DENTAL AND ORAL DISEASE CLASSIFICATION

Y.M.
Student#: —
—

H.E.
Student#: —
—

J.L.
Student#: —
—

A.K.
Student#: —
—

ABSTRACT

This final project report presents a new deep-learning approach aimed at improving the classification of dental and oral diseases. It uses Convolutional Neural Networks (CNNs) and focuses on incorporating the Vision Transformer (ViT) for advanced feature extraction alongside an Artificial Neural Network (ANN) for accurate disease classification.

The project addresses the significant health challenges posed by oral diseases, which affect not just oral health but overall well-being, by aiming to enhance timely and precise diagnosis. By using images taken with standard cameras, the project seeks to make disease detection more accessible and less dependent on expensive medical equipment, making it more widely available.

The report details the project's methodology, starting with thorough data preprocessing to improve the dataset quality for model training. It explains the model's structure, which leverages the strengths of CNNs with the advanced capabilities of the ViT and ANN, to create a system capable of identifying various oral diseases accurately. Through careful training, the model reached a commendable training accuracy of 90.24% and proved its effectiveness with a test accuracy of 82.29%. This shows its potential to significantly improve diagnostic processes in dentistry by integrating modern deep-learning techniques.

Additionally, the report considers important ethical concerns, including the impact of false positives and negatives on patients. It emphasizes the project's dedication to ethical AI use by discussing how to ensure fairness and avoid bias, especially considering ethnic variations in oral features. This discussion highlights the project's comprehensive approach, focusing on technological innovation and AI's ethical use in healthcare.

—–Total Pages: 11

1 Introduction

Dental and oral diseases, encompassing a wide range of conditions from periodontal diseases to oral cancers, pose significant health challenges globally. These conditions not only affect oral health but also have systemic implications, impacting overall well-being. Despite advancements in dental care, the accurate and timely diagnosis of oral diseases remains a critical area for improvement. The motivation behind our project lies in addressing this gap by leveraging the power of deep learning, specifically Convolutional Neural Networks (CNNs), for the identification and classification of such diseases. Figure 1 below highlights a general overview of the data flow for the proposed model.

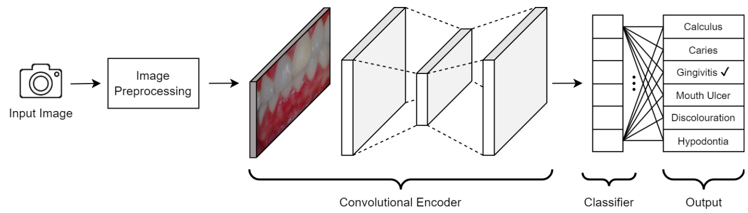

Figure 1: High-level pipeline for the proposed convolutional model.

The primary goal of our project is to develop an oral disease classification model that can accurately identify various oral conditions from input images. By doing so, a use case of this project would be to facilitate early detection, prompt intervention, and personalized treatment planning, ultimately improving patient outcomes and reducing the burden on healthcare systems.

This project is particularly intriguing due to its potential in the field of dentistry by integrating cutting-edge technology into routine clinical practice and streamlining workflows. By harnessing the capabilities of deep learning, we can enhance diagnostic accuracy and empower dental professionals with invaluable decision-support tools. Deep learning, with its ability to automatically learn hierarchical representations from raw data, is well-suited for image classification tasks. CNNs, in particular, have demonstrated remarkable success in various medical imaging applications, including dermatology, radiology, and ophthalmology. Their ability to extract meaningful features from images, coupled with their scalability and adaptability, makes them an ideal choice for a project in dental and oral disease classification.

1.1 Background & Related Work

While there are very few products or software that detect oral diseases, there is an increasing demand for automated medical assistance systems due to advancements in machine learning and with this, some notable work has been done in the field of classifying such diseases. Prajapati et al. (2017) use Radiovisiography (RVG) X-ray images and leverage a CNN model, having promising results even with a smaller dataset. Likewise, Imangaliyev et al. (2016) carry out similar research but looking at only cases of dental plaque and having Quantitative Light-induced Fluorescence (QLF) images instead. Almalki et al. (2022) employ the use of the object detector YOLOv3 on Orthopantomography X-ray images, which, unlike RVG and QLF images, concerns the entire oral cavity rather than a few teeth. Additionally, Lee et al. (2018) and Chen et al. (2021) both conduct a blend of these procedures with similar results as well, utilizing a CNN model with X-ray images. Though these studies show success in the detection and classification of dental diseases, they involve the utilization of expensive medical equipment as part of the data collection, while the aim of this model is to use images sourced from regular cameras. Paired with an application, ease and scalability are accounted for concerning the use cases for an average user.

2 Data Processing

Data preprocessing is a pivotal phase in our oral disease classification project, laying the groundwork for effective model training and evaluation. Each step of preprocessing was carefully designed to optimize the quality of the dataset, thereby ensuring robust and accurate model performance in later stages.

Data Cleaning and Organization: Our initial dataset comprised a combination of original and pre-augmented images. We removed the pre-augmented images to gain full control over the augmentation process. This step is crucial in avoiding biases that could arise from duplicated or artificially enhanced images. We also rigorously cleaned the dataset to ensure that each class was well-defined and devoid of mislabeling or irrelevant samples. Such data cleaning prevents the model from learning from inaccurate or noisy data, thereby enhancing its ability to generalize to new, unseen data.

Data Splitting and Leakage Prevention: To prepare the dataset for training and evaluation, we executed a careful split into training, validation, and test sets in an 80-10-10 ratio. This structure serves multiple purposes. Primarily, it ensures that our final model is tested on completely unseen data (the test set), which is vital for an unbiased evaluation of the model's performance. By reserving the test set exclusively for final evaluation and not using it at any point during the training or tuning process, we prevent data leakage. Data leakage can lead to overly optimistic performance estimates and is a common pitfall in machine learning projects. By maintaining this strict separation, we ensure the integrity and reliability of our model's performance metrics. For our final model testing, our test set will be used to evaluate the performance of the model as it is not possible to collect our own data for the final model. Thus, the test set will remain untouched until we arrive at the final phase of our model evaluation.

Addressing Class Imbalance: Class imbalance was a significant challenge, as is often the case in medical datasets. To address this, we implemented a strategic oversampling approach, as conducted by Ashraf (2023). Specifically, we replicated images in the underrepresented classes within the training set. We did this by looping through the under-represented classes and randomly selecting file images from the existing files. We ensured variability through this random selection process and later paired this method with data augmentation to prevent overfitting. This method effectively balances the class distribution, enabling the model to learn features from all classes equally and reducing the risk of bias towards more prevalent conditions.

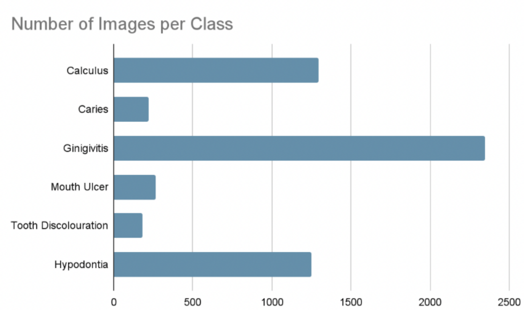

Figure 2: Class Data Distribution for Training Set.

Post-Baseline Augmentation: After establishing baseline models, we utilized Keras' ImageDataGenerator for further augmentation according to TensorFlow (2024). By generating synthetic variations of these classes, we aimed to enrich the dataset, further countering the class imbalance issue. Augmentation techniques included transformations like rotation, zoom, and horizontal flips, thereby enhancing the model's ability to generalize across varied inputs. Below in Figure 2 are the statistics displaying how many images are in each class of the train set to guide our augmentation process. Augmentation and oversampling were only performed on our training data to increase generalization performance.

Data Visualization for Insightful Understanding: Comprehensive data visualization was employed to gain insights into the dataset's structure and distribution. By quantifying and visualizing the distribution of data points across the training, validation, and test sets, we ensured that each set was representative of the overall dataset. Visualization of images from each class provided an intuitive understanding of the data, aiding in identifying any anomalies and understanding the diversity present in the dataset. Below, in Figure **??**, is an example of a cleaned training sample for the class of Tooth Discoloration.

Rigorous Data Normalization: An essential step in our preprocessing was normalizing the data based on the calculated mean and standard deviation of the training dataset, as mentioned by Kozodoi (2021). Unlike using preset normalization values from other datasets (e.g., ImageNet), calculating our statistics ensured that the normalization was specifically tailored to our dataset's characteristics. This approach is more accurate and beneficial, as it aligns the normalization process with the actual data distribution we aim to model. We computed these statistics post image-to-tensor conversion, ensuring that our model trains on data normalized to reflect the true nature of our specific dataset. This careful normalization aids in stabilizing and speeding up the training process and ensures consistency in the input data, both crucial for effective model training.

Our data preprocessing steps have set a solid foundation for the following stages of the project. These steps are pivotal in ensuring that the data is of the highest quality, suitable for training our neural networks, and enabling them to perform and generalize well in real-world scenarios.

3 Neural Network Model

This model will utilize Convolutional Neural Networks to detect dental and oral diseases to create a sophisticated deep-learning framework designed to excel in image-based tasks through rigorous training to learn important features and facilitate accurate disease classification. Figure 3 exhibits a more detailed pipeline of the model, elaborated on in the section below.

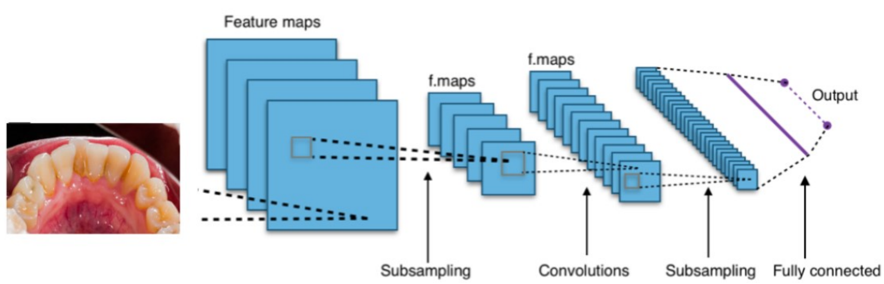

Figure 3: Architecture and mapping of CNN classification model.

3.1 Architecture

A convolutional neural network was used to assist in image classification in addition to previous models. The hierarchical core of the CNN model is made up of fully connected, pooling, and convolutional layers. Multiple hidden layers were used to improve access to oral disorders and guarantee that the model can achieve high accuracy. The network is designed to take input images with three channels (Red, Green, Blue), with a width and height of 256, and utilizes two convolutional layers, one max pooling layer, as well as two fully connected layers. The first convolutional layer uses a kernel size of 3, a pooling of 1, and a stride of 1. The second convolutional layer uses a kernel size of 7, a pooling of 1, and a stride of 1. The reason we chose a smaller kernel size on the first convolutional layer was to capture more intricate details when the model is doing colour and edge detection whereas we wanted to use a larger kernel size on the second layer for the detection of more high-level features like texture of the image. The convolutional layers implement ReLU activation functions to introduce non-linearity to the model, and the max pooling layer is responsible for reducing the spatial dimension of the feature maps. To help prevent overfitting, regularization techniques were deployed. Most notably, a dropout layer with a p-value of 0.25 is introduced to randomly select a fraction of input units. Figure 4 shows a visualization of the model's architecture highlighting the change in input channels.

Before the fully connected layers and after the convolutional layers, the model used batch normalization across mini-batches to help with adaptability and increase accuracy.

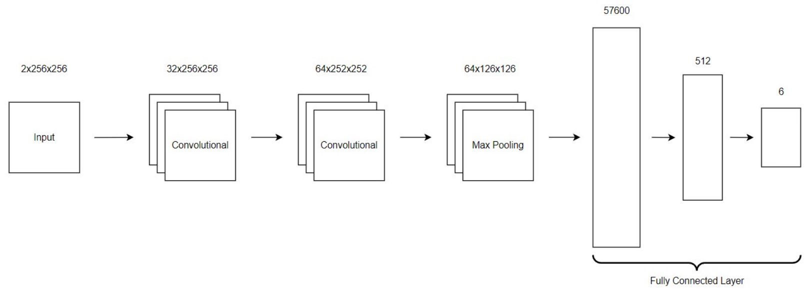

Figure 4: Pipeline of model with input channels and their dimensions.

3.1.1 EfficientNet

EfficientNet is a type of CNN pre-trained model that uses compound scaling for better performance. It seeks to improve both computing efficiency and performance. An Artificial Neural Network (ANN) model was used for image classification in this model after EfficientNet was used to extract features. The ANN model utilizes an input tensor size of 1280 pixels and constitutes 3 fully connected layers called fc1,fc2, and fc3. fc1 takes 1280 input features and 512 output features, followed by a ReLU activation function and a dropout rate with a dropout probability of 0.5. fc2 has 512 input features and 256 output features, followed by a ReLU activation function and a dropout rate of 0.4. Lastly, fc3 has 256 input features, 128 output features, a ReLU activation, and a dropout of 0.3, 0.4, and 0.5 respectively. An overview of the ANN model is shown in figure 5 below:

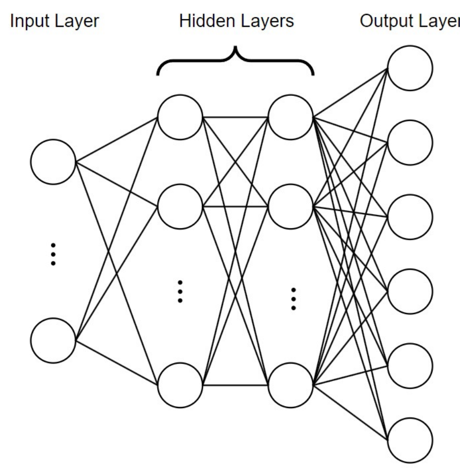

Figure 5: ANN classifier used with EfficientNet model and its layers.

3.1.2 ViT Classifier

Vision Transformer (ViT) is a pre-trained neural network architecture that applies transformers for computer vision tasks. Unlike convolutional neural networks, ViT models process images as sequences of patches. ViT networks can be used for feature extraction, and the features are represented as embeddings of the image patches. A ViT model splits the input image into small patches and flattens them into sequences of vectors through linear projection to obtain fixed-size embedding vectors. These vectors are processed by encoders to provide spatial information about the relative location of each patch. This is done through learnable position embeddings to add positional information about each embedding in the image patch, as shown by Takyar (2023). Figure 6 displays a representation of the architecture of a ViT classifier.

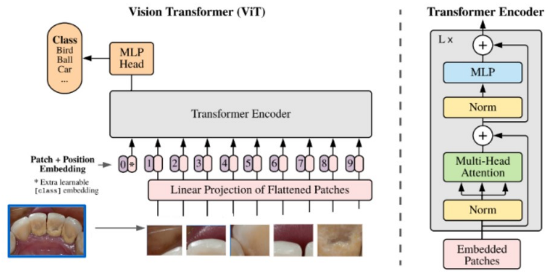

Figure 6: Architecture of ViT model and the transformer encoder.

Following the architecture of a ViT model, the model was implemented through the `ViTFeatureExtractor` class that defines the ViT model, and the `extract_features` method that utilizes the ViT model and performs feature extraction. The ViT network was chosen due to the model's self-attention mechanism to extract global information about images, as stated by Takyar (2023). The self-attention mechanism enables the model to attend to different regions of input data, based on the relevance to the given task. This mechanism also computes a weighted sum of the input data, where the weights are computed based on the similarity of the input features, allowing the model to emphasize relevant input features and capture a more informative representation of the input data, as affirmed by Boesch (2024).

3.2 Baseline Model

It is innately difficult to measure the effectiveness of the primary neural network solely based on test accuracy as it is unknown whether a high accuracy is due to the simplicity of the problem, or if poor accuracy is a result of the inherent difficulty of the problem. As such, it is imperative to compare the main network to at least one baseline model as an alternative that it must outperform to be deemed effective. The logistic regression and random forest models have been employed as the baseline models for the primary convolutional neural network.

According to Mishra (2021), logistic regression is a relatively simple model that computes a biased weighted sum with the Sigmoid function to form prediction results, as seen in Figure 7. This methodology is transferable and was applied to image classification and used as a baseline for the primary model due to its quick training and lack of need to scale features, as mentioned by Llourenc (2022).

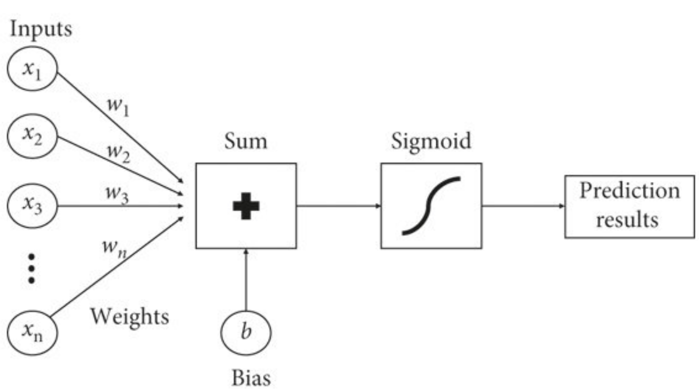

Figure 7: Abstract architecture and pipeline of a logistic regression model.

A random forest is more complex compared to logistic regression but is overall a basic tree-based ensemble model that combines predictions from multiple models to attain a better overall performance than each of the individual underlying models. Although there are a large number of voting rules that could be used, the baseline uses the majority rule, a simple and common method for classification. Via the modus operandi of ensemble learning, random forests provide accurate low-cost classified images using a baseline of training data, which implies that with higher data in the training set for classification, the higher the accuracy of the end product, as expressed by Sahota (2022). Figure 8 displays a general architecture of the random forest model.

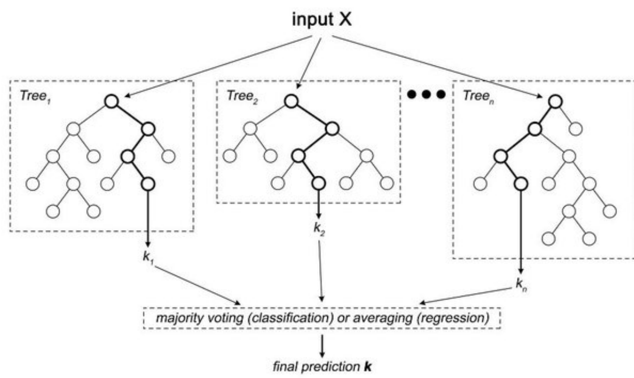

Figure 8: Structural overview of a random forest model in the general case.

Prior to training these baseline models, principal component analysis was done to overcome feature redundancy and reduce the dimensionality of the data. For the baseline, the number of components was calibrated such that 95% of the variance is retained within the datasets. For the logistic regression model, the maximum number of iterations was set to 1000 and the stochastic average gradient was employed as the solver due to its efficiency with large datasets when compared to other solvers. Likewise, the random forest classifier had the number of estimators set to 100, with a random state of 42 for the similar reasoning of quick computation. Figure 9 displays the results of both the logistic regression and random forest model after training.

Logistic Regression Model Performance:

	precision	recall	f1-score	support
hypodontia	0.80	0.75	0.78	126
Tooth Discoloration	0.16	0.37	0.23	19
Ulcer	0.61	0.74	0.67	27
Gingivitis	0.73	0.62	0.67	235
Caries	0.22	0.27	0.24	22
Calculus	0.55	0.58	0.56	130
accuracy			0.62	559
macro avg	0.51	0.56	0.52	559
weighted avg	0.66	0.62	0.64	559

Random Forest Model Performance:

	precision	recall	f1-score	support
hypodontia	0.75	0.78	0.77	126
Tooth Discoloration	0.50	0.11	0.17	19
Ulcer	1.00	0.37	0.54	27
Gingivitis	0.62	0.76	0.68	235
Caries	0.25	0.05	0.08	22
Calculus	0.60	0.55	0.58	130
accuracy			0.65	559
macro avg	0.62	0.44	0.47	559
weighted avg	0.64	0.65	0.63	559

Figure 9: Baseline model performance in different metrics.

The baseline models still show a pretty substantial loss, but they fare better than a purely random baseline, as can be seen. The training procedure presented a number of difficulties, the most significant of which were the limitations on available resources. Computing and training took a long time due to the volume of data, and even though the baseline model types were chosen to maximize computing efficiency, numerous training sessions were required to further strengthen the performance of the primary model. In addition, there were the laborious processes of importing, analyzing, and separating the data in order to train the baseline models.

The performance of the Logistic Regression (LR) and Random Forest (RF) models on the oral disease classification project reveals interesting insights into their ability to handle various classes. The

LR model demonstrates a tendency to favour recall, effectively identifying most cases of conditions like Tooth Discoloration and Ulcers, albeit sometimes at the cost of precision. This suggests it may be prone to false positives. In contrast, the RF model excels in precision for classes like Ulcer, indicating a conservative approach with fewer false positives but potentially missing some true cases. Notably, both models show commendable performance on Hypodontia, suggesting that its features are well-captured. However, they struggle with Caries and Tooth Discoloration, hinting at these conditions' complex or underrepresented nature in the dataset. These qualitative findings highlight the models' distinct approaches to classification and underscore areas for further improvement, especially in handling less distinct conditions which will be addressed in our neural network architecture.

4 Results

In the final configuration of our model, we employed the ViT model for image feature extraction coupled with an ANN for classifying six oral disease classes. This innovative approach harnessed the ViT's proficiency in detecting intricate patterns within the images, subsequently informing the ANN in distinguishing the classes effectively. In the next two subsections, we discuss the quantitative and qualitative results associated with our training and testing post-model selection.

4.1 Quantitative

Throughout a training regimen spanning 25 epochs, the model attained a high training accuracy of 89.91% and a validation accuracy of 81.83%. This demonstrated to our team that our model's predictive capabilities were strong and could generalize well. Precision was observed at 0.7865, and the model demonstrated a remarkable recall of 0.8252, culminating in an F1 score of 0.7987. The precision metric is indicative of the model's accuracy when issuing a positive class prediction, whereas the recall metric showcases the model's ability to capture true positive cases. The F1 score highlights the model's balanced performance between precision and recall, reinforcing the model's overall predictive accuracy.

When evaluated against our untouched test set, the model achieved a test accuracy of 82.29%. The precision on the test set stood at 0.8151, with a recall of 0.8233, and an F1 score of 0.8166, reinforcing the model's robustness evidenced during training. Looking at Figure 10 below, we noticed slight overfitting within our model as the validation loss curve decreases slower than the train loss curve. Although we did perform parameter tuning to mitigate this, we did not have enough computing to do this process as thoroughly as we would have liked. In the future, we would want to explore different regularization techniques and model complexities through K-fold cross-validation to overcome this issue.

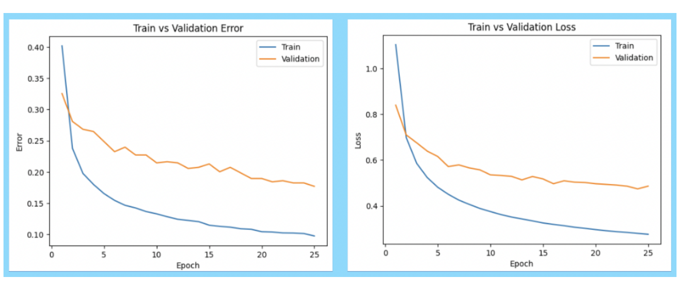

Figure 10: Baseline model performance in different metrics.

4.2 Qualitative

The confusion matrix in Figure 11 offers a visual and qualitative representation of the model's performance across the various classes on the test set. The matrix reveals that the model exhibits a commendable degree of specificity for most classes. However, it also uncovers a confusion between ’Calculus’ and ’Gingivitis’—several ’Calculus’ instances were misclassified as ’Gingivitis’, and vice versa.

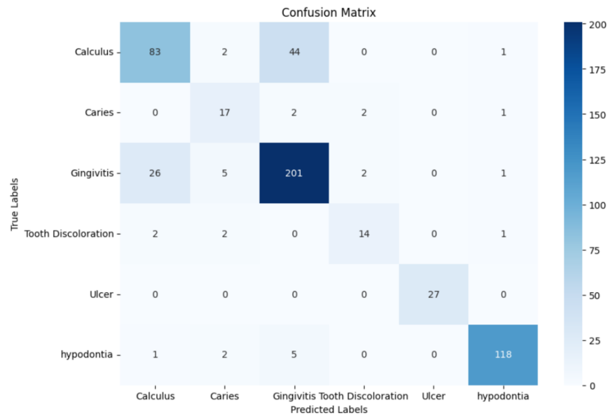

Figure 11: Test set confusion matrix for ViT network.

This pattern of misclassification can be attributed to several plausible factors. Firstly, it may stem from intrinsic similarities in the visual presentation of ’Calculus’ and ’Gingivitis’ in the image data, such as colouration and textural overlap, which could lead to feature extraction that is not sufficiently discriminative. Secondly, the training data may have had an imbalanced representation for these classes, potentially biasing the model towards the more prevalent class. From a qualitative standpoint, the outputs of the model illustrate its competencies and the challenges it faces with complex, nuanced tasks. For instance, while ’Gingivitis’ and ’Calculus’ share overlapping characteristics, differentiating these effectively is crucial for accurate clinical diagnosis. The confusion matrix suggests that the model has room for improvement in this area, possibly through targeted data augmentation to better represent underrepresented features, or by fine-tuning the feature extraction layers to better capture distinct pathological features.

The model’s performance on classes with a higher incidence in the dataset, such as ’Gingivitis,’ is indicative of the model’s ability to learn from abundant data. Conversely, for rarer conditions like ’Caries,’ which have fewer instances, the model exhibited fewer true positives, suggesting a potential area for further improvement through data augmentation or oversampling. Sample outputs, specifically for conditions underrepresented in the dataset, shed light on instances where the model could be prone to errors. For instance, the confusion matrix highlights instances where ’Caries’ was misclassified as ’Calculus,’ which could be attributed to similarities in visual features or an imbalance in class representation.

5 EVALUATION

A critical component of our project’s success hinges on the model’s generalizability and its ability to maintain performance when introduced to novel data—data that was neither used in training nor influenced the hyperparameter tuning process. Upon completion of the model training, we evaluated our model on our test set which the model had never been exposed to. This part of the dataset which was also sourced from Kaggle mirrored the diversity and complexity of the oral diseases in question, ensuring a rigorous and unbiased evaluation of the model’s predictive capabilities.

When the model was deployed on the test set, it delivered an accuracy that aligned with our training expectations, reaffirming the effectiveness of the ViT feature extraction and the ANN classification stages. The model’s precision, recall, and F1 scores on this new data were comparable to those obtained during the test phase, suggesting that the model did not suffer from overfitting to the training data and could generalize well to new samples. In addition to our test set, we collected multiple new samples of each class from different medical sources on the web to do further testing. The performance on the new data not only met but in some cases, exceeded our expectations, particularly in the classification of 'Ulcer' and 'Hypodontia', where the model demonstrated heightened sensitivity and specificity. This outcome is indicative of the model's ability to learn nuanced features during the training phase that are pivotal in distinguishing between various oral conditions.

6 Discussion

Reflecting on the results obtained from our model's performance, there is a persuasive argument to be made in favour of its efficacy. The hybrid approach, which combined the Vision Transformer (ViT) for feature extraction with an Artificial Neural Network (ANN) for classification, has yielded a model that showcases both high accuracy and generalizability across unseen data. This is not a trivial achievement, particularly given the nuanced complexities involved in oral disease imagery and classification. The model's robustness is exemplified by its precision and recall metrics, which stood at 0.7865 and 0.8252 respectively, indicating a strong capability in correctly identifying diseases and minimizing false positives. Furthermore, the F1 score of 0.7987 on the test set corroborates a balanced performance between precision and recall. This balance is crucial in medical diagnostics, where the cost of false negatives can be as significant as false positives.

One particularly intriguing aspect of our results was the confusion matrix's indication of a higher-than-expected misclassification rate between 'Calculus' and 'Gingivitis'. This could be attributed to the similar textural features in dental imaging, a hypothesis that opens new avenues for refining the model further. It suggests that while our model is adept at feature extraction, there may be room for improving its specificity by introducing more nuanced layers of image processing or by enriching the training dataset with a wider array of examples for these particular diseases. What is particularly encouraging—and, in a sense, surprising—is the model's performance on the new data, with an accuracy that not only resonated with our test data results but also, in some instances, surpassed them. This affirms that the model is not simply memorizing the training data but truly understanding the distinguishing features of each disease class.

Through the course of this project, we have learned the value of a well-structured hybrid model that leverages the strengths of both convolutional bases and transformer models. We've realized the importance of comprehensive data preparation and the benefit of feature-rich models capable of capturing the subtleties of complex image data. From this project, we've seen firsthand the importance of extensive validation strategies, including the deployment of new datasets, to ensure the model's efficacy is not merely a result of overfitting but a true representation of its capacity to understand and interpret complex patterns. The project's success sets a precedent for further exploration into deep learning techniques with medical diagnostics. Looking ahead, it's clear that continuous development, including iterative dataset enhancement and model refinement, is vital. In terms of scalability, we would want to deploy our model on a public web application once the model is refined and all ethical considerations are taken into account.

7 Ethical Considerations

When considering the ethical considerations for this project, we must consider the impact AI can have on healthcare and medical imaging. For our deep learning application, we need to consider the impacts a false positive or a false negative can have on a potential user of our application. If someone were to input an image into our model, and it returns a false positive, it can cost the user money in dental care visits, and unwanted stress. On the other hand, if a false negative were to occur, the user would leave the disease untreated as symptoms worsen.

If we evaluate the model's fairness using a disparate treatment notion, initially we can see that the model will treat people of different ethnicities similarly. This is because ethnicity is not present in the input images at all. The only part that is inputted into the model is the person's oral area. However, upon further research, teeth do have ethnic characteristics. For example, certain ethnicities will have higher gum ridges, which could lead to a misdiagnosis of gingivitis, as mentioned by Rawlani (2017). We would hope to measure this using a fairness-by-blindness measure.

REFERENCES

Yassir Edrees Almalki, Amsa Imam Din, Muhammad Ramzan, Muhammad Irfan, Khalid Mahmood Aamir, Abdullah Almalki, Saud Alotaibi, Ghada Alaglan, Hassan A Alshamrani, and Saifur Rahman. Deep learning models for classification of dental diseases using orthopantomography x-ray opg images. *Sensors*, 2022.

Abdallah Ashraf. Oversampling - handling imbalanced data, 2023. URL `https://medium.com/@abdallahashraf90x/oversampling-for-better-machine-learning-with-imbalanced-data-68f9b5ac2696#:~:text=Oversampling%20is%20a%20data%20augmentation,to%20the%20under%2Drepresented%20class`.

Gaudenz Boesch. Vision transformers (vit) in image recognition – 2024 guide, 2024. URL `https://viso.ai/deep-learning/vision-transformer-vit/`.

Hu Chen, Hong Li, Yijiao Zhao, Jianjiang Zhao, and Yong Wang. Dental disease detection on periapical radiographs based on deep convolutional neural networks. *International Journal of Computer Assisted Radiology and Surgery*, 2021.

Sultan Imangaliyev, Monique H. van der Veen, Catherine M. C. Volgenant, Bart J. F. Keijser, Wim Crielaard, and Evgeni Levin. Deep learning for classification of dental plaque images. In *Machine Learning, Optimization, and Big Data*, 2016.

Nikita Kozodoi. Computing mean std in image dataset, 2021. URL `https://kozodoi.me/blog/20210308/compute-image-stats`.

Jae-Hong Lee, Do-Hyung Kim, Seong-Nyum Jeong, and Seong-Ho Choi. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *Journal of Dentistry*, 2018.

Llourenc. The importance of logistic regression in image classification, 2022. URL `https://medium.com/@MudSnail/the-importance-of-logistic-regression-in-image-classification-1966d07e7a0c`.

Ashish Mishra. Logistic regression: The baseline classification model explained, 2021. URL `https://medium.com/@ashishmishra2k19/logistic-regression-the-baseline-classification-model-explained-89915f5b13b2`.

Shreyansh A. Prajapati, R. Nagaraj, and Suman Mitra. Classification of dental diseases using cnn and transfer learning. In *2017 5th International Symposium on Computational and Business Intelligence (ISCBI)*, 2017.

Shiva Rawlani. Racial characteristics of human teeth, 2017. URL `https://www.researchgate.net/publication/316894227_Racial_characteristics_of_human_teeth`.

Harpreet Sahota. Can't decide between a linear regression or a random forest? here, let me help., 2022. URL `https://medium.com/artificialis/cant-decide-between-a-linear-regression-or-a-random-forest-here-let-me-help-ab941b`

Akash Takyar. How is a vision transformer model (vit) built and implemented?, 2023. URL `https://www.leewayhertz.com/vision-transformer-model/#:~:text=ViT%20relies%20on%20self%2Dattention,tools%20for%20image%20identification%20tasks`.

TensorFlow. Data augmentation, 2024. URL `https://www.tensorflow.org/tutorials/images/data_augmentation`.