Aiding Classification and Diagnostic Interpretation through Image Recognition in Dental Radiography

Chandrakanth Bodapati, Bharath Irigireddy, Sri Sai Charan Velisetti
Department of Computer Science
University of Maryland
College Park
chand308@umd.edu, bharathc@umd.edu, svellise@umd.edu

1 Abstract

An estimated 2 billion people suffer from decay/caries of permanent teeth and 514 million children suffer from caries of primary/pediatric teeth worldwide. The natural full count of teeth which is 10 for a child keeps increasing until it reaches 32 at around the age of 25, the different classes of teeth present and the different degrees of invasion of pathology and/or anomaly create large chunks of data, especially in the radio-graphic realm. There is a need for an adaptive model to classify with accuracy and speed that would make diagnosis an efficient and effective process. The field of research would also benefit immensely from such a trained model-making classification, identification, and creation of exhaustive databases thereby finding meaning in mountains of data. Pediatric molars, Adult Molars, and Adult Bicuspids were taken as study groups since their occlusal terrain encourages caries progress and also due to the availability of data.

2 Introduction

Although primary teeth are considered to be essentially miniature versions of permanent teeth numerous differences exist between them. They have a larger pulpal volume, a higher degree of crown bulbosity, and root splaying with lesser hard tissue thickness. They additionally accommodate balls of tissue called tooth buds of corresponding permanent teeth under their roots which would be at different stages of development depending on the dental age of the child. The proper maintenance of the health of all the teeth in dentition cannot be stressed enough. However, from a functional and developmental point of view, the molars are one of the most important teeth contributing to the appearance and symmetry of the face and playing a key role in occlusion and Temporo-Mandibular Joint function. Unfortunately, they are often removed due to decay. Removing these teeth, especially at an early age can cause misalignment of the rest of the teeth, even leading to asymmetric development of the face. The teeth next to the newly vacant space can start tipping toward the empty space, the opposing teeth start to grow out, and bone is lost where the tooth used to be. Up to 90% of the chewing function takes place in the molar area. This affects negatively their Quality of life and their ability to enjoy food.

Any illness that can manifest inside the mouth, including the salivary glands or jaws, is referred to as oral pathology. It is always important to get an early examination to obtain the right treatment, even if the majority of oral pathology is benign and not harmful. With the introduction of deep learning in contemporary computers, image processing and segmentation have made incredible strides. With the goal of automating diagnosis and treatment, deep learning-based picture recognition has made tremendous progress in the interpretation of dental radiographs. Deep learning-based image analysis in the context of dental imaging has demonstrated nearly 90% accuracy in the segmentation, classification, and diagnosis of various prevalent dental disorders. These findings offer a window of opportunity for improved dental medicine diagnosis and treatment planning.

CMSC 733: Computer Processing of Pictorial Information (Fall 2022), .

The application of deep learning techniques in the health sector especially in the detection of cancer is prevalent nowadays. Skin cancer is one of the most common forms of cancer and also one of the most dangerous forms out there therefore, early detection is key to better chances of survival. This has not only inspired us to do a comparative study of different Convolution Neural Network models by using transfer learning, but also made us think about designing our own model and testing its performance on the data set.

3 Data

A total of 30,000 images were procured from dental hospitals and we manually picked the best and scoured through the data to filter out the images of the teeth that were not being used for this study. We did this with the help of a few professionals in the field who could visually identify the given four types of teeth and most of the data was low resolution and which did not focus on the teeth that we were doing our study on. We were able to get around 550 images with an adequate resolution for each type. We have reserved 10% data for testing our model.

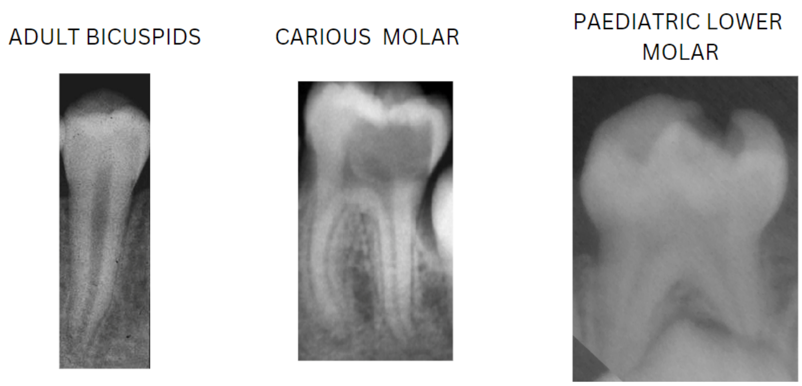

Figure 1: Sample Images from the data-set

3.1 Data Augmentation

In data analysis, there are methods for increasing the amount of data by adding copies of already existing data that have been slightly modified or by generating new synthetic data from existing data. When training a machine learning model, it functions as a regularizer and helps minimize overfitting. Oversampling in data analysis is directly related to it.

- Flipping (both vertically and horizontally)
- Rotating
- Zooming and scaling
- Adding Gaussian noise (distortion of high-frequency features)
- Translating (moving along the x or y-axis)

4 Problem Statement

-This project mainly focuses on providing a highly accurate deep learning architecture to predict the type of tooth the X-ray scan is given. For the initialization of the model, we first clean the images and perform data pre-processing. For this purpose, we have collected a data set on our own by visiting dental hospitals. The input for the project will be a sequence of x-ray images of 4 different types of teeth from 600 patients . We will be building our own CNN architecture and doing some hyper-parameter tuning to come up with the best model that gives the highest accuracy. These methods will be analyzed based on the following metrics:

- F1 - Score

- Confusion Matrix
- Overall Loss

5 Related Work

- The past few years, the increase in medical imaging and scientific knowledge has caused exponential growth in databases and repositories. Details of clinical symptoms to various types of biochemical data and outputs of imaging devices are now widely available. Das k, et al [1] This paper explains how skin cancer detection still lags behind and mentions how AI could be the solution as early detection of cancer is crucial. Elgamal, M. [2] This paper focuses on dimensionality reduction, by relating features through a discrete wavelet transformation and feed-forward backpropagation. Bisla D [3] This paper implements data purification and augmentation to overcome occlusions and data that are heavily imbalanced.

6 Method

The architecture of this dental data classification model is based on the ViT (Vision Transformer) model, which is a transformer-based model that has been modified for image classification tasks. The model takes in images of size 224x224 pixels and processes them through several patch-based transformer layers, each with a patch size of 28x28 pixels. The model has a depth of 6 and uses 8 heads for attention. It also has a hidden dimension of 128 and an MLP (multi-layer perceptron) dimension of 256, with a dropout rate of 0.1 for regularization. The model is trained using the Adam optimizer with a learning rate of 0.001 and the categorical cross-entropy loss function.

Vision transformers are a type of deep learning model that are based on the transformer architecture, which is a type of model that has been successful in natural language processing tasks. The transformer architecture uses self-attention mechanisms to process input sequences, which allows the model to capture long-range dependencies in the data. In the case of vision transformers, the input sequences are images, which are divided into patches and processed through the transformer layers. The use of transformers allows for the model to capture both global and local features in the images, which can be beneficial for image classification tasks.

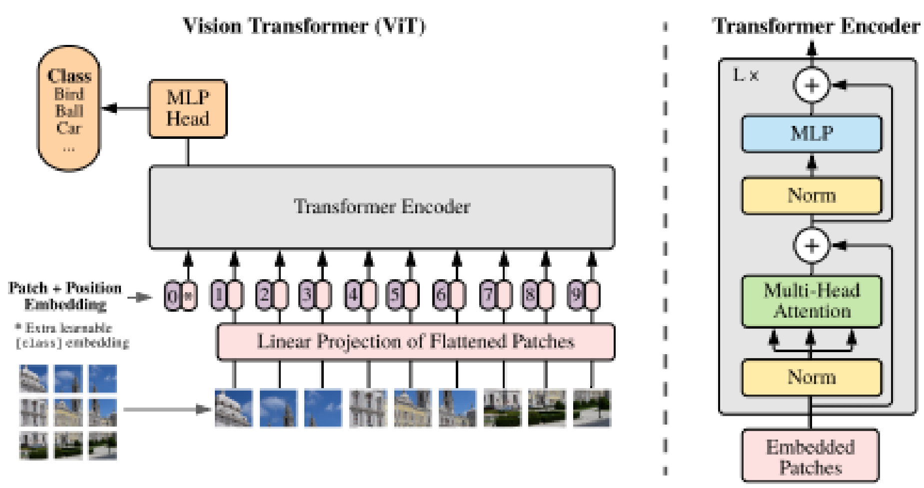

Figure 2: Model Architecture

They also have a relatively simple architecture compared to other convolutional neural network (CNN) based models, with fewer parameters and no need for pooling layers. Overall, vision transformers offer an alternative approach to traditional CNNs for image classification tasks and may be particularly useful for tasks with large input sizes or complex relationships between features.

The training loss and testing loss during training are as follows:

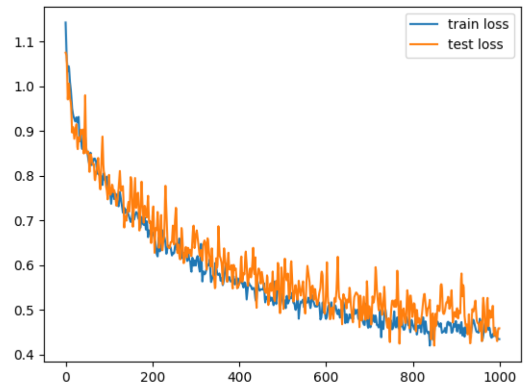

Figure 3: Loss per epochs

6.1 Evaluation Metrics

6.1.1 F1 - Score

Whenever the accuracy metric is used, we aim to learn the closeness of a measured value to a known value. It's therefore typically used in instances where the output variable is categorical or discrete — Namely a classification task. In instances where we are concerned with how exact the model's predictions are we would use Precision. The precision metric would inform us of the number of labels that are actually labeled as positive in correspondence to the instances that the classifier labeled as positive. Recall measures how well the model can recall the positive class (i.e. the number of positive labels that the model identified as positive). Precision and Recall are complementary metrics that have an inverse relationship. If both are of interest to us then we'd use the F1 score to combine precision and recall into a single metric. Compared to the raw accuracy metric, the F1 score provides more robust evaluations of imbalanced datasets in

Given the True Positives (TP), False Positives (FP), True Negatives (TN), False Negatives (FN) between ground truth and predictions, Following is the mathematical formulation of F1 Score:

$$Precision = \frac{TP}{TP + FP} Recall = \frac{TP}{TP + FN}$$

	Precision	Recall	F1- Score
Paediatric Lower Molar	0.69	0.65	0.67
Adult Mandibular Molar	0.98	0.97	0.97
Adult Bicuspids	0.97	0.98	0.98
Carious Molar	0.66	0.70	0.68

Table 1: Evaluation Metric

6.1.2 Confusion Matrix

An algorithm's performance may be seen using a particular table structure called a confusion matrix, also known as an error matrix. This type of method is commonly one for supervised learning (in unsupervised learning it is usually called a matching matrix). Both variations of the matrix, where each row represents examples in an actual class and each column represents instances in a predicted class, are documented in the literature. The name was chosen since it is simple to determine whether the system is conflating two classes (i.e. commonly mislabeling one as another).

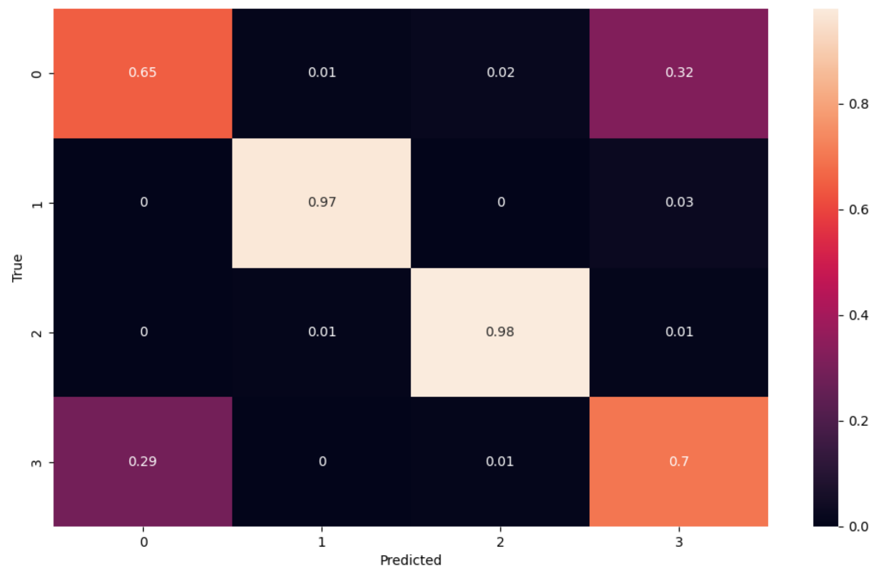

Figure 4: Confusion Matrix

7 Results

Based on the confusion matrix provided, it appears that the model is performing relatively well on the data. The diagonal elements of the matrix, which represent the number of correct predictions made by the model, are generally larger than the off-diagonal elements, which represent incorrect predictions.

In the first row of the matrix, the model correctly classified 65% of the paediatric lower molar samples. Similarly, the model correctly classified 97% and 98% of the adult mandibular molar and adult bicuspid samples, respectively.

There are a few areas where the model appears to be struggling, such as in the classification of carious molars. In this case, the model correctly classified 70% of the samples, but made incorrect predictions for the remaining 30%.

Overall, the model seems to be performing well. However, there is room for improvement, particularly in the classification of carious molars and Paediatric Lower Molar. Further analysis and optimization of the model may be necessary to improve its performance in this area.

8 Challenges faced

Procuring the data posed a number of challenges including access to the data, data interpretation, and sieving through a myriad number of radiographs to gather useful information. Access to the data presented a two-pronged challenge- Hardly any private clinics stored this kind of data and special permission was taken to obtain the data itself after reaffirming patient confidentiality. The data that was obtained was available in dental imaging formats of .raw and .rvg among others which were not readily readable and had to be converted to readable formats. Readable data itself was quite big and sorting it into usable data for the project was incredibly exhausting and time-consuming, to say the least. It cannot be left unsaid that the strain taken to gather usable data for this project could have been avoided, had a similar trained model been available such as this project aims for.

9 Conclusion

We have explored applications of various Machine Learning models to identify and classify dental images of paediatric lower molar, carious molar, adult bicuspids, and adult mandibular molar. We have experimented with various machine-learning architectures such as CNNs, ResNet, and ViT we have concluded that ViT had the best performance on our dataset. This is a proof of concept that dental analysis and classification through a Vision Transformer would greatly reduce the amount of time and manpower required to document and label dental imaging, allowing doctors to focus on more complex problems that classifying teeth.

References

[1] Das K, Cockerell CJ, Patil A, Pietkiewicz P, Giulini M, Grabbe S, Goldust M. Machine Learning and Its Application in Skin Cancer. International Journal of Environmental Research and Public Health. 2021; 18(24):13409. https://doi.org/10.3390/ijerph182413409

[2] Elgamal, M. (2013). AUTOMATIC SKIN CANCER IMAGES CLASSIFICATION. International Journal of Advanced Computer Science and Applications, 4.

[3] Devansh Bisla, Anna Choromanska, Russell S. Berman, Jennifer A. Stein, David Polsky; Towards Automated Melanoma Detection With Deep Learning: Data Purification and Augmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2019, pp. 0-0.

[4] Shweta Suresh Naik , Dr. Anita Dixit, 2019, Cancer Detection using Image Processing and Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH and TECHNOLOGY (IJERT) Volume 08, Issue 06 (June 2019),

[5] Nripendra Kumar Singh, Khalid Raza, Progress in deep learning-based dental and maxillofacial image analysis: A systematic review, Expert Systems with Applications, Volume 199, 2022, 116968, ISSN 0957-4174,