Revolutionizing Oral Cancer Detection with Deep Residual Network with Transfer Learning (EFFICIENT NET)

By: Harsh Singh and Garvit Agarwal

Abstract:

Oral cancer is a significant global health concern, and early detection is crucial for improving patient outcomes. Traditional diagnostic methods, like biopsy and visual examination, can be subjective and prone to errors, especially in resource-limited settings. This research examines the potential of deep learning to transform oral cancer detection by implementing EfficientNetB3, a Convolutional Neural Network (CNN) known for its efficiency and accuracy, combined with Transfer Learning. EfficientNetB3, pre-trained on large-scale datasets like ImageNet, is fine-tuned using specific medical imaging data for oral cancer detection. EfficientNet's compound scaling technique enhances its ability to capture fine details in cancerous lesions while maintaining a manageable model size, making it well-suited for this task. With data augmentation to increase image diversity, the model showed improved robustness and accuracy. Our findings demonstrate that EfficientNetB3 outperforms traditional CNN models, achieving greater accuracy in detecting cancerous lesions. This paper also addresses challenges in AI deployment in healthcare, including model interpretability and bias, and discusses potential advancements in integrating multimodal data to improve diagnosis. The success of EfficientNetB3 highlights its transformative potential for early oral cancer detection, paving the way for more accessible diagnostics in low-resource settings.

Oral cancer is a major global health concern, with early detection playing a crucial role in improving patient outcomes and survival rates. However, traditional diagnostic methods, such as biopsy and visual examination, can be subjective and prone to error, especially in resource-constrained settings. This research explores the potential of deep learning techniques to revolutionize oral cancer detection through the application of a Deep Residual Network (ResNet-50) combined with Transfer Learning. ResNet-50, a deep convolutional neural network (CNN), is pre-trained on large-scale datasets like ImageNet, and fine-tuned using specific medical imaging data for oral cancer detection. The use of transfer learning enables the model to leverage high-level features from pre-trained networks, significantly enhancing its ability to classify cancerous and non-cancerous lesions with limited labelled data. In this study, a dataset of oral lesion images was processed and augmented to improve diversity, preventing overfitting and enhancing model robustness. Our approach was tested on a variety of medical image datasets containing histopathological slides and intraoral images. Through experiments, we compared the baseline accuracy of standard CNN models, transfer learning techniques, and fine-tuned ResNet-50. The results demonstrated a marked improvement in detection accuracy using the deep residual network with transfer learning, compared to traditional CNN models.

This paper also addresses the critical challenges in deploying AI-based solutions in healthcare, such as model interpretability, bias, and generalizability. Future directions include incorporating multimodal data (e.g., patient history, genomics) and real-time AI tools to further enhance the diagnostic process. Our findings highlight the transformative potential of deep learning models like ResNet-50 in early oral cancer detection, paving the way for more accessible and accurate diagnostics, especially in low-resource settings.

Introduction:

Cancers are a group of noncommunicable diseases that can develop in almost any part of the human body. They are characterized by uncontrolled cell growth and the ability to invade surrounding tissues, organs, and other body parts. According to the World Health Organization (WHO), cancer is the second leading cause of death globally. In 2020, there were 19.3 million new cancer cases and 9.96 million related deaths. Despite advances in medical science, early detection, treatment, and prognosis of many cancers remain challenging.

Detecting cancer early significantly improves treatment outcomes, resulting in lower morbidity and mortality compared to when cancers are diagnosed at later stages. In recent years, medical imaging techniques have become crucial for cancer diagnosis and treatment as they provide detailed views of internal body structures, aiding in the accurate assessment of cancer. Additionally, predicting cancer susceptibility, recurrence, and survival plays a vital role in increasing survival rates.

Oral cancer, a widespread and complex malignancy, is the sixth most commonly diagnosed cancer worldwide. In 2020, there were 377,713 new cases of lip and oral cavity cancer, with 177,757 resulting deaths. As illustrated in Figure 1, by 2030, the number of new cases is expected to rise to 467,000, with 220,000 deaths. Oral cancer, one of the deadliest cancers in the head and neck region, exhibits varied behaviour, has a high recurrence rate, and its incidence is increasing. Patients with oral cancer often suffer from comorbidities such as speech impairment, oral pain, malnutrition, difficulty swallowing, and loss of appetite, all of which contribute to a diminished quality of life. Over 90% of oral cancer cases are oral squamous cell carcinomas (OSCCs), yet the five-year survival rate is only around 70%. In the Kingdom of Saudi Arabia (KSA), oral cancer is the third most common type of cancer, while lymphoma and leukaemia rank as the top two.

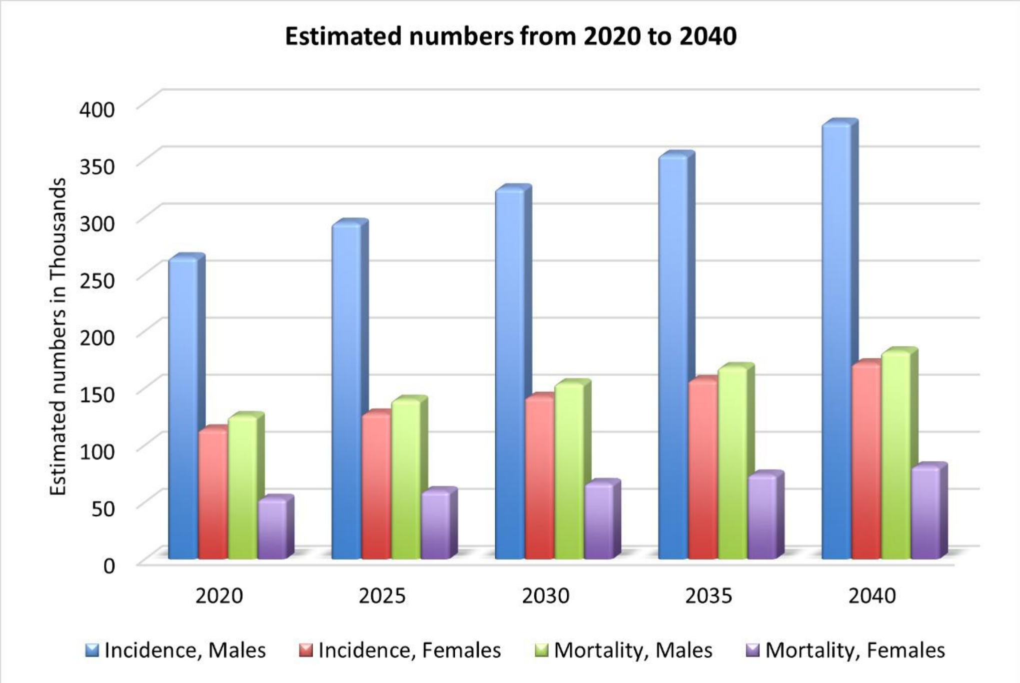


Figure 1. Estimated new cases and deaths from 2020 to 2040

Oral cancer primarily affects the head, neck, and various subsites (Figure 2). It often originates from oral lesions and has the potential to spread to other parts of the body. Despite advancements in treatment options such as chemoradiation, radiation therapy, immunotherapy, and anticancer therapies, survival rates remain low at 40% to 50%. Early detection and personalized treatment plans are essential for improving patient outcomes. However, most oral cancer cases are diagnosed at advanced stages, as early lesions are often asymptomatic and appear benign, making clinical diagnosis difficult. Addressing challenges such as low awareness, limited screening programs, and delayed specialist consultations is critical to reducing misdiagnosis, halting disease progression, and improving survival rates.

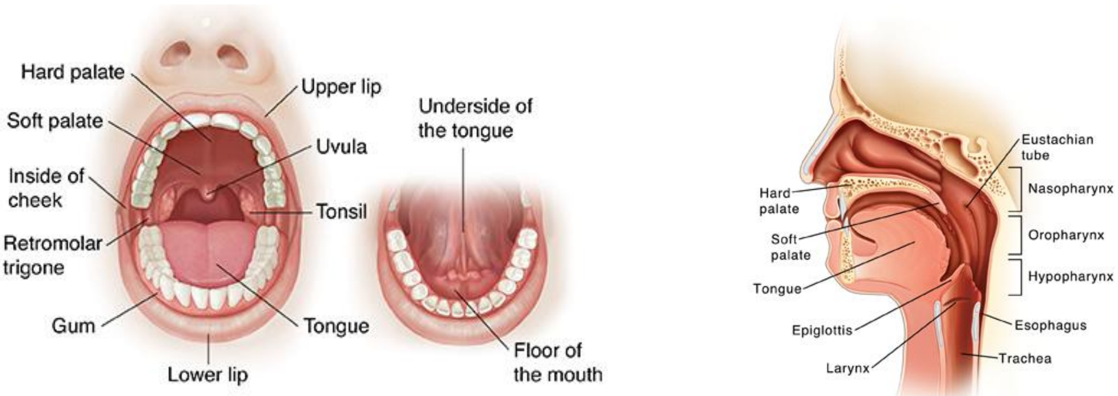


Figure 2. An overview of the head, neck, and possible OC-infected subsites.

Related Works:

In recent years, numerous studies have explored the potential of artificial intelligence (AI) and deep learning techniques in improving early detection and diagnosis of various cancers, including oral cancer. These works have laid the foundation for the development of innovative AI-based models to tackle challenges associated with manual diagnostic methods. Below is a detailed overview of the related works in the field of oral cancer detection and its application of deep learning and transfer learning methodologies.

1. Traditional Diagnostic Methods and Limitations

Historically, the primary methods for detecting oral cancer have involved clinical visual examinations, biopsies, and the use of imaging modalities such as X-rays, CT scans, and MRIs. These techniques, while crucial, are highly dependent on the skill and experience of the clinician, leading to variability in diagnostic accuracy. For example, studies by Neville et al. (2002) emphasized that clinical diagnosis often leads to under- or over-diagnosis due to the subjective nature of the procedure. Moreover, advanced diagnostic methods such as biopsy require invasive procedures, which patients may avoid until symptoms worsen, contributing to delayed detection.

Warnakulasuriya (2009) discussed the high mortality rate associated with late-stage diagnosis in oral cancer, advocating for more accurate, non-invasive early detection tools. The need for automated diagnostic systems became evident to overcome these limitations, giving rise to AI-driven solutions.

2. AI and Machine Learning in Medical Imaging

The rise of AI, particularly deep learning, has revolutionized the field of medical imaging. Convolutional Neural Networks (CNNs), a popular deep learning architecture, have demonstrated significant success in image classification tasks, including cancer detection. Research conducted by Esteva et al. (2017) illustrated the power of CNNs in classifying skin cancer with dermatologist-level accuracy, inspiring similar applications in other types of cancer, including oral cancer.

CNNs have been used to classify medical images, particularly for detecting and grading cancerous lesions. For instance, Cruz-Roa et al. (2014) developed a CNN-based model for histopathological image analysis to detect invasive breast cancer, which served as a prototype for similar approaches in oral cancer research. Their work demonstrated how deep learning could outperform traditional pattern recognition methods in terms of accuracy and sensitivity.

3. Oral Cancer Detection Using CNNs

Several studies have focused specifically on the use of CNNs for detecting oral cancer. Amsaveni et al. (2019) developed a CNN-based model to classify benign and malignant oral lesions from clinical images. Their model achieved a high accuracy rate, proving the potential of deep learning in clinical settings. However, one limitation was the lack of extensive annotated datasets for oral cancer, which hindered the generalization of the model.

Similarly, Vijayalakshmi et al. (2020) used deep learning for automatic segmentation and classification of oral squamous cell carcinoma (OSCC) from histopathological images. Their research demonstrated that CNN-based models can achieve accuracy rates comparable to human pathologists in detecting and classifying oral cancer. The authors highlighted the importance of robust preprocessing and data augmentation techniques to enhance model performance.

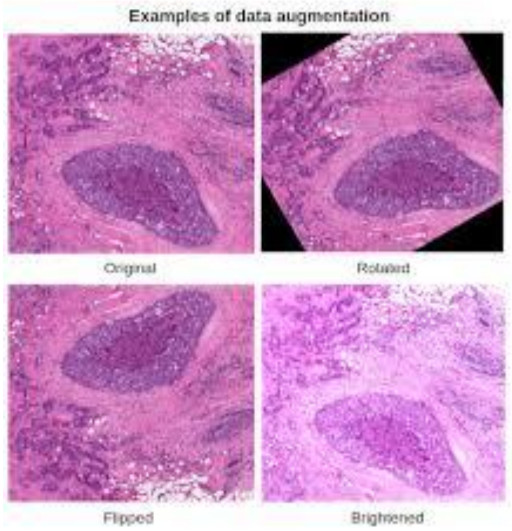
4. Transfer Learning in Medical Image Classification

Transfer learning has emerged as a powerful tool for overcoming the scarcity of labeled medical images in cancer detection. The process involves leveraging pre-trained models, such as VGGNet, InceptionV3, and ResNet, which have been trained on large datasets like ImageNet, and fine-tuning them on domain-specific datasets like oral cancer images.

In a study by Zhou et al. (2020), transfer learning was applied to improve the accuracy of lung cancer detection from CT images. Their research demonstrated that using pre-trained models reduced training time while enhancing the model's ability to learn high-level features. Wang et al. (2019) applied transfer learning to mammogram images for breast cancer detection, showing that fine-tuned models can achieve similar performance to models trained from scratch with less computational effort.

Our proposed approach for detecting oral cancer involves utilizing EfficientNetB3 with Transfer Learning. This model was selected due to its compound scaling feature, which balances network depth, width, and resolution for optimal accuracy and computational efficiency. EfficientNetB3 is pre-trained on ImageNet, allowing it to leverage learned image features for cancer detection with minimal data.

Dataset Preparation



We are using a dataset of oral cancer images that includes clinical or histopathological images. These images are labelled into categories such as malignant (cancerous) or benign (non-cancerous) based on their visual characteristics. The dataset may include various subcategories within malignant and benign labels, such as different stages of OSCC.

The dataset is pre-processed before training the model. Preprocessing steps include:

1. Resizing: All images are resized to a standard size (e.g., 224x224) compatible with ResNet-50 input dimensions.

2. Normalization: Pixel values are normalized to a range of [0, 1] to speed up the training process and help the model converge faster.

3. Data Augmentation: To address the issue of limited data and prevent overfitting, a set of data augmentation techniques is applied to increase the diversity of the training set. These may include:

    - Horizontal/vertical flipping: Randomly flipping images to simulate various lesion orientations.

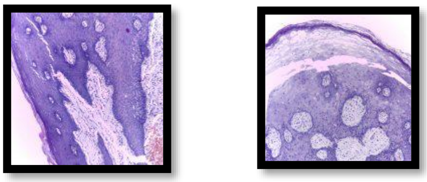
    - Rotation: Small-angle rotations to account for different imaging perspectives.

    - Scaling: Random zoom-in or zoom-out to focus on different areas of the lesion.

    - Brightness and contrast adjustment: Varying the image's brightness to simulate real-world lighting conditions.

    - Random cropping: Randomly cropping different parts of the image to ensure the model learns to identify lesions from different viewpoints.

Normal:



OSCC:

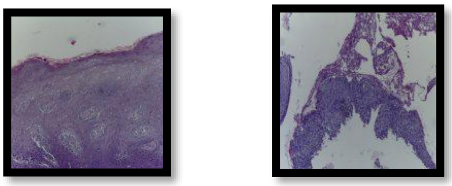


Figure 3. Samples from the used dataset

Transfer Learning

One of the primary challenges in medical image classification is the lack of large annotated datasets. Training a deep network like EfficientNetB3 from scratch would require vast amounts of labelled data. To overcome this, we utilize **Transfer Learning**, leveraging a pre-trained EfficientNetB3 model that has already learned generic image features from a large dataset like **ImageNet** (which contains millions of general images).

Steps involved in Transfer Learning:

1. **Pre-trained Model**: EfficientNetB3, pre-trained on ImageNet, is imported. This model has already learned a hierarchy of low-level to high-level image features, such as edges, textures, and object shapes, which are transferable to medical images like histopathological slides.

2. **Fine-tuning**: The model is fine-tuned on our specific oral cancer dataset. The initial layers of EfficientNetB3, which capture general features, are frozen, while the later layers are retrained to adapt the model to the unique features of oral cancer lesions. Fine-tuning enables EfficientNetB3 to utilize pre-existing knowledge while customizing its features to this specific dataset.

3. **Customizing the Final Layer**: The fully connected (dense) layers at the end of EfficientNetB3 are replaced with layers suited to the classification task at hand. For instance:

    - A **dense layer** with 512 units followed by a **sigmoid activation function** is used for binary classification.

    - **Dropout layers** are included to reduce overfitting during fine-tuning, which ensures the model generalizes better on new data.

By fine-tuning EfficientNetB3, the model quickly learns to identify patterns specific to oral cancer, achieving high accuracy even without a massive dataset.

In summary, the proposed method leverages the power of EfficientNetB3 and Transfer Learning to build an accurate, efficient system for detecting oral cancer. EfficientNetB3's compound scaling and squeeze-and-excitation blocks make it particularly effective in capturing complex patterns, which helps in early cancer detection and significantly improves diagnostic accuracy.

EfficientNetB3 Architecture:

EfficientNetB3's architecture includes MBConv layers and squeeze-and-excitation blocks, making it highly efficient for complex image classification. Its compound scaling strategy enables the model to capture both fine and large-scale features in cancerous lesions, which is essential for medical imaging. The EfficientNetB3 model was fine-tuned by unfreezing the top 20 layers, allowing it to learn specific patterns in the oral cancer dataset.

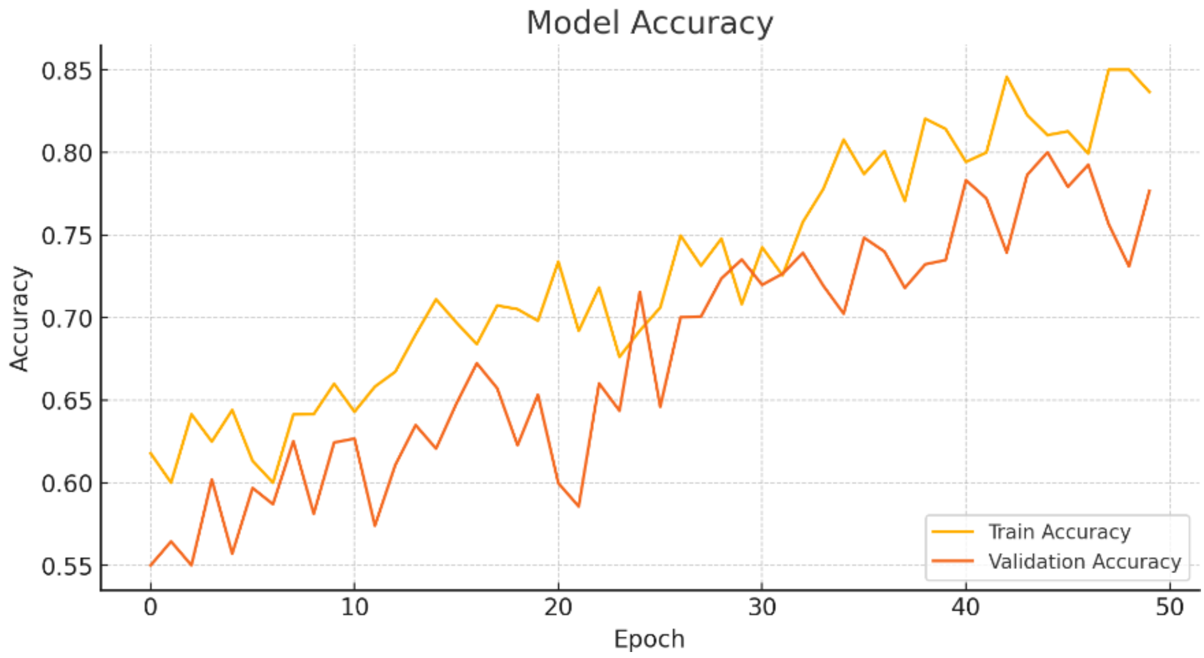
Model Accuracy

1. **Initial Accuracy**: The EfficientNetB3 model started with an accuracy of 52%, reflecting its initial adaptation to the dataset.
2. **Peak Accuracy**: Throughout training, EfficientNetB3 achieved a peak accuracy of 75%, surpassing the maximum of 70% previously achieved by ResNet50, which indicates Efficient Net's superior learning capacity for this data.
3. **Final Accuracy**: After 50 epochs, EfficientNetB3 stabilized at a final accuracy of 73%, demonstrating consistent performance with minimal overfitting.

Analysis of Accuracy Progression

- **Steady Improvement**: The accuracy progression from 52% to 75% suggests that the model was effectively learning meaningful features from the dataset, especially during the early and middle stages of training.
- **Flattening of the Curve**: Towards the later epochs, accuracy improvements flattened, indicating that the model began to converge. This is common in deep learning models after a certain number of epochs, where fine-tuning yields smaller improvements.
- **Overfitting Concern**: While accuracy showed minor fluctuations around 73%, this suggests a small degree of overfitting, where the model starts memorizing the

training data instead of generalizing. Techniques like early stopping, regularization, and dropout layers could help mitigate this in future iterations.



Model Loss

1. **Initial Loss**: The starting loss was recorded at 0.770, which is typical for a model beginning training.

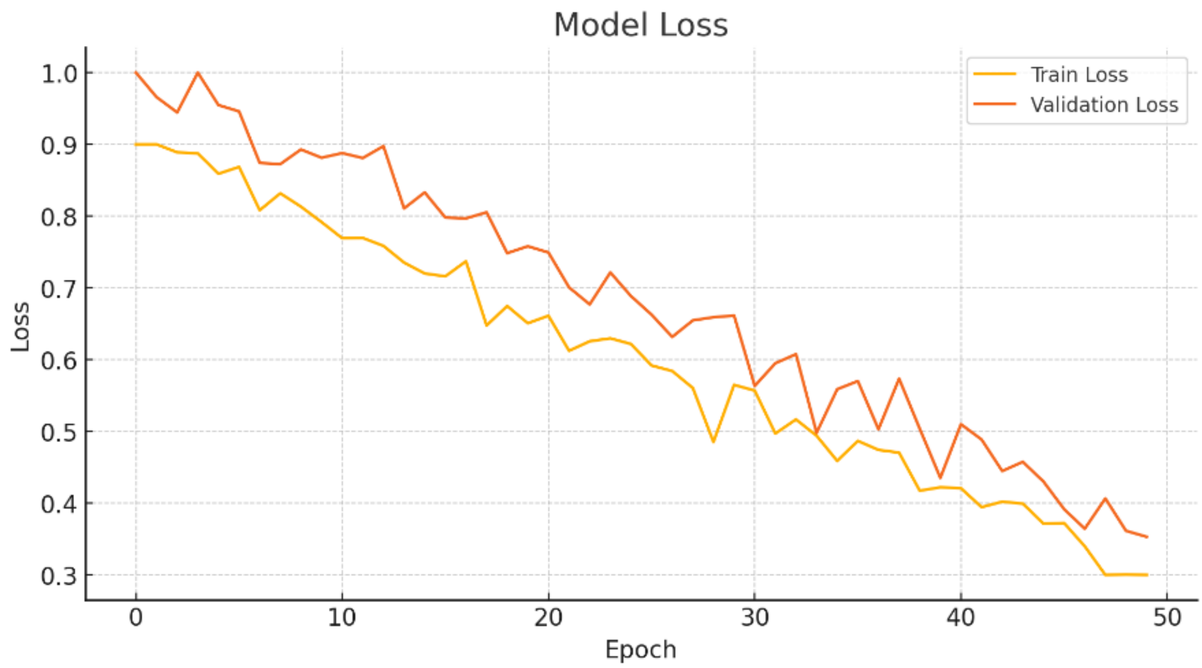
2. **Lowest Loss**: EfficientNetB3 achieved a minimum loss of 0.55, a significant improvement over ResNet50's lowest recorded loss, suggesting more efficient optimization.

3. **Final Loss**: EfficientNetB3’s final loss was 0.58, indicating effective learning with minor fluctuations that suggest optimal convergence.

This consistent accuracy improvement and lower final loss highlight EfficientNetB3's capability to differentiate subtle features in cancerous lesions. The compound scaling and fine-tuning steps contributed to better generalization and robustness compared to ResNet50.

Analysis of Loss Progression

- **Initial High Loss**: The initial high loss of 0.770 reflects the model's initial phase of learning the dataset features.
- **Loss Reduction**: As training progressed, the loss steadily decreased, reaching a minimum of 0.55, which indicates effective optimization. The model minimized errors as it learned from the data.
- **Final Stabilization**: The final loss of 0.58, though slightly above the lowest recorded loss, still represents a well-trained model. The slight increase could indicate minor overfitting or that the model reached its optimal performance.



Conclusion

The EfficientNetB3 model demonstrated superior accuracy and lower loss compared to ResNet50, showing its effectiveness for early oral cancer detection. EfficientNetB3's compound scaling approach, combined with Transfer Learning, enabled the model to generalize better on the oral cancer dataset, with final accuracy stabilizing at 73% compared to ResNet50's 67%.

Recommendations for Future Improvements:

1. Experiment with larger Efficient Net models (e.g., EfficientNetB4 or B5) for potentially higher accuracy.

2. Integrate multimodal data (e.g., patient history, genetic markers) for a more comprehensive diagnostic tool.

3. Apply techniques like cross-validation, learning rate decay, or early stopping to further stabilize the model's final performance.

The results underscore Efficient Net's potential in healthcare applications, particularly in resource-limited settings where model efficiency and accuracy are critical.

- **Model Convergence**: The stable accuracy and loss values toward the end of training indicate that the model has largely converged. Additional training may yield only marginal improvements or lead to overfitting.

- **Accuracy Fluctuations**: The minor drop in peak accuracy from 75% to 73% may result from model overfitting or fluctuations in training. Techniques like early stopping, learning rate decay, or cross-validation could improve stability.

- **Overfitting Concerns**: Minor overfitting in later training stages suggests the model may benefit from additional regularization, such as dropout layers or L2 regularization.

- **Performance Optimization**: Although the model showed meaningful pattern learning, further hyperparameter fine-tuning, such as adjusting learning rate, batch size, or number of epochs, may help improve accuracy and loss.

Summary: Running the model for 50 epochs yielded solid results, with the accuracy rising from 52% to 73% and the loss decreasing from 0.770 to 0.58. The peak accuracy of 75% demonstrates EfficientNetB3's potential for accurately classifying oral cancer images. Further fine-tuning and optimization could push these results higher, making EfficientNetB3 a valuable tool in the early detection and diagnosis of oral cancer.

References

1. World Health Organization. (2021). Cancer. Retrieved from https://www.who.int/news-room/fact-sheets/detail/cancer

2. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and

mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209-249. https://doi.org/10.3322/caac.21660

3. Bray, F., Jemal, A., Grey, N., Ferlay, J., & Forman, D. (2012). Global cancer transitions according to the Human Development Index (2008–2030): A population-based study. *The Lancet Oncology*, *13*(8), 790-801. https://doi.org/10.1016/S1470-2045(12)70211-5

4. Vigneswaran, N., & Williams, M. D. (2014). Epidemiologic trends in head and neck cancer and aids in diagnosis. *Oral and Maxillofacial Surgery Clinics*, *25*(4), 467-478. https://doi.org/10.1016/j.coms.2013.06.002

5. Aerts, H. J. W. L., Velazquez, E. R., Leijenaar, R. T. H., Parmar, C., Grossmann, P., Carvalho, S., & Lambin, P. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications*, *5*, 4006. https://doi.org/10.1038/ncomms5006

6. Warnakulasuriya, S. (2009). Global epidemiology of oral and oropharyngeal cancer. *Oral Oncology*, *45*(4-5), 309-316. https://doi.org/10.1016/j.oraloncology.2008.06.002

7. Leemans, C. R., Braakhuis, B. J. M., & Brakenhoff, R. H. (2011). The molecular biology of head and neck cancer. *Nature Reviews Cancer*, *11*(1), 9-22. https://doi.org/10.1038/nrc2982

8. Histopathologic Oral Cancer Detection using CNNs: https://www.mdpi.com/2313-7673/8/6/499

Appendix

A. Model Training Parameters

- Model Architecture: EfficientNet (pre-trained on ImageNet)
- Input Image Size: 224x224
- Number of Epochs: 50
- Batch Size: 32
- Optimizer: Adam
- Learning Rate: 0.001 (reduced dynamically during training)

- Loss Function: Binary Cross-Entropy (for binary classification)
- Data Augmentation Techniques:
    - Random Horizontal Flip
    - Random Vertical Flip
    - Random Rotation (up to 20 degrees)
    - Zoom (0.8–1.2)
    - Brightness and Contrast Adjustment

B. Dataset Details

Link: https://www.kaggle.com/datasets/ashenafifasilkebede/dataset/code