FGS Press

Journal of Computer Allied Intelligence
ISSN: 2584-2676
Vol.02(02), April 2024, pp.22-34

Research Article

Early Detection and Diagnosis of Oral Cancer Using Deep Neural Network

K. Vinay Kumar[1,*] , Sumanaswini Palakurthy [2], Sri Harsha Balijadaddanala[3], Sharmila Reddy Pappula[4] and Anil Kumar Lavudya[5]

[1-5] Department of CSE, Kakatiya Institute of Technology and Science, Warangal, Telangana,506001, India.

[*]Corresponding Author: K. Vinay Kumar. Email: kvk.cse@kitsw.ac.in

Received: 26/03/2024; Accepted: 22/04/2024.

Abstract: Oral squamous cell carcinoma (OSCC) is a malignancy that destroys the ability of the tissues surrounding the mouth to develop layers and membranes. Automated early diagnosis of oral histopathological images has allowed for the successful diagnosis of oral cancer, thanks to recent advancements in Deep Learning (DL) for biomedical image classification. By using a convolutional neural network (CNN) model based on deep learning for the initial analysis of oral squamous cell carcinoma (OSCC), this work aims to automate the classification of benign and malignant oral biopsy histopathological images. For this study, the CNN model Inception-Resnet-V2 is selected using the transfer learning approach. To enhance OSCC detection, additional layers are incorporated into this pre-trained model. By mining a repository of oral cancer histopathology images, we can gauge how well these tweaked models perform. We examine the modified structure of the pre-trained Inception-Resnet-V2 model and suggest a DL-CNN model that uses it. With an accuracy of 91.78%, it has outperformed in terms of performance metrics.

Keywords: Convolutional Neural Network (CNN); Oral squamous cell carcinoma (OSCC); oral cancer detection; Deep Learning (DL); Inception-Resnet-V2.

1 Introduction

Millions of individuals worldwide are impacted by the deadly illness known as oral cancer. Any part of the mouth, including the lips, cheeks, tongue, mouth floor, soft and hard palate, throat, and sinuses, can develop this type of cancer. The highest rate of mouth cancer in the world is known to affect women more frequently than men. An estimated 6,60,000 new cases of mouth cancer are reported each year, and an estimated 3,40,000 people die from the disease worldwide as a result of delayed diagnosis.[1][2] When it comes to oral cancer, the lips, mouth cavity, and pharynx all contain malignant tissues. As a result, the mouth region loses its developing layer structure and membranes. Even though it is common, oral cancer frequently remains undiagnosed until it has reached a more advanced stage.

This is due to the fact that initial symptoms are frequently painless and simple to ignore. Because of this, a lot of cases go undiagnosed until they have spread to other body parts, complicating and decreasing the efficacy of treatment.

One class of deep learning algorithms that excels at picture processing and recognition is convolutional neural networks (CNNs). A number of layers are comprised of it, including pooling, convolutional, and fully connected layers. Central to convolutional neural networks

(CNNs) are convolutional layers, which use filters to extract input image features such as edges, textures, and shapes. The third feature maps are subjected to down sampling after the pooling layers process the output from the convolutional layers. This reduces the spatial dimensions while retaining the most important data. The output of the pooling layers is subsequently processed by one or more fully connected layers to produce a prediction or image classification.

Motivation

By leveraging the capabilities of convolutional neural networks (CNNs), our work aims to establish a novel approach for the early diagnosis and detection of oral cancer. We envision a time when deep learning capabilities are applied to turn technology into a potent weapon in the fight against this horrible illness. Our work is primarily motivated by the possibility that it will dramatically alter the early detection of cancer, which could ultimately lead to outcomes that could save lives. [4] Our commitment to promoting technological developments underscores our belief in the positive knock-on effects that technology can have on the medical community as we investigate the intricate field of oral cancer. Moreover, our trajectory is not autonomous; rather, it is interwoven into the ongoing tapestry of medical investigation. A symbiotic relationship between our study and the broader field of research is required. We must continuously research, validate, and enhance our processes to ensure the long-term effectiveness and dependability of our plan.

Objectives

- To forecast the outcome for oral cancer detection by employing models that can be applied to past data.
- Preventing and early detection of oral cancer is the major goal.
- To support the medical professionals and radiologists diagnoses
- To ensure a high level of precision in the oral cancer prognosis.

2 Literature Survey

Recent advancements in artificial intelligence (AI) are increasingly making their way into the healthcare industry. Among these AI techniques, Convolutional Neural Networks (CNNs) have gained prominence, particularly for their exceptional accuracy in tasks like texture classification, making them well-suited for medical image analysis. Utilizing deep learning (DL) methods, various approaches have been developed and proposed for analysing medical data, including applications in breast cancer and lung cancer detection. These DL techniques have demonstrated enhanced accuracy, specificity, and sensitivity in medical image classification tasks. Additionally, the adoption of transfer learning has become widespread in medical image analysis, further improving the performance of DL models. New studies have shown that DL methods can effectively categorize oral lesions using a variety of medical image sources, such as histopathological data and live pictures of the mouth. The use of deep learning techniques (DLTs) to the analysis of histopathological images for the purpose of oral cancer detection has been the focus of numerous researchers. Our goal is to use deep learning to extract classification features from suspicious oral lesions in order to detect oral squamous cell carcinoma (OSCC) early from histopathological images. This will help in early detection of the disease.

For patients with oral squamous cell carcinoma (OSCC), Fujima [5] proposed using F-fluorodeoxyglucose PET images to predict how long it will be before they contract an infection. In order to evaluate h-index, metabolic tumor volume, and total lesion glycolysis, ResNet-101 was used to analyze FDG-PET images. The highest level of accuracy achieved was an 80%

accuracy rate, which was achieved through the application of DL classification.The DL-CNN model developed by Nandita et al. [6] combines the best features of the Resnet-50 and VGG-16 architectures. With a 96.20% success rate, this merged model was trained using a dataset that included enhanced pictures of mouth lesions. This accuracy surpassed that of other prominent DL-CNN models in classifying oral squamous cell carcinoma (OSCC).

Researchers used EfficientNet-B0 to perform binary classification of 716 real-time clinical images into potentially malignant or benign categories. They developed a streamlined deep learning convolutional neural network (DL-CNN) using transfer learning techniques [7]. The DL- CNN model proposed by the authors achieved an accuracy of 85.0%.With the use of cascaded deep learning (DL) techniques, Fu et al. [8] were able to differentiate between 44,409 photographic images of squamous cell carcinoma (SCC) confirmed by biopsy and normal clinical images. A sensitivity level of 94.90% and a specificity level of 88.70% were attained by the utilized deep learning technique.

After initially employing the transfer learning method, Das et al. [9] used deep learning (DL) to classify oral squamous cell carcinoma (OSCC) into its four distinct categories. Among the pre-trained models they used, ResNet-50 yielded the best classification accuracy (92.15 percent). Other models included VGG-16 and VGG-19. Then, they improved upon that with 97.50% classification accuracy by using a bespoke CNN model built on the VGG-19 architecture.The possibility of using deep learning techniques (DLT) for the detection of oral malignant disorders (OMD) was explored by Tanriver et al. [10]. In order to identify oral lesions and classify them as either benign, OMD, or carcinoma, they proposed a two-stage model. The dataset including pictures of mouth lesions was collected from the Oncology Institute's Tumor Pathology department at Istanbul University. When taking semantic segmentation into account, the EfficientNet-B7 model obtained the maximum accuracy of 92.90%, as pointed out by the authors.

Welikala et al. [11] utilized ResNet-101 and Fast R-CNN models to classify oral squamous cell carcinoma (OSCC) from images of the oral cavity annotated with bounding boxes. They reported an F1 score of 87.07% for identifying OSCC, showcasing the effectiveness of deep learning techniques in early oral cancer detection.By using six models trained using transfer learning to identify early-stage oral cancer from annotated images, the authors demonstrated the efficacy of deep learning techniques (DLT) [12]. After training VGG-19 with a 98.00% classification accuracy and ResNet50 with a 98.00% accuracy, they were able to differentiate between five distinct oral lesion types, with a focus on those that affected the tongue. These findings proved that the system could detect oral cancer in its early stages with an accuracy comparable to that of a human expert.

Figueroa et al. [13] incorporated the Grad-CAM technique from [14] for interpretability and opted for the GAIN [15] architecture instead of a standard DL-CNN for classification purposes. By sequentially integrating the GAIN classification and attention map, they achieved great success. With VGG-19 as their base CNN for training and GAIN as their output pipeline, the authors were able to achieve an estimated accuracy of 86.38%. Oral squamous cell carcinoma (OSCC) prognosis prediction using deep learning convolutional neural networks (DL-CNN) was the subject of a review by Alabi et al. [16]. They explored the application of deep learning techniques across diverse medical data types, including histopathological, clinic-pathological, Raman spectroscopy, gene expression, and CT images. The review highlighted the potential of emerging imaging modalities such as CT or enhanced CT, as well as spectral data, to yield substantial insights in OSCC prognosis.

In their review, Santisudha Panigrahi and Tripti Swarnkar [17] zeroed in on various deep

learning techniques (DLTs) that have been used to classify oral histopathology images. Additionally, they evaluated different methods for applying DL-CNN models to forecast the outcome of early-stage oral cancer.

Swetha and colleagues [18] introduced a neural network- based model for cancer detection, specifically targeting oral cancer. Employing a CNN architecture, the researchers developed a system capable of automatically identifying key indicators such as temperature, saliva pH value, and CO2 levels to aid in the diagnosis of oral cancer. A hybrid model specifically for oral cancer was presented by Rajaguru and Kumar Prabhakar [19]. Their method for classifying input sets of cancer images uses a combination of the Bayesian Linear Discriminant Analysis (LDA) and the artificial bee colony optimization algorithms. In the tested setting, the model's classification accuracy was 83.13% (out of a potential 100%).

Jiang and colleagues [20] introduced an innovative model for the detection of oral cancer cells utilizing fluorescent images. Their methodology revolves around leveraging fluorescent images as the primary input data, employing an image fusion algorithm to meticulously process the information. This approach aims to pinpoint and identify regions within the oral cavity that are particularly susceptible to cancer development. The integration of the image fusion algorithm is anticipated to enhance both the detection rate and overall efficiency of the system significantly. By intricately analyzing the fluorescent images, this model strives to provide a more thorough and precise means of identifying potential cancerous lesions within the oral cavity, thus offering a promising avenue for early detection and intervention in oral cancer cases.

3 Proposed System

A cutting-edge deep learning architecture called InceptionResNetV2 blends components from the ResNet and Inception models. It achieves superior performance in a variety of computer vision tasks, such as segmentation, object detection, and image classification, by utilizing the strengths of both architectures. InceptionResNetV2's complex architecture combines residual connections, multi- scale feature extraction, and effective computing resource management to enable deeper networks with higher accuracy and efficiency. This model's effectiveness in handling complex visual data has been demonstrated by its widespread use in research and real-world applications. Figure 1 shows Flow Chart of the Implementation Model

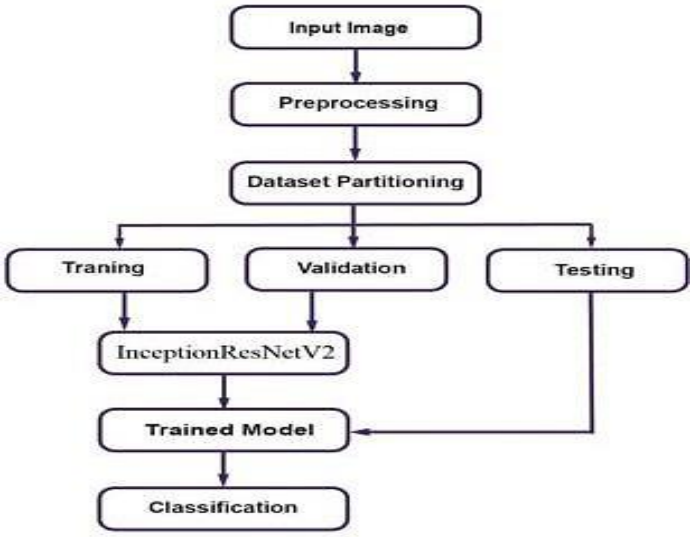

Figure 1: Flow Chart of the Implementation Model

3.1 Methodology

- A preliminary processing step will be performed on the gathered data.
- All algorithms are compared, and the algorithm with the highest accuracy is chosen.
- Building a prediction model using the gathered information.
- The data set will be used to apply the algorithm.
- Data analysis is carried out in accordance with implementation outcomes.
- Assess how accurate the forecasts were.

3.2 Dataset

Effective use of computational techniques, such as Deep learning techniques, is made to find the solution for OSCC detection. The dataset used determines how successful the detection is. The datasets used in this paper is prepared by combining accessible to the public and were released by Tabassum et al. There are 5685 oral histopathological images in the dataset; 3099 of them are cancerous, and the remaining 2586 are not. Biopsy slides are used to extract histopathological images, which are then analyzed using various cytological techniques under a microscope. Consequently, the dataset has passed clinical validation and can be used with various deep-learning models. We obtained the dataset from the website kaggle.com. To test and validate our hypotheses, we randomly split this data into two sets. Pictures of benign and malignant conditions were included in each dataset with a similar distribution of classes. Figure 2 shows Noncancerous image and Figure 3 shows Cancerous image

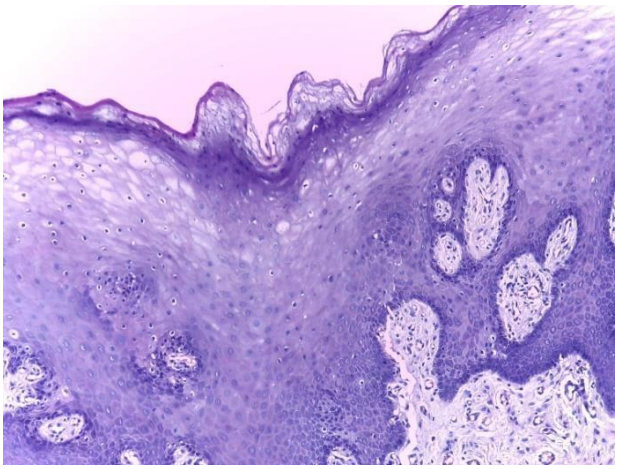

Figure 2: Noncancerous image

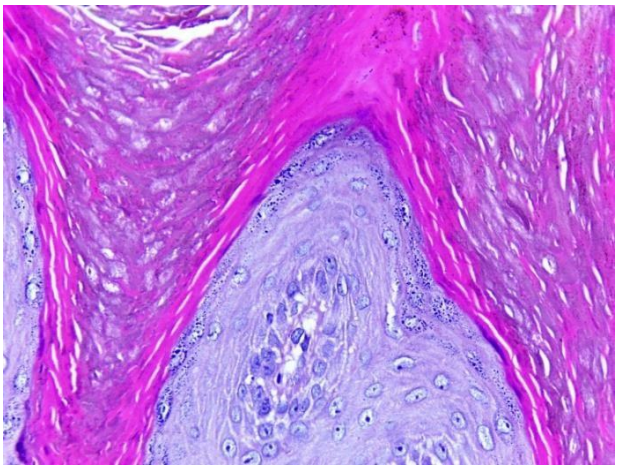

Figure 3: Cancerous image

3.3 Data Preprocessing

3.3.1 Image Scaling

Python is used to enlarge the dataset to 224 by 224 pixels. This will significantly cut down on processing time, but the accuracy of the model will suffer.

3.3.2 Data augmentation

The picture data generator function from the Python Keras package was utilized to increase the diversity of the training set and avoid over-fitting. Reducing the variance in pixel values would enhance the computer's performance. Pixel values are limited to the interval [0,1] by default due to the parameter value (1./255). The images were rotated to face a 25-degree target orientation. The width shift range transformation enables arbitrary rotation of the image to the left and right with a width shift value of 0.1. Training images were shifted by 0.1 in a vertical direction to the top or bottom.

3.3.3 Dataset splitting

A careful dataset splitting strategy is essential when preparing data for the Inception-ResNet V2 model, which is used to detect oral cancer. The training set, validation set, and test set are the three separate subsets of the dataset that are created during this process. Model training is built upon the training set, which makes up the majority of the dataset. To enable effective learning across conditions, it must include a wide range of oral images, such as samples of oral cancer lesions, precancerous lesions, and healthy oral tissue. The validation set is then used to monitor training results and adjust model hyperparameters, reducing the possibility of overfitting by offering a separate dataset for assessment. Finally, the test set—which consists of entirely unknown data—is essential for evaluating the model's ultimate performance in real-world situations. To prevent bias and preserve class balance, data splitting must be random, especially if stratified sampling is being used if the dataset is unbalanced. The Inception-ResNet V2 model for oral cancer detection is robust, reliable, and generalizable thanks to this careful dataset splitting technique, which also makes it easier for the model to be eventually used in clinical practice for early diagnosis and intervention.

3.4 Pre-trained Model

A synopsis of InceptionResNetV2 a deep convolutional neural network (CNN) is used in Google Research's InceptionResNetV2 architecture. It takes the best concepts from the ResNet and Inception designs and expands upon them to create a structure that is even more powerful. What distinguishes InceptionResNetV2 are its inception modules, residual connections, and network depth. Let's examine the fundamental components of the structure in more detail:

3.4.1 Inception Modules

InceptionResNetV2 is made up of a group of elements referred to as "Inception modules." To capture details across a range of scales, the model is constructed using parallel convolutional branches with different filter sizes. These nodes employ convolutions of 1×1, 3×3, and 5×5 sizes in addition to max pooling operations. By merging the results of these forks, the model is able to capture details at both the regional and global levels effectively.

3.4.2 Residual Connections

Using ResNet-style residual connections is crucial for training deep neural networks and fixing the vanishing gradient problem. BeginningResNetV2 enhances gradient flow during backpropagation by introducing residual connections between inception modules. These correlations facilitate the training of more intricate networks and aid in the preservation of crucial data.

3.4.3 Stem Block

The main entry point into the network is the stem block. It consists of several convolutional and pooling layers that perform the initial feature extraction and down sampling

tasks. Through the reduction of spatial dimensions in the input, the stem block enables the model to better capture low-level information.

3.4.4 Blocks of Reduction

Inception Reduction blocks in ResNetV2 are purposefully constructed to reduce the spatial dimensions and increase the number of channels. These reduction blocks typically consist of convolutional layers, followed by strides and max pooling procedures. They enable the model to capture higher-level information while reducing its computational load.

3.4.5 Initialization of Auxiliary Classifiers

Auxiliary classifiers are incorporated into ResNetV2 at intermediate nodes to improve regularization and training. These auxiliary classifiers assist the model in learning more robust and informative representations by permitting gradients to propagate from different depths.

3.4.6 Global Average Pooling and Classification

A global average pooling layer transforms the feature maps into a vector of fixed length before classification. The final classification probabilities are generated by connecting this vector to a fully connected layer using softmax activation. By reducing the number of parameters and providing spatial invariance, global average pooling improves the model's efficiency.

3.4.7 Transfer Learning with InceptionResNetV2

Improving a previously-trained model using its weights on a different dataset is called transfer learning. The tool used for this task is InceptionResNetV2. By freezing the early layers in the process and training only the last few levels or extra classifier layers with the learned representations from pre-training, we can improve performance on the new task. When there is little to no labeled data available, this method performs exceptionally well.

Taking the best aspects of both the Inception and ResNet models, the InceptionResNetV2 architecture is state-of-the- art. Because of its deep structure, residual connections, and inception modules, it is very good at classifying pictures. When working on image classification projects, developers and researchers can save time and money by utilizing InceptionResNetV2 transfer learning. They can use pre- trained models to achieve superior outcomes thanks to it. By applying InceptionResNetV2 to a variety of computer vision tasks, including semantic segmentation, object detection, and medical image analysis, researchers established that transfer learning could improve performance.

The success of InceptionResNetV2's transfer learning for computer vision tasks has prompted the creation of similar systems using various pre-trained models. Figure 4 depicts the fundamental building blocks of the InceptionResNetV2 architecture.

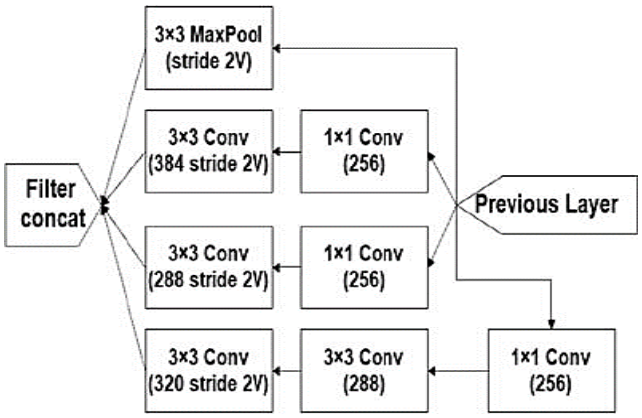

Figure 4: Architecture of InceptionResnetV2 model

4 Implementation

4.1Framework

Following dataset collection, preprocessing, and augmentation, the implementation framework for oral cancer detection using InceptionResNetV2 is divided into training, validation, and test sets. Transfer learning is used to refine the model, which is then trained, validated, and tested toevaluate performance using metrics like accuracy. Ongoing efficacy is ensured by deployment in clinical settings as well as ongoing evaluation and enhancement.

4.2 Objective

InceptionResNetV2 is being applied to oral cancer detection in an effort to achieve the following goals: high accuracy separation of cancerous lesions from normal tissue; early detection of precancerous abnormalities; generalization across a variety of patient populations; improved interpretability for clinical decision-making; scalability to process large volumes of images quickly; easy integration into clinical workflows; rigorous performance validation; and assurance of accessibility to healthcare facilities. When taken as a whole, these objectives aim to improve patient outcomes for oral cancer detection, enable timely intervention, and raise diagnostic accuracy.

4.3 Comparative Evaluation

We compared the accuracy of the model with many algorithms. The comparison is done among Inception v3, Xception, NASNet and InceptionResnet. Among all of them InceptionResnet has more accuracy. The details of other algorithms which were compared with InceptionResnet are below:

4.3.1 InceptionV3

Image classification and object recognition are two applications of InceptionV3, a robust convolutional neural network (CNN) architecture. Developed by Google research team, it is renowned for its intricate inception modules that allow the network to capture and process information at multiple scales.

With advanced features like batch normalization and factorized convolutions, InceptionV3 achieves impressive accuracy on various image datasets. Its architecture promotes efficient training and high-performance image recognition, making it a popular choice in the realm of computer vision. Better Feature Extraction: InceptionV3 makes use of the Inception module, which parallelly applies three different filter sizes (1x1, 3x3, and 5x5). This enables the model to effectively capture features at various scales. Rich feature representations are preserved while computational cost is decreased with the usage of 1x1 convolutions.

4.3.2 Xception

François Chollet, who developed the Keras framework for deep learning, also introduced Xception, an architecture for deep neural networks. Xception stands for "Extreme Inception." In 2017's "Xception: Deep Learning with Depthwise Separable Convolutions," the idea was first presented. One of Xception's selling points is its revolutionary use of convolutional layers—specifically, depthwise separable convolutions—to better capture spatial and channel-wise dependencies. A set of depthwise separable convolutions, comprising depthwise and pointwise convolutions, is one of Xception's defining characteristics. These allow for a substantial reduction in the parameter count when compared to conventional convolutional layers.

Efficiency and training speed are both enhanced by reducing the number of parameters. Xception is an impressive architecture in the field of deep learning, having figure 5 shown competitive performance on a range of computer vision tasks like object detection and image classification.

4.3.3 NasNet

Neural Architecture Search Network (NASNet) is an innovative deep learning architecture designed to automate the process of neural network architecture design. Developed by Google researchers, NASNet employs a neural architecture search algorithm to discover optimal network architectures for specific tasks. One distinctive feature of NASNet is its ability to to automatically search for efficient and effective architectures, leading to improved performance and reduced computational requirements. The network is known for its versatility and adaptability across various tasks, thanks to its ability to seamlessly combine elements from different architectures during the search process. NASNet has demonstrated state-of-the-art performance on image classification, object detection, and other computer vision tasks, showcasing its potential to revolutionize the way neural network architectures are designed and tailored to specific applications.

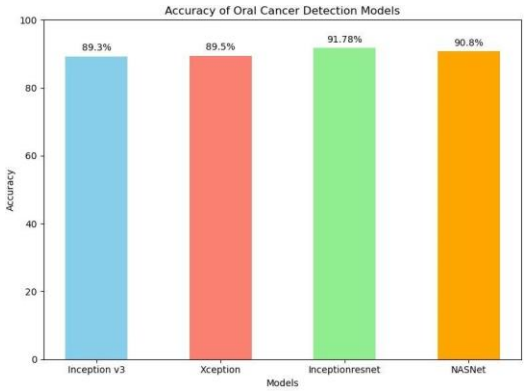

Figure 5: Comparision of Accuarcies of various pre-trained Models (InceptionV3, Xception, InceptionResnet and NASNet)

4.4 Algorithm

4.4.1 Step 1

Gather a variety of oral image datasets, apply uniform preprocessing, and divide them into test, validation, and training sets.

4.4.2 Step 2

Select the InceptionResNetV2 model architecture Initialize it using weights that have already been trained, and adjust it using the training set.

4.4.3 Step 3

With the hyperparameters adjusted, train the optimize model on the training set while keeping an eye on its performance on the validation set.

4.4.4 Step 4

Analyse the trained model using the test set, calculating the F1-score, recall, accuracy, and precision.

4.4.5 Step 5

Analyse the model's predictions for clinical application, then apply it to actual situations to detect oral cancer.

4.5 Graphical User Interface

Analyse the model's predictions for clinical application, then apply it to actual situations to detect oral cancer. Using InceptionResNetV2, a graphical user interface (GUI) for oral cancer detection entails building a user-friendly platform that facilitates efficient user interaction with the model. The first step in the process is creating the layout, making sure it is intuitive to use and includes necessary functions like uploading images and displaying results. Programming languages and frameworks such as Python with Tkinter, PyQt, or Flask are used in the implementation phase to create the user interface. It is imperative to integrate with the InceptionResNetV2 model so that users can upload oral images for analysis. Next, preprocessing features such as resizing and normalization are applied to the images in order to prepare them for inference. Furthermore, feedback systems and logging features are integrated to monitor user interactions and enhance model performance in the long run. In order to guarantee functionality and accuracy, deployment options include web-based interfaces and standalone applications that have undergone extensive testing and validation. For user help and troubleshooting, extensive documentation and support channels are offered. All things considered, the GUI simplifies the process of detecting oral cancer and makes the InceptionResNetV2model available to researchers and medical professionals for better diagnostic results. Figure 6 shows Web Interface [1](Loading an image) and Figure 7 shows Web Interface (Detecting the cancer)

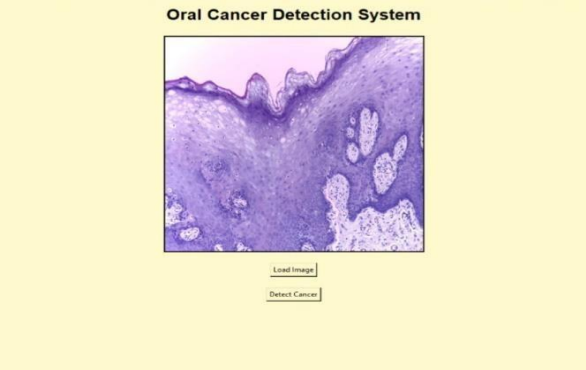

Figure 6: Web Interface [2](Loading an image)

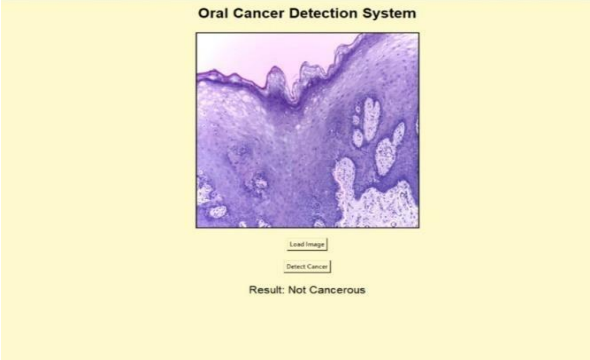

Figure 7: Web Interface (Detecting the cancer)

5 Experimental Evaluation

5.1 Validation

To ensure generalization to new data and avoid over fitting, keep an eye on the model's performance during training on the validation set. Analyse important performance indicators like recall, accuracy, precision, F1-score and ROC Curve to determine how well the model detects oral cancer.

5.2 Testing

To gauge the trained InceptionResNetV2 model's performance in real-world situations, analyse it on a different test set. Compute assessment metrics using the test set to assess the model's overall performance in identifying oral cancer lesions as well as its sensitivity, specificity, accuracy, and overall performance.

5.3 Comparison

Analyse the InceptionResNetV2 model's performance in relation to current protocols or baseline models for the identification of oral cancer. Examine variations in robustness, efficiency, and accuracy to determine the advantages and disadvantages of the suggested strategy.

5.4 Interpretation

To understand the behaviour of the model, interpret its predictions and decision-making procedure. To determine which areas of the oral images, contribute most to the model's classification decisions, visualize feature maps and activation patterns. Examine both false positives and false negatives to find areas that need work and to improve the architecture or training procedure of the model. Figure 8, 9 and 10 shows results.

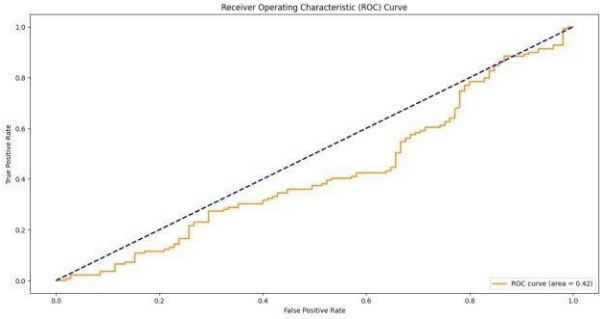

Figure 8: ROC Curve

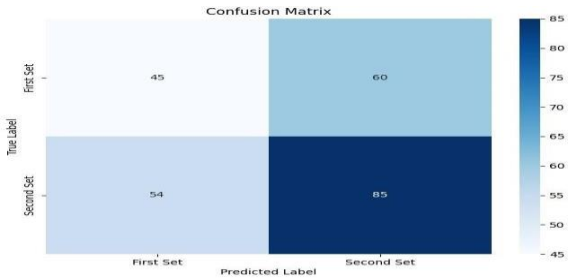

Figure 9: Confusion Matrix

Classification Report:				
	precision	recall	f1-score	support
First Set	0.45	0.43	0.44	105
Second Set	0.59	0.61	0.60	139
accuracy			0.53	244
macro avg	0.52	0.52	0.52	244
weighted avg	0.53	0.53	0.53	244

Figure 10: Classification Report

6 Conclusion & Future Work

In conclusion, the implementation of Inception-Resnet- V2 has shown promising results in oral cancer detection, achieving an accuracy of 91.25%. This success forms a robust foundation for further developments and real- world applications, with a suggested exploration of model compression techniques like quantization or pruning to enhance efficiency. Diversifying the training dataset is crucial for increasing the model's robustness, encompassing a broader spectrum of patient conditions and demographics for wider applicability. Additionally, the proposal to create web and mobile applications with streamlined user interfaces aims to improve accessibility and usability, making the model a valuable diagnostic toolfor healthcare providers and individuals.

As for future work, expanding the dataset to include various oral and dental disorders beyond oral cancer is recommended. This approach, coupled with multimodal data integration and extensive data augmentation, can enhance the diversity of datasets, ultimately improving diagnostic accuracy. Leveraging large-scale datasets for pretraining and transfer learning will expedite model training and boost generalization. Furthermore, validation in feedback-loop clinical settings is crucial for ensuring usability improvements and model relevance, fostering early diagnosis and detection for better patient outcomes. The combination of these efforts forms a comprehensive approach to oral and dental disorder detection, advancing healthcare technology and providing valuable tools for accurate and timely diagnoses.

Acknowledgement: Not Applicable.
Funding Statement: The author(s) received no specific funding for this study.
Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. F. Bray and J. Ferlay, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J Clin*, vol. 68, pp. 394–424, 2018.
2. R. D. Coletta, W. A. Yeudall and T. Salo, "Grand challenges in oral cancers," *Front Oral Health*, vol. 1, pp. 1–3, 2020.
3. A. Duggento and A. Conti, *"*Deep computational pathology in breast cancer," *Seminars in cancer biology*, vol. 72, pp. 226–237, 2020.
4. J. Gigliotti, S. Madathil and N. Makhoul, "Delays in oral cavity cancer," *Int J Oral Maxillofac Surg*, vol. 48, pp. 1131–1137, 2019.
5. N. Fujima and V. C. Andreu-Arasa, "Deep learning analysis using fdg-pet to predict treatment outcome in patients with oral cavity squamous cell carcinoma," *Eur Radiol*, vol. 30, pp. 6322–6330, 2020.
6. B. R. Nanditha and A. Geetha, "An ensemble deep neural network approach for oral cancer screening," *International Association of Online Engineering*, 2021.
7. F. Jubair and O. Al-karadsheh, "A novel lightweight deep convolutional neural network for early detection of oral cancer," *Oral Diseases*, vol. 28, pp. 1123–1130, 2021.
8. F. U. Qiuyun and Yehansen Chen., *"*A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: A retrospective study," *E-Clinical Medicine*, vol. 27, 2020.
9. Navarun Das and Elima Hussain, *et al*, "Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutionalneural network," *Neural Networks*, vol. 128, pp. 47–60, 2020.
10. G. Tanriver and Soluk Tekkesin, "Automated detection and classification of oral lesions using deep learning to detect oral potentially malignant disorders," *Cancers*, vol. 11, 2021.
11. Welikala and R. A. Remagnino, "Automated detection and classification of oral lesions using deep learning for early detection of oral cancer," *The Multidisciplinary Open Access Journal*, vol. 8, pp. 132677–132693, 2020.
12. Shamim and M.Z.M., "Automated detection of oral precancerous tongue lesions using deep learning for early diagnosis of oral cavity cancer," *The Computer Journal*, vol. 65, pp. 91–104, 2022.
13. Kevin Figueroa and Bofan Song, "Interpretable deep learning approach for oral cancer classification using guided attention inference network," *Journal of Biomedical Optics*, vol. 27, 2022.

14. R. Ramprasaath and Selvaraju, "Grad-cam: Visual explanations from deep networks via gradient-based localization," *IEEE International Conference on Computer Vision*, pp. 618–626, 2017.
15. Kunpeng Li and Ziyan Wu., "Tell me where to look: Guided attention inference network," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9215–9223, 2018.
16. R. O. Alabi and A. Almangush, "Deep machine learning fororal cancer: From precise diagnosis to precision medicine," *Frontiers in Oral Health*, vol. 2, 2022.
17. S. Panigrahi and T. Swarnkar, "Machine learning techniques used for the histopathological image analysis of oral cancer-a review," *Journal of Multimedia Information System*, vol. 13, pp. 106–118, 2020.
18. Sahanaz Praveen Ahmed and Lekshmy Jayan, "Oral squamous cell carcinoma under microscopic vision: A review of histological variants and its prognostic indicators," *SRM Journal of Research in Dental Sciences*, vol. 10, pp. 90–97, 2019.
19. R. Ramprasaath and Selvaraju, "Grad-cam: Visual explanations from deep networks via gradient-based localization," *IEEE International Conference on Computer Vision*, pp. 618–626, 2017.
20. Shipu Xu and Chang Liu, "An early diagnosis of oral cancer based on three-dimensional convolutional neural networks," *The Multidisciplinary Open Access Journal*, vol. 7, pp. 158603–158611, 2019.