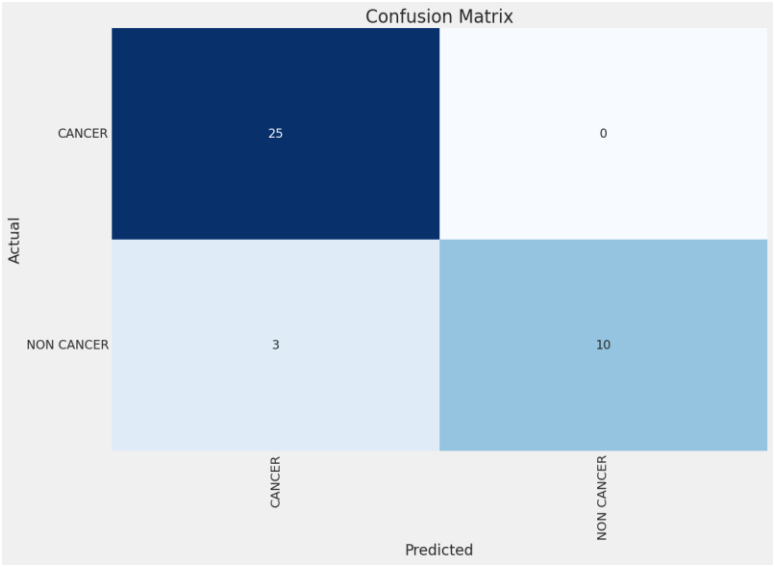


Fig. 6. Confusion Matrix.

3.4. Detailed Analysis of Classification Performance

The classification report provides a comprehensive evaluation of our model's performance on the test dataset, broken down by class. These metrics offer valuable insights into the model's diagnostic capabilities for oral cancer detection:

Table 2. Classification report

Performance	Precision	Recall	F1-Score
Cancer	0.8929	1.0000	0.9434
No Cancer	1.0000	0.7692	0.8696

Cancer Class Performance:

- Precision: 0.8929 - When our model predicts a case as cancerous, it is correct approximately 89.29% of the time. This indicates that 10.71% of cancer predictions are false positives.
- Recall: 1.0000 - The model correctly identifies 100% of all actual cancer cases in the dataset, suggesting no false negatives (missed cancer cases).
- F1-Score: 0.9434 - This harmonic mean of precision and recall provides a balanced assessment of the model's performance, with a high value indicating strong overall

performance for cancer detection.

Non-Cancer Class Performance :

- Precision: 1.0000 - When the model classifies a case as non-cancerous, it is correct 100% of the time, indicating no false negatives in this category.
- Recall: 0.7692 - The model correctly identifies approximately 76.92% of all actual non-cancer cases, with 23.08% being misclassified as cancer.
- F1-Score: 0.8696 - This represents balanced performance for the non-cancer class, slightly lower than for the cancer class.

These results have important clinical implications for the application of our model in oral cancer screening:

High Cancer Recall - The perfect recall for cancer detection is particularly valuable in clinical settings, as it minimizes the risk of missing cancer cases. This characteristic aligns with the medical principle that false negatives (missed diagnoses) should be minimized for potentially life-threatening conditions.

Trade-off in Non-Cancer Recall - The lower recall for non-cancer cases indicates that some patients without cancer may be flagged for further investigation. While this could lead to some unnecessary follow-up procedures, this is generally preferable to missing actual cancer cases in screening applications.

Balanced Precision - The high precision values for both classes demonstrate the model's reliability when making positive predictions, which helps maintain clinician confidence in the system's recommendations.

4. CONCLUSION

In this study, we developed an automated approach for oral cancer detection using deep learning. By applying transfer learning with the Inception ResNet V2 architecture and optimizing activation functions, we created a robust classification model for distinguishing between cancerous and non-cancerous oral tissues.

Our findings demonstrate that the Sigmoid/ReLU/Softmax configuration provides optimal performance, prioritizing cancer detection sensitivity while maintaining strong overall accuracy. This approach successfully addresses the clinical priority of minimizing missed cancer cases, making it suitable for deployment as a computer-aided diagnostic tool in screening environments.

Data augmentation proved effective in addressing the challenge of limited medical imaging data, while our systematic investigation of activation functions highlighted their critical importance in model performance. The results suggest that our methodology offers a viable solution for enhancing early detection capabilities in oral cancer screening programs.

Future work should focus on further improving the model's performance on non-cancer cases

while maintaining its excellent cancer detection capabilities, as well as validating these findings across more diverse datasets.

5. REFERENCES

[1] Deepa, S. N., & Devi, B. A. (2011). A survey on artificial intelligence approaches for medi-cal image classification. Indian Journal of Science and Technology, 4(11), 1583-1595.

[2] Patil, S., Albogami, S., Hosmani, J., Mujoo, S., Kamil, M. A., Mansour, M. A., ... & Ah-med, S. S. (2022). Artificial intelligence in the diagnosis of oral diseases: applications and pitfalls. Diagnostics, 12(5), 1029.

[3] Bisht, S. R., Mishra, P., Yadav, D., Rawal, R., & Mercado-Shekhar, K. P. (2021). Current and emerging techniques for oral cancer screening and diagnosis: a review. Progress in Biomedical Engineering, 3(4), 042003.

[4] Shamim, M. Z. M., Syed, S., Shiblee, M., Usman, M., Ali, S. J., Hussein, H. S., & Farrag, M. (2022). Automated detection of oral pre-cancerous tongue lesions using deep learning for early diagnosis of oral cavity cancer. The Computer Journal, 65(1), 91-104.

[5] Fati, S. M., Senan, E. M., & Javed, Y. (2022). Early diagnosis of oral squamous cell carci-noma based on histopathological images using deep and hybrid learning approaches. Dia-gnostics, 12(8), 1899.

[6] Song, B., Zhang, C., Sunny, S., Kc, D. R., Li, S., Gurushanth, K., ... & Liang, R. (2023). Interpretable and reliable oral cancer classifier with attention mechanism and expert knowledge embedding via attention map. Cancers, 15(5), 1421.

[7] Panigrahi, S., Nanda, B. S., Bhuyan, R., Kumar, K., Ghosh, S., & Swarnkar, T. (2023). Classifying histopathological images of oral squamous cell carcinoma using deep transfer learning. Heliyon, 9(3).

[8] Rashid, M. Z. (2024). Oral Cancer Dataset: Oral cancer and non-cancer [Data set]. Kaggle. Apache License, v2.0. https://www.kaggle.com/datasets/zaidpy/oral-cancer-dataset/data

[9] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017, February). Inception-v4, incep-tion-resnet and the impact of residual connections on learning. In Proceedings of the AAAI conference on artificial intelligence (Vol. 31, No. 1).

[10] McNeely-White, D., Beveridge, J. R., & Draper, B. A. (2020). Inception and ResNet fea-tures are (almost) equivalent. Cognitive Systems Research, 59, 312-318.

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[12] Shafiq, M., & Gu, Z. (2022). Deep residual learning for image recognition: A survey. Ap-plied sciences, 12(18), 8972.

[13] Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann ma-chines. In Proceedings of the 27th international conference on machine learning

(ICML-10) (pp. 807-814).
[14] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1, No. 2). Cambridge: MIT press.
[15] Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network train-ing by reducing internal covariate shift. In International conference on machine learning (pp. 448-456). pmlr.
[16] Llugsi, R., El Yacoubi, S., Fontaine, A., & Lupera, P. (2021, October). Comparison be-tween Adam, AdaMax and Adam W optimizers to implement a Weather Forecast based on Neural Networks for the Andean city of Quito. In 2021 IEEE Fifth Ecuador Technical Chapters Meeting (ETCM) (pp. 1-6). IEEE.