An Approach to Classify Alzheimer's Disease using Vision Transformers

MSc Research Project
Data Analytics

Anitha Drewitt
Student ID: X21122954

School of Computing
National College of Ireland

Supervisor: Qurrat Ul Ain

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Anitha Drewitt
Student ID:	X21122954
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Qurrat Ul Ain
Submission Due Date:	15/12/2022
Project Title:	An Approach to Classify Alzheimer's Disease using Vision Transformers
Word Count:	4481
Page Count:	14

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	1st February 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

An Approach to Classify Alzheimer's Disease using Vision Transformers

Anitha Drewitt
X21122954

Abstract

Over the past decades, different deep-learning approaches in Medical Imaging have shown promising results and performance. Classifying the stages of a disease and thus its early treatment may avoid its further spread resulting in a reduced fatality rate. A neurodegenerative condition called Alzheimer's disease (AD) is likely to grow, spread, and deteriorate. It causes our brain cells to die leading to complete loss of memory and physical impairment. According to estimates, nearly 6 million Americans aged 65 and older have Alzheimer's disease, which is a prominent cause of death in developing nations. The study investigates a novel approach for classifying Alzheimer's disease called "Vision Transformer". The research employs publicly available datasets on Kaggle by classifying images into four stages which are Mild Demented, Very Mild Demented, Non-Demented, and Moderate Demented and the accuracy obtained from the model is compared with other deep learning approaches and the limitations and future scope of this research methodology are presented. Accuracy and F1-Score has been used to evaluate the performance of the model. The model has an accuracy rate of 87.5% and a loss of 0.34 in classifying Alzheimer's disease. On comparing the model with the other historical CNN methods, the algorithm has produced encouraging results for classifying AD and it can be enhanced for wider application. The proposed model can help the physician to classify AD and provide treatment to the affected person which can eventually help in reducing the fatal rate.

1 Introduction

Alzheimer's disease (AD) is a complicated, fatal condition of the nervous system that gradually destroys brain cells, weakening memory and thinking, and eventually rendering patients incapable of performing even the most basic tasks. Dementia is the result of its effects on cognitive function. The disease in its earlier stage does not have any symptoms but grows, deteriorates, and worsens over time as a neurodegenerative form of dementia. Alzheimer's disease must be identified with a comprehensive medical examination that includes the patient's medical history, a short mental state examination (MMSE), physical exams, and neurobiological studies. Resting-state functional magnetic resonance imaging (rs-fMRI) and structural magnetic resonance imaging are non-invasive methods for studying the structure of the brain, functional brain activity, and changes in the brain in addition to these assessments. Patients are positioned prone on the MRI table for the duration of scanning utilizing the structural and rs-fMRI procedures. This makes it possible to collect data without worrying about how a particular task may affect the brain's ability to function. The cerebral cortex and hippocampus shrink as a result of Alzheimer's, while the ventricles of the brain grow. The cerebral cortex and hippocampus have seen considerable shrinkage in the late stages of AD, while the ventricles have significantly increased. The brain areas and networks responsible for thinking, remembering, planning, and judgment are impacted by this impairment. Both MRI and rs-fMRI image intensities are low because brain cells in the injured regions have deteriorated. Depending on the stage of the disease's development, the degree of alteration in various brain regions varies. Using MR imaging, it is simple to spot a considerable reduction in the volume of the cerebral cortex and hippocampus as well as a major enlargement of the ventricles.

For clinicians, it is crucial to create an algorithm that can distinguish between a brain that is normal and healthy and one that is dysfunctional and disordered. The medical community is already making significant contributions to the field of AD classification and is already seeing astounding results. To differentiate between AD and normal aging by examining the observable changes in brain regions on an MRI scan, where a diagnosis is still susceptible to interpretation, much expertise and experience are needed. Therefore, for a more precise data classification, the imaging data should be integrated with additional clinical outcomes like MMSE.

Vision Transformers are used in the experiment to classify Alzheimer's disease.In this study, an unique classification method utilizing advanced CNNs known as transformers is applied. The availability of the labeled dataset, Alzheimer's MRI scans, served as the research's motivating factor. The Dataset is divided into test train and validation and is categorized into 4 stages: Mild Demented, Moderate Demented, Non-Demented, and Very Mild Demented. Medical experts grade the dataset, and it is then assessed by a second professional expert. Due to a lack of data to train the model on in this research, classifying Alzheimer's from magnetic resonance imaging is a challenging task.Therefore, Data Augmentation is used for this purpose.

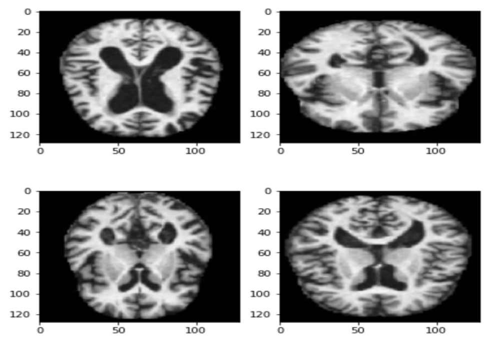

Figure 1: Images of MRI Data set containing four stages

The main objective of the research study is to classify different stages of Alzheimer's disease (AD) into four stages and to evaluate the Severity of Alzheimer's disease (AD) using a novel approach Vision Transformers and to compare the model performance with the historical methods.

The research question of the project is given below:

How accurately the stages of Alzheimer's disease could be classified from the Medical Resonance Imaging (MRI) using Vision Transformers?

The sections of the report study are as follows. The need is explained and an overview of the research is given in the first section. Section 2 of the research paper covers the related work that was done. Section 3 of this article provides a description of the research methodology that was used. Design specifications for the employed architectures are presented in Section 4.The project's implementation is covered in Section 5. Section 6 provides explanations of evaluations, which concludes with a discussion of the research's future work.

2 Related Work

The first stage of the five-stage approach involved pre-processing the images and segmenting them into GM, WM, and CSF. The GM segmented ROIs were used to build similarity matrices in the second stage, and statistical features were obtained in the third stage. In the last two rounds of the investigation, the functional activities questionnaire (FAQ) and a support vector machine (SVM) classifier were used to categorize the data into the AD vs. normal group. In (Grady et al.; 2016), a 3D displacement-field estimation was used to distinguish between AD patients and healthy subjects. The student t-test, Welch's t-test, and the Bhattacharyya distance were the three methods that were utilized to choose the features. The data was classified using an SVM classifier, and the classification accuracy was 93.5%. By comparing AD patients to healthy controls, voxel-based morphometry identified local and overall GM atrophy (VBM) (Lowe; 2004). The VOIs were divided into areas with noticeably reduced GM volume. These voxel values were used to create a feature vector, which was then ranked using t-test results and a genetic algorithm to choose the best subset of features.The accuracy of the classification for the AD vs. normal and MCI vs. normal classes, respectively, was 84.17% and 70.38% when utilizing SVM with 10-fold cross-validation.

GM volume was evaluated using VBM for AD patients and healthy controls, and regions that had undergone a significant amount of GM atrophy were designated as VOIs, according to (Simonyan and Zisserman; 2014). Seven different feature ranking techniques were employed to assess the voxel values from these regions. SVM was used to classify the subjects, with an accuracy rate of 92.48%.

The AD was categorized in (He et al.; n.d.) using a Laplace Beltrami eigenvalue shape descriptor. By employing a response diffusion level set approach to segment T1-weighted MRI data, the changes in the corpus callosum's structure were examined.Information gain ranking was used to choose the important features, which were subsequently categorized using SVM and K-nearest neighbor (KNN). Using the KNN classifier, a maximum classification accuracy of 93.37% was attained.However, because it is difficult to quantify variations in the microstructure of the corpus callosum, the approach is less practical in practice. Jack2008(Jack et al.; 2008) offer a paradigm for the extraction of features from inter-subject variability represented by low-dimensional subspaces. The manifold subspace was built using ROIs that were driven by data. A sparse regression with the MMSE score was used to find these areas. By performing variable selection using sparse regression, the sample bias was decreased in conjunction with a resampling strategy. The categorization accuracy rate was 71%.

In citepSuk2013(Suk and Shen; n.d.), a method for classifying AD and cognitively normal (CN) patients using the sulcal medial surface was proposed. 24 distinct sulci were retrieved from every patient using the Brain-VISA sulcal identification pipeline, and SVM was employed to classify AD and CN as well as compute the sulcal medical surface features and a accuracy of 87.9% was achieved.

In (Suk et al.; 2015) a bio-marker was created using information from MRI scans by combining different parameters, including cortical thickness and the texture and shape of the hippocampus.The method was trained using MR images from the Alzheimer's disease Neuroimaging Initiative (ADNI) database using linear discriminant analysis (LDA). The study revealed that this MRI biomarker combination achieves a multi-class classification accuracy of 62.7%.

The hippocampus and posterior cingulate cortex's circular harmonic functions (CHFs) were used in CitepYin2022(Yin et al.; 2022) to extract regional features. The task comparing AD and MCI had a classification accuracy of 62.07%. To categories AD subjects, a deep learning approach was described in (Suk et al.; 2014). Convolutional neural networks and auto-encoders were used to predict the output classes as AD and normal, with a classification accuracy of 98.4%. Other classifications and multi-class classification were not considered. To extract distinguishing characteristics for the classification of AD and normal people, convolution neural networks were applied (Payan and Montana; 2015), (Liu et al.; 2015). Even while deep learning-based methods for big data analysis have achieved considerable successes, it still takes a significant amount of training and processing time to extract useful information from huge amounts of unstructured data.

Choosing the appropriate hyper-parameters and architecture is another difficult task. The binary classification of AD and MCI as well as the multi-class classification of AD, normal, and MCI are challenging tasks. The efficiency of the most approaches described in the literature for classifying data was constrained because they only utilized features that were directly retrieved from brain pictures.

The way that Alzheimer's disease alters brain structure and function has drawn the attention of numerous academics and research organizations. Numerous studies have been done on the classification and prognostication of the phases of Alzheimer's disease, especially in diagnostic imaging. Using learning-based techniques (Suk and Shen; n.d.)), (Suk et al.; 2015) , and (Suk and Shen; 2015) classified structural MRI and PET data from AD magnetic current imaging (MCI) and MCI-converter with accuracy rates of 95.9%, 85.0%, and 75.8% respectively. Using an auto-encoder network, the researcher removed low- to mid-level features from photographs. A multi-task and multi-kernel Support Vector Machine (SVM) learning strategy was then used for classification. To improve this workflow, multimodal MRI/PET data and more intricate SVM kernels were employed. However, the study's highest accuracy rate stayed the same (Suk et al.; 2014). To distinguish the imaging of AD MCI from that of normal, healthy control patients, (Payan and Montana; 2015) of Imperial College London developed a prediction algorithm. This experiment used an auto-encoder with a 3D convolutional neural network design. In separating AD from NC individuals, the researcher achieved an accuracy rate of 95.39%. The research team also experimented with a 2D CNN design, and the results showed that the accuracy rate was essentially the same. A multimodal neuroimaging feature extraction technique for multiclass AD diagnosis was also developed by (Liu et al.; 2015). This deep-learning framework was developed using a zero-masking method to keep every piece of information in the imaging data. With regard to multimodal and multiclass MR/PET data, high-level features were retrieved using stacked auto-encoder (SAE) networks, and classification was carried out using SVM. In that investigation, the maximum accuracy rate attained was 86.86%.

A unique image pre-processing technique has been developed to address the problems of lower accuracy in prediction challenges since the necessity for the Image processing technique is not focused in most of research publications with the Visionbased Transformers modelling.

3 Methodology

From the beginning of data collecting until the conclusion, the research process demands a well-organized framework. The step-by-step procedure used to carry out this research is covered in this section. Figure 2 below indicates the KDD methodology and its various steps employed in the research work.

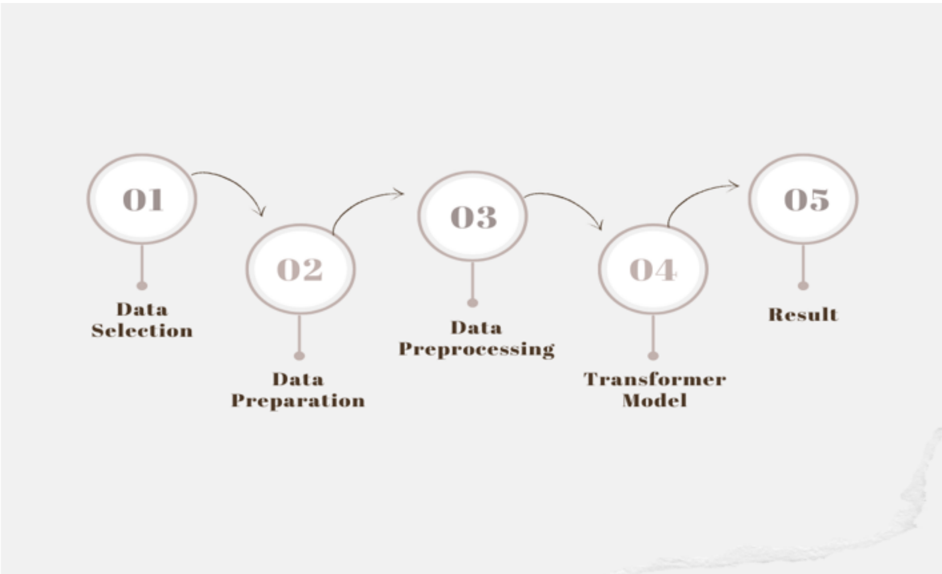

Figure 2: Steps carried out in the research methodology

Data used in this research work were obtained from an online source, Kaggle. The dataset contains 6400 images of four stages: Mild Demented, Very Mild Demented, Non-Demented, and Moderate Demented.

The dataset is split for training, validating, and testing the model and divided into 80% training and 20% testing. The training data is further divided into 20% validation and 80% training. As the data were not prepared for the model to train on directly so we prepared the data which were not available in uniform resolution. Data preprocessing is the next phase. The primary goal of pre-processing is an enhancement of the image data that enhances a key aspect of the image that will be used in subsequent preprocessing.

Data Augmentation is used in preprocessing the data. It is a technique which is used to increase the amount of data and it helps in avoiding the issue of overfitting. The details of the applied augmentation operations are listed below.

Scaling: In scaling or resizing, the image is scaled to the given size. The idea of resizing the image is to produce a lower data size in order to hastens the processing time.

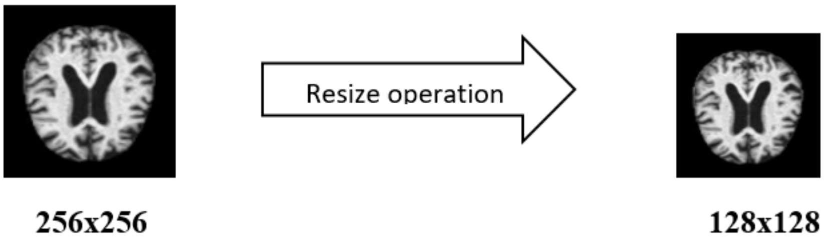

Figure 3: Scaling

Horizontal Flip: In horizontal Flipping the image is flipped horizontally. This operation reverses the entire rows and columns of an image pixels in horizontal direction.

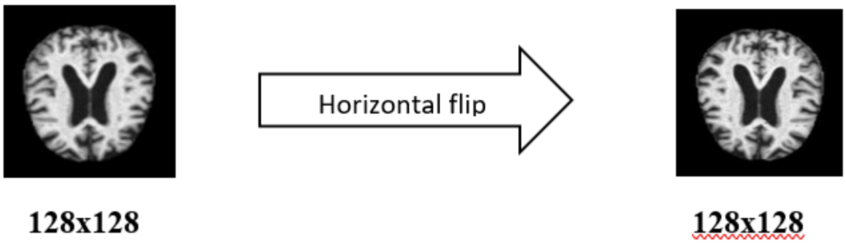

Figure 4: Horizontal Flip

Random Rotation: This augmentation technique will randomly rotate the image from 0 to 360 degrees in clockwise direction. The factor of rotation as a parameter must be mentioned and this technique will rotate the image with the same factor.

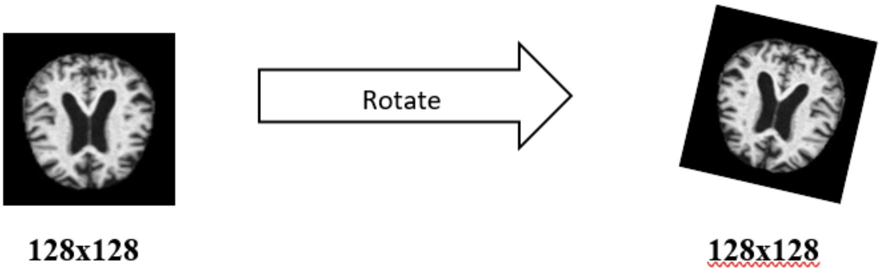

Figure 5: Random Rotation

Random Zoom: Random zoom is used to zoom the input image; this technique randomly zooms the image in and out on each axis independently.

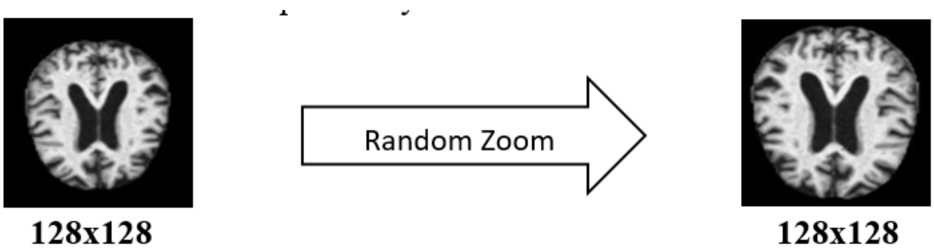

Figure 6: Random Zoom

The augmented data is then passed through the pre-trained ViT architecture. A linear projection layer is used to project the image into a sequence of vectors. A convolutional layer performs the projection, which is then followed by a GELU activation function, batch normalization, and normalization. Classification head is used to predict the class of the image. The output of the transformer encoder is mapped linearly to the number of classes in the dataset by this layer. A positional encoding layer is used to add positional information to the sequence of vectors, which means that the model knows the position of each vector in the sequence. The positional encoding is done by adding a sinusoidal signal to the input sequence. Transformer encoder has been used which is the core of the vision transformer model. It is a stack of N identical layers.A MLP(Multi-Layer Perceptrons) layer is used to implement the feed-forward network in the transformer encoder. A dropout layer is used to prevent overfitting. In order to avoid overfitting during training, it randomly sets a fraction rate of input units to 0 at each update. For evaluation, Accuracy and F1 score has been used and have compared the model with the historical models.

4 Design Specification

4.1 Vision Transformer (ViT)

A technique for classifying images called Vision Transformer applies a Transformer-like design to selected areas of the image. The vision Transformer (ViT) model was initially stated in a study titled "An image worth 16*16 words" presented at the ICLR 2021 conference by Neil Houlsby et al. On the ImageNet-21k dataset and the ImageNet, their ViT model was pre-trained. To fully understand ViT model, it should be kept in consideration that it is both a deep learning and machine learning model. Moreover, its mechanism necessitates attention and diligence and ways differently the significance of the input data parts. The multiple self-attention layers that build the model are used in natural language processing(NLP) and in computer vision which is why it is said to hold strong promises toward generic learning methods when applied to different data modalities.

4.2 Comparative review of Vision Transformer

When comparing the performance of Vision Transformers with Convolutional Neural Networks (CNN), the former requires less resources while producing superior results.

Additionally, when the ViT is trained on a smaller data set, its weaker inductive bias causes it to rely more on either data augmentation or model regularization. The Vision Transformer model's image inputs are represented as a series of image patches. To further enhance the efficiency of the Vision Transformer model, it needs to be trained on enough data. Likewise, where Vision Transformer splits the images into visual tokens during the computation, the CNN usually uses pixel arrays. This embeds information globally in the overall image because of the self-attention layer. In addition, the ViT model can reconstruct the structure of an image once it learns from training data to encode the relative location of the image. Further Vision Transformer model has a multi-Head layer that concatenates all the outputs in the right dimensions and most of the attention heads are used by the model to train the global dependencies in an image and the local dependencies.

4.3 Multi-Layer Perceptron (MLP) Layer

This layer contains two layers with linear Gaussian errors (GELU). Layer Norm (LN) is added before each block since it does not create any extra dependencies between the training images. Training time is reduced, and performance is enhanced. Remaining connections are also created after each block since they allow direct component flow across the network without going via non-linear activations. The MLP layer implements the classification head in the context of image classification.It accomplishes this by pre-training with a single hidden layer and fine-tuning with a single linear layer. The data is additionally augmented during data pre-processing because the transformer model requires more data to train. Image resizing, flipping, zooming, and rotation are all part of the data augmentation process. It aids in preventing the model from overfitting.

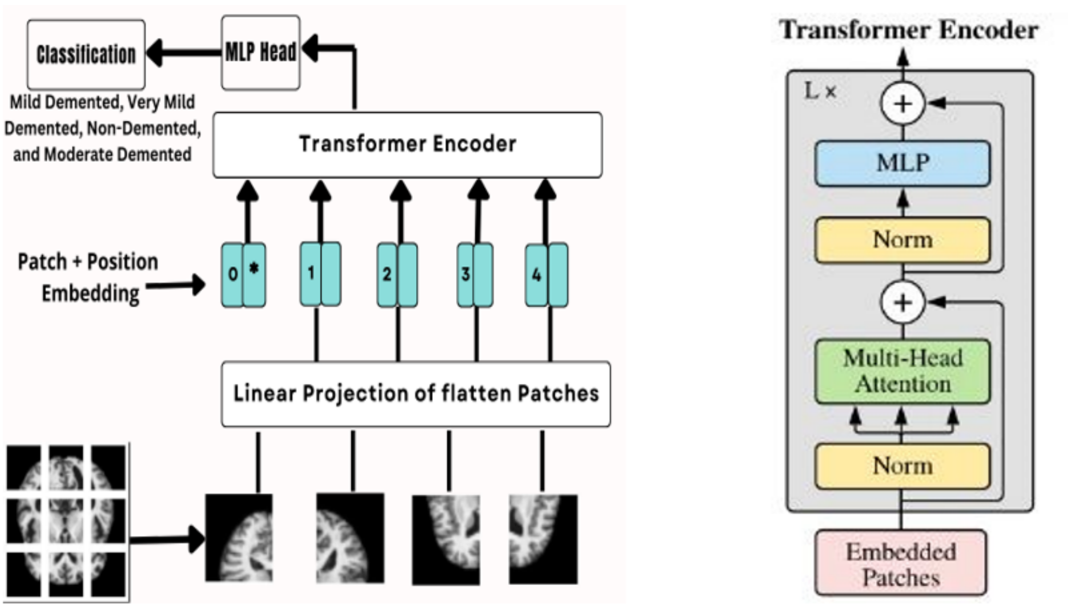

Figure 7: Workflow of Vision Transformer

5 Implementation

Using the Keras library, which includes Tensorflow as its backend, research is carried out in Python. Python-based Keras is an open-source neural network library. Tensorflow is used as the backend and is used to create and assess deep learning models. The arrays were processed with using the Numpy library, and the data visualization was done using Matplotlib. The MultiHeadAttention layer is used as a self-attention mechanism applied to the sequence of patches by the various Transformer blocks that make up the ViT model. The Transformer blocks produce a (batch size, num_patches, projection_dim) tensor, which is processed via a classifier head with softmax to produce the final class probabilities output. Unlike the technique described in the paper, which prepends a learnable embedding to the sequence of encoded patches to serve as the image representation, all the outputs of the final Transformer block are reshaped with layers. As the transformer Model is computational heavy due to its Encoder decoder architecture so we used google Colab which provide free GPU to run our code in that environment. The data is downloaded from Kaggle and performed some augmentation to increase the dataset. The augmented data is used to train the Model. Adam optimizer is used with learning rate of 0.001 and a batch size of 16 is used. As the images need to be in patches, the image was converted into 6x6 patches and then embed to the Model. Model has been trained for 200 epoch and got a testing accuracy of 87.5%. For evaluation the model confusion matrix is created for the model. F1-Score evaluates the model's performance, and the model provided an F1-score of 84%. In the evaluation section, the performance of the model is compared to other models in terms of the matrices.

6 Evaluation

In this research, the dataset was first split into training and testing with an 80:20 ratio. The training data was divided into a train set and a validation set with a ratio of 80:20 during the training phase. Every time a new epoch is trained, the model first learns the data from the training data and then validates the model using the validation data. Only the patch embedding layers and classifier will be updated during training in accordance with the AD/CN classification loss, leaving the transformer encoder's parameters alone. The transformer Model will learn from the training images and test the model on every epoch on the validation set. Accuracy and F1-Score are the important performance parameter in classification problem. The vision transformer model was employed in the experiment, and the outcomes were compared with various historical methods. The effectiveness of various model settings on the test set was reported, and the F1-Score was selected as the model evaluation metric for classification accuracy.

Comparison of Vision Transformer with the other historical models are given in the below table:

Study	Methods	Modalities	Accuracy	F1-score
Valliani and Soni (2017)	ResNet-50	MRI	0.81	-
Wen et al. (2020)	4 CNN blocks + 2 Fully Connected (FC) layer	MRI	0.85	-
Lian et al. (2018)	Hierarchical FCN + automatic discriminative localization	MRI	0.85	-
Korolev et al. (2017)	CNN	MRI	0.8	-
Bäckström et al. (2018)	CNN	MRI	0.9	-
Pan et al. (2020)	Spatially-constrained Fisher representation	MRI	0.894	0.88
Pan et al. (2020)	Spatially-constrained Fisher representation	MRI + PET	0.91	0.90
Qiu et al. (2020)	FCN+MLP	MRI+clinical data	0.968	0.965
Our model	Vision Transformer	MRI	0.875	0.84

Figure 8: Comparison of Model with historical Models

The performance of transformer is shown in the Figure 9. The model has been trained for 200 Epoch, as can be seen from the graph, where the yellow line denotes validation accuracy and the blue line, training accuracy. The accuracy of training and validation both rises with increasing epoch.

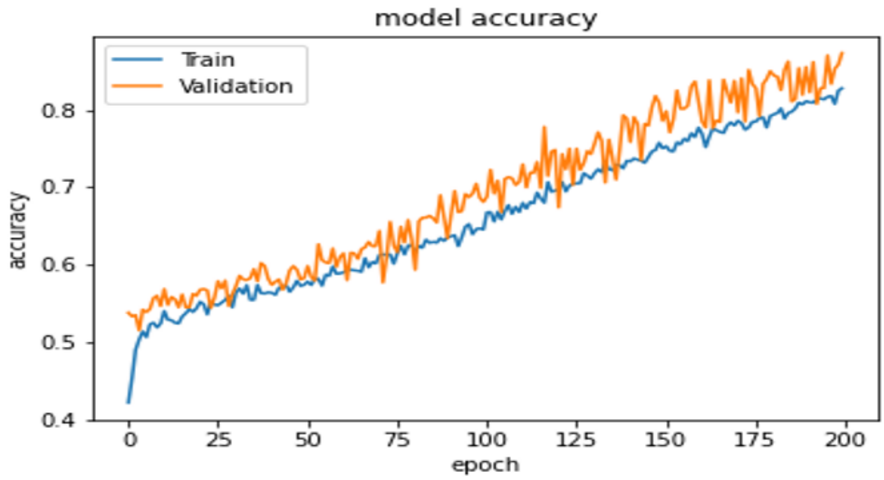

Figure 9: Accuracy Graph

And the loss curve is shown in the below Figure 10. The blue line reflects training loss, while the yellow line shows validation loss. The graph below shows that, as expected, the loss is decreasing as the epoch increases.

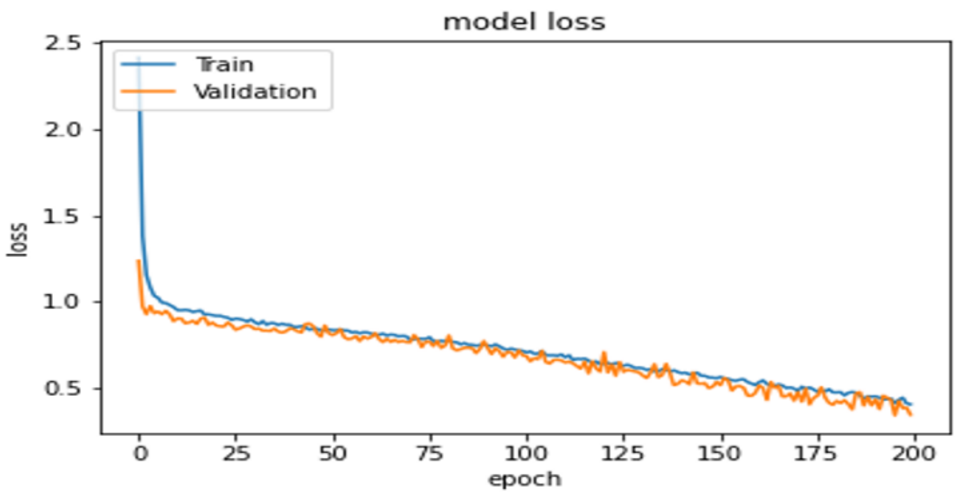

Figure 10: Loss Graph

A table called a confusion matrix is used to describe how well the classification report performed. Here in the below Figure 11 the confusion matrix shows the True predictions and false predictions for all the four classes.

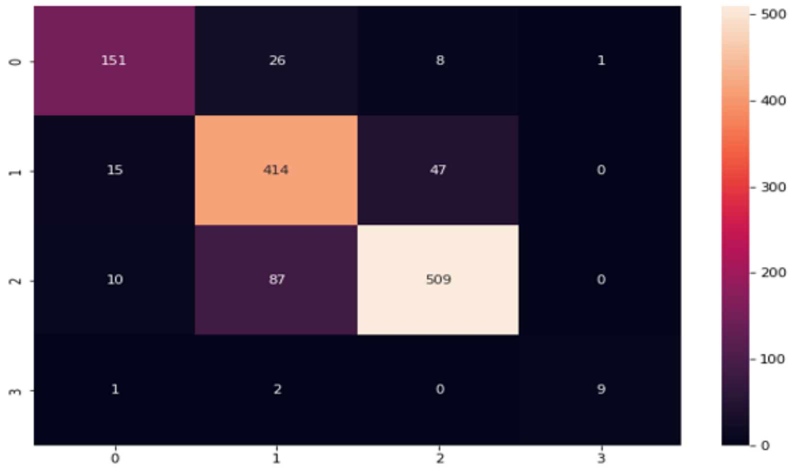

Figure 11: Confusion matrix

The Classification report of the model is given in the below Figure 12

	precision	recall	f1-score	support
Mild_Demented	0.85	0.81	0.83	186
Very_Mild_Demented	0.78	0.87	0.82	476
Non_Demented	0.90	0.84	0.87	606
Moderate_Demented	0.90	0.75	0.82	12
accuracy			0.85	1280
macro avg	0.86	0.82	0.84	1280
weighted avg	0.85	0.85	0.85	1280

Figure 12: Classification report

ViT has a promising accuracy of 87.5% with a loss of 0.34 and an F1-Score of 0.84 when compared to the other historical CNN models. According to the classification report, the model did a good job of predicting all the four classes, and the outcomes appear encouraging. Our model was trained using augmented data in comparison to previous models, which helps to prevent the overfitting problem and can improve the model's robustness and generalizability.

7 Discussion and Conclusion

From the above evaluation method, it is evident that Vision Transformers performs well and gives a good accuracy of 87.5% and a loss of 0.34. On comparing to the historical methods, the model produces less accuracy and F1 score but in this research the novel approach has been used which focuses on Image processing technique. Vision Transformer has shown a remarkable result in image classification with good accuracy. Different preprocessing technique has been used to increase the amount of data as transformer models are data hungry. The method can effectively classify the disease based on MRI images. This can help the physicians to classify the stages using the MRI image and provide related treatment to the affected person which can eventually reduce the fatal rate.

Deep learning in general but particularly for Transformer model requires a large amount of training data in order to obtain good accuracy. To meet these challenges, increasing the quantity of data is a common solution. The data augmentation technique helps us to increase the quantity of data and thus benefits in increasing the performance of the model. So far, in the research project the basic data augmentation techniques like zooming, resizing, rotating and horizontal flip were used. There are some other data augmentation techniques which can be beneficial in medical imaging that is shifting, shearing, cropping, etc. These augmentation techniques perform well in medical imaging. Secondly, to improve the performance of our model we can go with miscellaneous transforming that includes adding noise, applying Gaussians filter, increasing contrast and brightness etc. Finally, which is the most important and it focuses on the model architecture, we can use different model structure, from changing the depth of the model by increasing the number of transformer layer.

8 Future Work

A neurological condition known as Alzheimer's disease (AD), which is a more prevalent form of late dementia. It results in the death of our brain cells, which causes total memory loss as well as physical damage. This research report used a novel technique called the Vision Transformer to classify Alzheimer's disease, and the results are promising. In Future, the challenge will be to collect brain balanced and sufficient data related to Alzheimer's disease. Deep learning segmentation can be injected to segment the affected area or the region of interest only.

References

Grady, C., Sarraf, S., Saverino, C. and Campbell, K. (2016). Age differences in the functional interactions among the default, frontoparietal control, and dorsal attention networks, *Neurobiology of Aging* **41**: 159–172.

He, K., Zhang, X., Ren, S. and Sun, J. (n.d.). Deep residual learning for image recognition. **URL:** *http://image-net.org/challenges/LSVRC/2015/*

Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., Whitwell, J. L., Ward, C., Dale, A. M., Felmlee, J. P., Gunter, J. L., Hill, D. L., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C. S., Krueger, G., Ward, H. A., Metzger, G. J., Scott, K. T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J. P., Fleisher, A. S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J. and Weiner, M. W. (2008). The alzheimer's disease neuroimaging initiative (adni): Mri methods.

Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., Feng, D. and Fulham, M. J. (2015). Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer's disease, *IEEE Transactions on Biomedical Engineering* **62**: 1132–1140.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints.

Payan, A. and Montana, G. (2015). Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks.
URL: *http://arxiv.org/abs/1502.02506*

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition.
URL: *http://arxiv.org/abs/1409.1556*

Suk, H. I., Lee, S. W. and Shen, D. (2014). Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis, *NeuroImage* **101**: 569–582.

Suk, H. I., Lee, S. W. and Shen, D. (2015). Latent feature representation with stacked auto-encoder for ad/mci diagnosis, *Brain Structure and Function* **220**: 841–859.

Suk, H. I. and Shen, D. (2015). Book cover recent progress in brain and cognitive engineering pp 203–213cite as deep learning in diagnosis of brain disorders, *NeuroImage* pp. 569–582.

Suk, H.-I. and Shen, D. (n.d.). Deep learning-based feature representation for ad/mci classification.
URL: *http://www.loni.ucla.edu/ADNI*

Yin, W., Li, L. and Wu, F. X. (2022). Deep learning for brain disorder diagnosis based on fmri images, *Neurocomputing* **469**: 332–345.