AGED-ViT: A Novel Transformer-Based Framework for Diagnosis of Alzheimer's Disease by Leveraging Gene Expression Data

Albert Guo
School of Computing Science
Simon Fraser University
Burnaby, Canada
albertguo2006@gmail.com

Megan Fowler
School of Computing Science
Simon Fraser University
Burnaby, Canada
megan_fowler_2@sfu.ca

Kay C. Wiese, IEEE SM
School of Computing Science
Simon Fraser University
Burnaby, Canada
wiese@sfu.ca

***Abstract*—Alzheimer's disease (AD) is a growing global health concern and correct diagnosis is crucial for effective treatment. In this study, we present a novel method for the detection of AD using gene expression data from blood samples. We normalized and combined four publicly available Alzheimer's datasets and trained a Vision Transformer (ViT) model. This combined dataset had almost seven times more features than patient samples which can cause models to overfit on the training data. To overcome this issue, we employed Linear Discriminant Analysis (LDA) to reduce the dimensionality of the data and noise injection to encourage generalizability and robustness. We then compared our model to several state-of-the-art models that used Support Vector Machines (SVMs), Convolutional Neural Networks (CNNs), and Deep Neural Networks (DNNs). Our model, AGED-ViT, achieved an average accuracy of 88.4% and area under the curve (AUC) of 0.951 on the combined dataset, outperforming previous methods. Our results demonstrate the importance of preprocessing techniques for data with more features than samples to reduce overfitting, as well as the powerful predictive capabilities of ViTs, establishing a foundation for further exploration and optimization of the transformer architecture in the context of genomic diagnosis. This study can contribute to improving the accuracy of AD diagnosis, thus facilitating intervention and leading to a more promising outcome for patients.**

***Index Terms*—Machine Learning, Alzheimer's disease, Transformers, Gene Expression.**

I. Introduction

A. Background

Over 55 million people worldwide suffer from the effects of Alzheimer's Disease (AD) or other dementias [1] and this number is only expected to grow as the global population ages. Although some treatments are somewhat effective in slowing down the progression of AD [13], [16], they are effective only in the early stages of the disease. Early diagnosis of AD, however, remains especially challenging due to the lack of any singular metric or biomarker able to predict it effectively. Currently, medical professionals rely on questionnaires or mental tests involving counting, memory, or problem-solving to diagnose AD [18]. The use of such tests is not ideal, as they are subjective and only diagnose Alzheimer's after moderately severe symptoms develop, ultimately delaying treatments until after the effective window.

Previous studies have attempted to address this growing issue by applying a variety of statistical techniques and machine-learning algorithms to gene expression data from patient blood samples [5], [8], [9], [14]. However, these studies often run into an issue known as the "curse of dimensionality". In most genomic datasets there is a significantly higher number of features than the number of samples. This leads to lower accuracy scores due to neural network overfitting [9].

We employ two key techniques to help alleviate this tendency to overfit; Linear Discriminant Analysis (LDA) and noise injection. LDA was chosen for dimensionality reduction due to its use in a previous study for this task [9]. LDA projects high dimensional data to a lower dimension while maintaining the maximum possible variance between classes [4]. Noise injection helps reduce overfitting by adding random noise to each input sample, thus forcing the model to learn general patterns.

Transformers are a type of neural network originally developed for use in natural language processing (NLP) [17]. Recently, they emerged as an alternative to traditional convolutional neural networks (CNNs) in image recognition tasks [3]. One limitation of traditional CNNs is that they are only capable of finding relationships within each kernel, which is typically only a square of pixels comprising a small portion of the entire image. This limits their ability to perceive global relationships. The Vision Transformer (ViT) solves this issue by feeding linearly embedded image patches through a standard transformer model [3]. Due to the self-attention layer of the transformer, the ViT is able to elucidate relationships in the information contained in patches across the image. For our application, this allows us to find complex relationships between gene expression values, regardless of their location in the image. We theorize that the transformer's ability to focus on the most relevant information can significantly enhance model accuracy, allowing for more accurate diagnosis.

This study, which is the development of a short paper

presented at IEEE CIBCB 2023 [7], aims to develop a novel method for the diagnosis of Alzheimer's Disease. Our model, **AGED-ViT** (**A**lzheimer's **G**ene **E**xpression **Di**agnosis - **Vi**sion **T**ransformer) uses a vision transformer model and several pre-processing steps to predict whether a patient has Alzheimer's disease using blood gene expression data.

B. Purpose

As access to healthcare improves for the global population and the average life expectancy of the world increases, the incidence of aging-related diseases such as AD will continue to increase. The purpose of this study was to develop an accurate diagnostic method for AD using blood gene expression data. Accurate diagnosis of AD is imperative and would allow patients to receive necessary treatments early on, drastically improving their quality of life, slowing down AD progression, and reducing healthcare costs. An effective model for interpreting gene expression data could also be generalized for diagnosing other diseases including cancer, diabetes, and cardiovascular disease.

II. Materials and Methods

An overview of the model is given in Figure 1.

A. Datasets

Four publicly available and anonymized Alzheimer's datasets were used for this project: AddNeuroMed1 (ANM1) and AddNeuroMed2 (ANM2) [15], Nachun et. al (Nachun) [12], and the Alzheimer's Disease Neuroimaging Initiative (ADNI) [2]. These datasets consisted of gene expression tables extracted from blood samples from AD and healthy control patients within the same demographic. ANM1, ANM2, Nachun, and ADNI had 329, 388, 551, and 442 samples respectively, for a total of 1710 patient samples and 10678 common genes (Table I). Within the data, there are three classes; Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), and Control (CTL) (Table I). For the initial analysis, only AD and CTL samples were included.

TABLE I: Distribution of classes in the individual and combined datasets

Dataset	AD	MCI	CTL	Total
ANM1	145	80	104	**329**
ANM2	140	113	135	**388**
Nachun	198	124	229	**551**
ADNI	100	191	151	**442**
Total	**583**	**508**	**619**	**1710**

B. Pre-processing

1) Normalization: Prior to combining the datasets, the data was normalized in both dimensions to a mean of 0 and standard deviation of 1. That is, the gene expression data for each patient sample was normalized, as well as the gene expression data for each gene.

2) Data partitioning: After combining the normalized datasets, the data was split into a 20% test set and an 80% training set.

3) Pearson Correlation and Dimension Reduction: The correlational power of each gene with AD was calculated using the Pearson correlation coefficient, which measures the strength and direction of a linear relationship between two continuous variables. The formula for the Pearson correlation coefficient r is given by:

$$r = \frac{\sum(x - m_x) - (y - m_y)}{\sqrt{\sum(x - m_x)^2 \sum(y - m_y)^2}}$$

where m_x is the mean of the vector x and m_y is the mean of the vector y. The Pearson correlation coefficient ranges from -1 to 1, where an r-value of 1 indicates a perfect positive linear relationship between variables, -1 indicates a perfect negative linear relationship, and 0 indicates no linear relationship between the variables. Genes were then sorted by their mean Pearson correlation coefficient, and the dataset was reduced to only the 4,096 most discriminative genes.

4) Reshaping Data to 2D: Each patient sample was converted to a 2D image by wrapping the data row-by-row. That is, the first row contains the gene expressions for genes with the highest Pearson correlation coefficients, and the last row contains expressions for genes with the lowest Pearson correlation coefficients. The gene ordering was then randomized to prevent the introduction of biases.

5) Linear Discriminant Analysis Dimensionality Reduction: Despite reducing the number of genes down to the 4,096 most discriminative, our dataset still has more features than samples. We therefore reduce the dimension further by employing linear discriminative analysis (LDA). LDA is a supervised technique which aims to find a projection that best separates the data.

LDA was applied to the data in 256 batches, with each batch comprised of 64 genes. The genes used were selected on a rolling basis, with each batch selecting 8 new genes iteratively. This effectively allows for the dimensionality of the image to be reduced whilst simultaneously preserving the discriminative properties of each gene (Figure 2). In this way, each patient sample was reduced to a 32×32 image (Figure 3).

6) Noise Injection: To prevent overfitting of the neural network, Gaussian noise injection was used. Gaussian noise is a randomized form of noise that takes the form of a standard distribution in which the center of the curve represents the original value. For the training dataset, the standard distribution used was one with a standard deviation of 0.01, thus introducing 1% noise. In this way, noise injection was used to augment the training set ten-fold, allowing for different noisy representations of the same data.

C. Network Architecture

For the network architecture, a Vision Transformer (ViT) model was used [3]. Hyperparameter tuning of the layer count and size, patch size, and number of attention heads was performed manually on the training set. The optimal configuration found is given in Table II and a plot of the AUC for varying number of layers is given in Figure 4. The model was then retrained using the optimal hyperparameters.

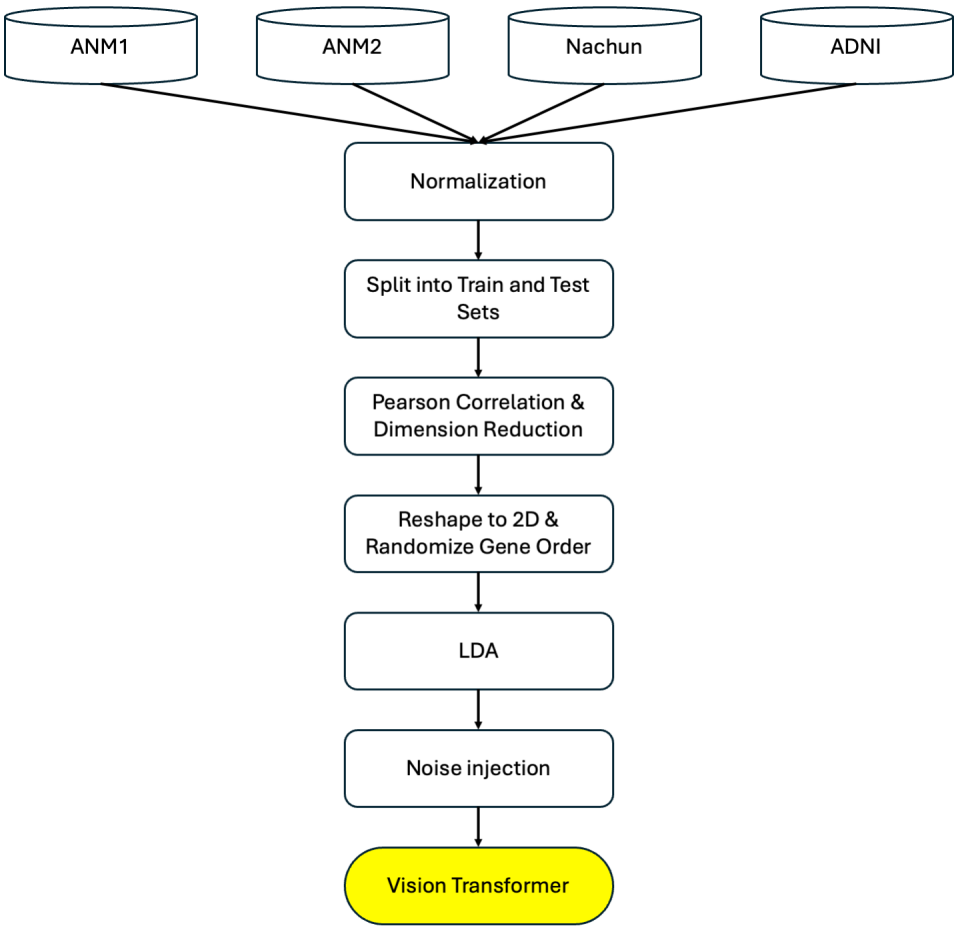

Fig. 1: Overview of the model, showing data pre-processing steps and Vision Transformer

TABLE II: ViT Model Architecture

Number of Layers	8
Layer sizes	96
Size of Patches	2
Number of Attention Heads	12

The model was implemented using a PyTorch Vision Transformer (ViT) binary classification model with 5.9M parameters. The model was trained for 10 epochs with batch size 100 using the Adam optimizer with a learning rate of 0.001. The loss criteria was L1 (mean absolute error) loss. Figure 5 shows the convergence of the model loss.

D. Evaluation Metrics

The model's performance was evaluated using two metrics: accuracy and Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) graph. The accuracy is the percentage of samples correctly predicted. The AUC is given by plotting the false positive rate against the true positive rate and serves as a robust measure of a model's performance. Notably, AUC is advantageous as it remains unaffected by imbalances in class distribution within testing sets. On the other hand, accuracy is valued for its simplicity and the ease with which it can be compared across different studies. The model was evaluated using 5-fold cross-validation.

III. Results

The dataset contained three classes, Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), and Control (CTL). We initially omitted the MCI class and only looked at the binary classification task of classifying AD vs CTL. In follow up experiments, we include the MCI class and perform pairwise binary classification. We also assess the effect of combining the MCI class with both the AD and CTL classes. Finally, we assess how well our model performs on individual datasets by re-training and evaluating our model on the ANM1, ANM2, and ADNI datasets.

A. Comparison of AD vs CTL

Our first experiment was the binary classification of AD or CTL from patients' blood gene expression data. The confusion matrix for the model evaluated on the test set is given in Table III.

AGED-ViT achieved an average accuracy of 0.883 ± 0.019, outperforming current state-of-the-art (SOTA) models, as seen

TABLE III: Confusion Matrix for AGED-ViT evaluated on the combined test set

	Predicted Positive	**Predicted Negative**
Actual Positive	0.92 ± 0.05	0.08 ± 0.05
Actual Negative	0.14 ± 0.07	0.86 ± 0.07

in Table IV. The proposed model was able to outperform Kalkan et al. [9], the previously best-performing model, as well as many other models, including DeepInSight [14].

B. Comparison of AD vs MCI vs CTL

In a follow-up experiment, we evaluate AGED-ViT's performance at discriminating between AD and MCI as well as MCI and CTL classes. We also assess the effect of combining the MCI class with the AD and CTL classes. These results were then compared to those obtained by Kalkan et al. [9] (Table V).

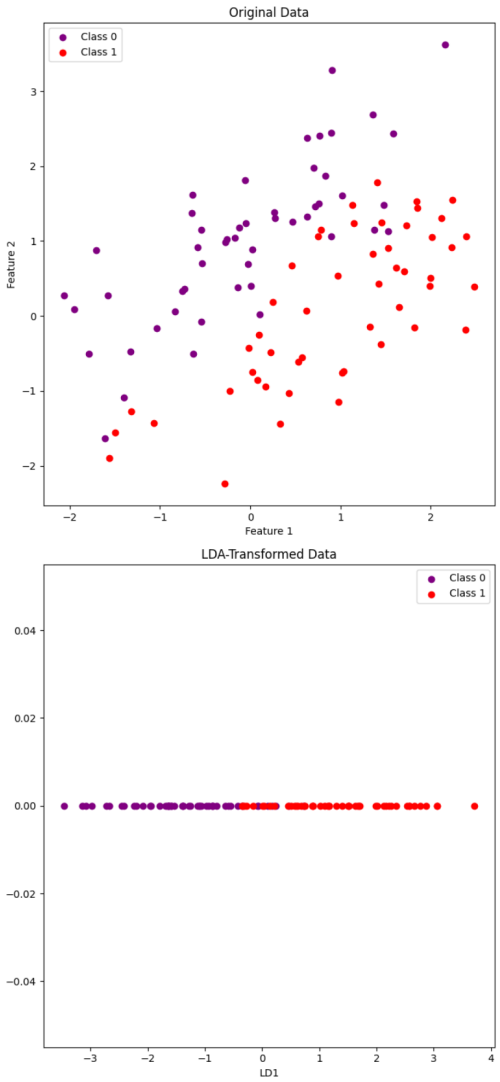

Fig. 2: Example of LDA. The multidimensional data is projected into one axis, separating the two groups with maximum variance.

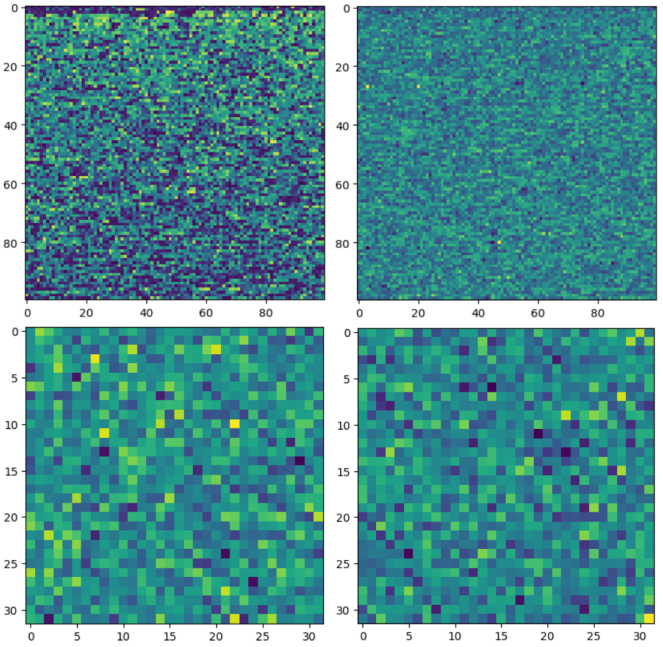

Fig. 3: Example of Model Pre-Processing. Top left: patient data before normalization. Top right: after normalization. Bottom left: After LDA. Bottom right: after noise injection. Data taken from the Alzheimer's Disease Neuroimaging Initiative (ADNI), showing a control patient.

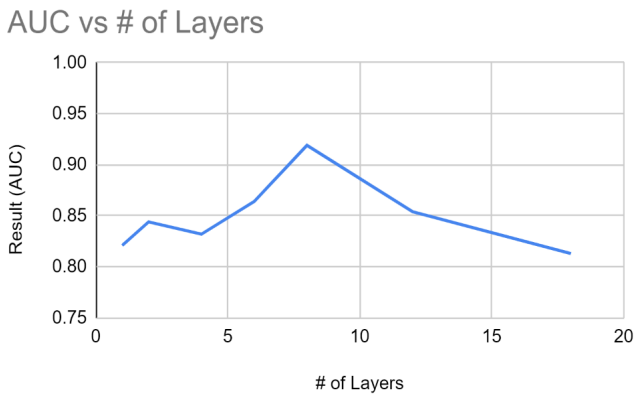

Fig. 4: Relative AUC vs the number of layers on ANM1.

Similar to Kalkan et al., our model's performance decreased when trying to differentiate between AD and MCI and MCI and CTL. This is likely due to the fact that the MCI class will share some features similar to the AD class, and some similar to the CTL class. We also found that combining the MCI class with either the AD or CTL classes leads to a larger decrease in performance, compared to the classification of AD vs CTL only. This indicates that despite being somewhat similar to both the AD and CTL classes, the MCI class is sufficiently different from both. Despite these declines in performance, AGED-ViT continues to outperform the Kalkan et al. model across all pairwise classifications (Table V).

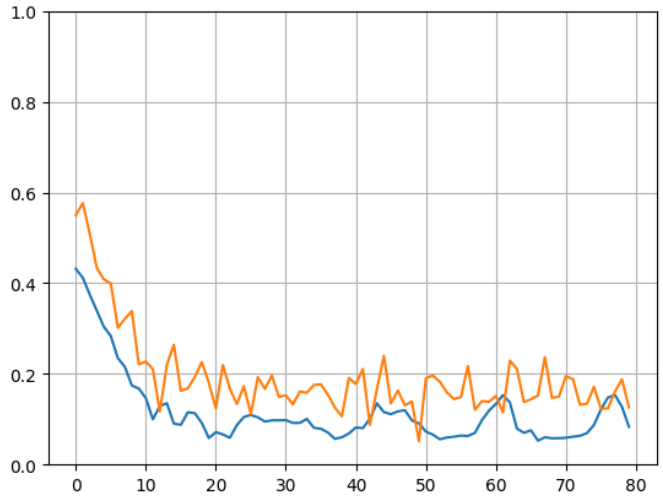

Fig. 5: Convergence of L1 loss of AGED-ViT on training (blue) and testing (orange) data. X-Axis is Epochs, Y-Axis is loss.

TABLE IV: Comparison of Machine Learning Model Performances

Study	Method	Accuracy	AUC
El-Gawady et al. (2022)	Multiple Feature Selection + SVM	0.690	0.690
Güçkıran et al. (2019)	LASSO + SVM	0.764	0.850
DeepInsight - Sharma et al. (2019)	DeepInsight (tSNE + CNN)	0.670	0.743
Kalkan et al. (2022)	LDA-based imaging + CNN	0.842	0.875
AGED-ViT	**Iterative LDA-based imaging + ViT**	**0.883 ± 0.019**	**0.950 ± 0.005**

Accuracy and AUC for other studies taken from comparison conducted by Kalkan et al. trained and tested on the ANM1, ANM2, Nachun, and ADNI datasets [9].

C. Performance Across Datasets

We then assessed how well AGED-ViT performs on three of the individual datasets. We retrained and tested the model on each of the ANM1, ANM2, and ADNI datasets. Table VI shows the AUC of AGED-ViT as well as a model by Lee & Lee [11]. The Lee & Lee study only assessed the ANM1, ANM2, and ADNI datasets, hence we only give the results for these three datasets (omitting the Nachun data). The model outperformed Lee & Lee across each of these datasets. The highest performance was achieved on ANM1, beating the previous SOTA model by 0.9%. On these datasets, the model performed better with lower layer counts and a higher dropout rate. We believe that due to the limited dataset for these experiments, these changes help prevent the model from overfitting.

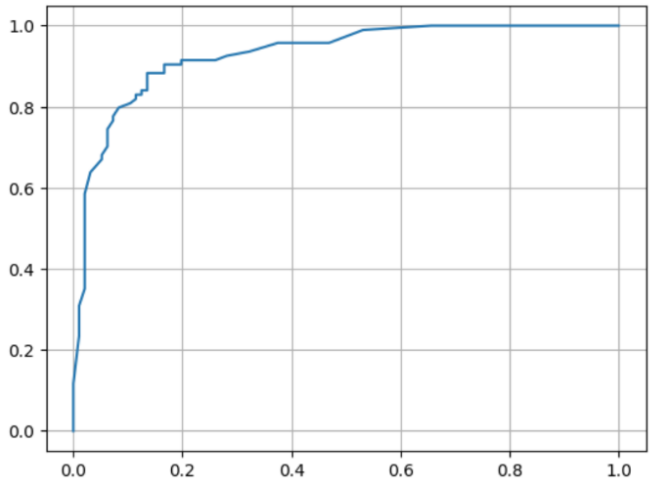

Fig. 6: AUC on combined dataset, X-axis is true positive rate, Y-axis is false positive rate.

TABLE V: Comparison of Accuracy for Alzheimer's Disease Classification Models for AD vs. MCI vs. CTL

Classification Pair	Kalkan et al. (Accuracy)	AGED-ViT (Accuracy)
AD vs CTL	0.842	**0.883**
MCI vs CTL	0.698	**0.833**
MCI vs AD	0.704	**0.848**
AD vs. (MCI and CTL)	0.707	**0.793**
(AD and MCI) vs. CTL	0.773	**0.815**

D. Ablation Study

An ablation study was conducted on the final model, as illustrated in Figure 7, to assess the impact of individual network components. We systematically removed each of LDA, noise injection, and normalization, and assessed the model's performance. We observed that the absence of LDA and noise injection resulted in a large drop in both accuracy and AUC. The drop in AUC indicates that the positive AD class was not as easily separable from the CTL class as previously, resulting in a higher number of false positives. Additionally, we saw a large decrease in the accuracy as well, we can assume that this is largely due to CTL samples being incorrectly classified as AD. Since the data was imbalanced and contained a higher number of CTL samples than AD samples, this is likely the result of the model overfitting. LDA

TABLE VI: Comparison of AUC and Accuracy for Different Datasets

Dataset	Lee & Lee (AUC)	AGED-ViT (AUC)	AGED-ViT (Accuracy)
ANM1	0.874	**0.882 ± 0.009**	0.869 ± 0.011
ANM2	0.804	**0.805 ± 0.014**	0.742 ± 0.008
ADNI	0.657	**0.669 ± 0.002**	0.703 ± 0.0

Reported AUC scores for Lee & Lee models are for a DNN for ANM1 and ADNI and a SVM for ANM2.

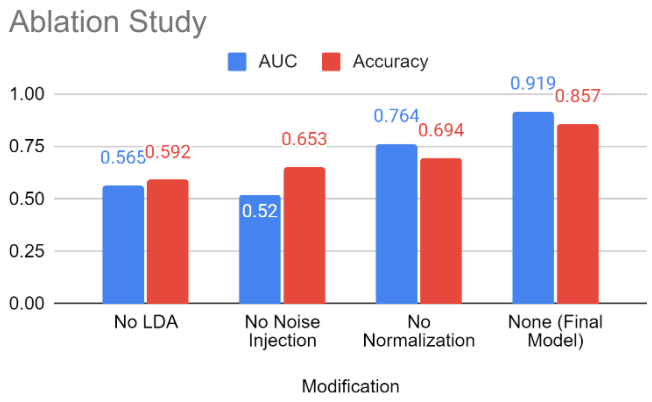

Fig. 7: Ablation study showing the effect of removing various components of the AGED-ViT model.

reduces the dimensionality while noise injection adds a degree of uncertainty to each expression value, and hence forces the model to learn more general patterns.

When we omitted normalization, we saw only a slight drop in AUC but a larger decrease in accuracy. The high AUC indicates that the model is still able to maintain a low count of false positives, however, the low accuracy tells us that this is at the expense of an increase in false negatives. This likely suggests that the data has a few genes for which the expression value is extremely low or high, to which the model is then attributing too much weight to. Normalizing the data helps ensure that the model learns the impact of each gene equally.

In conclusion, the incorporation of all three components, LDA, noise injection, and normalization, was deemed essential for optimizing the model's performance.

IV. Analysis

The Vision Transformer model is structured to process input images, or, by analogy, structured data arrays representing gene expressions, into a series of patches. These patches are then embedded into a high-dimensional space, where the transformer architecture captures complex patterns and relationships among the data points. Specifically, the model's key features include:

1) Patch Embedding: Converts each image patch into a high-dimensional vector, preparing it for processing by the transformer. This step is critical for understanding local features within the broader context of the entire dataset.

2) Transformer Architecture: Utilizes self-attention mechanisms to independently weigh different parts of the data, allowing the model to focus on the most informative features for diagnosing Alzheimer's disease. This adaptability is a significant advantage over more traditional analysis techniques that treat all data points equally.

3) Output Layer: Translates the transformer's complex representations into a final prediction. This step is crucial for making the model's insights actionable, providing a direct link between the data and a tangible diagnostic outcome.

We theorize that this model architecture directly impacts performance through its ability to process and analyze data at multiple scales. By focusing on both local and global features within the data, the Vision Transformer can identify subtle patterns that might be missed by models that analyze the data in a more uniform manner. Moreover, the use of self-attention allows for dynamic adaptation to the most informative features, which is particularly beneficial for the complex and varied nature of gene expression data associated with Alzheimer's disease.

The specific parts of the model contributing to its effectiveness in diagnosing Alzheimer's disease include the embedding layer, which ensures that data is properly pre-processed and represented; the transformer mechanism, which dynamically adapts to focus on the most relevant features of the data; and the output layer, which translates complex patterns into a clear diagnosis outcome.

V. Discussion

A. Implications of Findings

The application of the Vision Transformer model to Alzheimer's disease diagnosis represents a significant advancement in computational biology. By adapting an architecture originally designed for visual data to interpret complex gene expression patterns, our research demonstrates the flexibility and power of transformer models. This approach not only enhances our ability to diagnose Alzheimer's more accurately but also opens the door to novel applications of transformer models in understanding other complex diseases. The ability of the ViT model to dynamically focus on the most informative features of the data could revolutionize how we approach the analysis of biological datasets, moving towards a more nuanced understanding of disease mechanisms.

B. Potential Limitations and Future Work

While the Vision Transformer model shows promising results, it's important to acknowledge potential limitations. The model's performance is highly dependent on the quality and quantity of training data. As mentioned in the Visual Transformer paper, transformers do not generalize well when trained on insufficient data [3]. Additionally, biases in dataset composition can lead to skewed interpretations and potentially limit the model's applicability across different populations.

To mitigate these risks, future research should focus on diversifying and enlarging data sources and implementing techniques like data augmentation to ensure robust model performance. Additionally, the complexity of transformer models can lead to challenges in interpretability. Developing methodologies for model explainability will be crucial for translating these advanced computational analyses into actionable insights for clinicians.

Improvements can also be made to the pre-processing pipeline. Our model employs LDA for dimensionality reduction, however, there are other techniques that could be explored such as Uniform Manifold Approximation and Projection (UMAP) or Heteroscedastic Discriminant Analysis (HDA).

UMAP is a nonlinear dimensionality reduction technique and thus may be able to capture more complex relationships between the genes, while HDA negates the assumption of equal sample covariance within each class. Using these more powerful techniques may lead to an improvement in model accuracy and AUC.

VI. Conclusion

This study demonstrates the potential for employing Vision Transformers with a robust pre-processing pipeline to detect Alzheimer's disease using gene expression data. AGED-ViT outperforms several existing approaches which use SVMs [5], [8], CNNs [9], [14], and DNNs [11], showing promise for accurate detection and improved patient outcomes. Given the globally aging population and ever-increasing rates of Alzheimer's, effective ways to combat AD have never been of greater importance. With innovative treatments entering the market, doctors need an effective and efficient way to diagnose AD. With this powerful tool, healthcare providers would be able to diagnose AD using only a simple blood sample from a patient and an inexpensive gene expression bead chip. This work also provides a useful transformer framework for the diagnosis and management of other diseases through the novel approach to interpreting gene expression data. This could help in the diagnosis and treatment of other diseases more generally. As gene expression data is very hard to interpret by humans, a sophisticated transformer-based approach has the potential to open up new frontiers in medicine by allowing for greater diagnostic accuracy and the ability to process and understand more data than possible with traditional methods.

References

[1] "2020 Alzheimer's Disease Facts and Figures," *Alzheimer's & Dementia*, vol. 16, no. 3, pp. 391–460, 2020. [Online]. Available: https://doi.org/10.1002/alz.12068.

[2] Alzheimer's Disease Neuroimaging Initiative, "ADNI," Alzheimer's Disease Neuroimaging Initiative, 2017. [Online]. Available: https://adni.loni.usc.edu/.

[3] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv.org*, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2010.11929.

[4] J. L. S. Díaz, "Linear discriminant analysis (LDA)," *pyDML 0.0.1 documentation*, 2018. [Online]. Available: https://pydml.readthedocs.io/en/latest/dml.lda.html.

[5] A. El-Gawady, M. A. Makhlouf, B. B. S. Tawfik, and H. Nassar, "Machine learning framework for the prediction of Alzheimer's disease using gene expression data based on efficient gene selection," *Symmetry*, vol. 14, no. 3, 491, 2022. [Online]. Available: https://doi.org/10.3390/sym14030491.

[6] A. Guo, "Albertguo2006/2023-science-fair," GitHub, Dec. 3, 2022. [Online]. Available: https://github.com/albertguo2006/2023-science-fair.

[7] A. Guo, M. Fowler, and K. C. Wiese (2023). "Alzheimer's in 16x16 Words: A Novel Transformer-based Framework for the Interpretation of Gene Expression Data." *IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology.* [Online]. Available: https://cmte.ieee.org/cis-bbtc/wp-content/uploads/sites/172/IEEE_CIBCB_2023_paper_3984.pdf

[8] K. Güçkıran, İ. Cantürk, and L. Özyılmaz, "DNA Microarray Gene Expression Data Classification Using SVM, MLP, and RF with Feature Selection Methods Relief and LASSO," *Journal of Natural and Applied Sciences*, pp. 115–121, 2019. [Online]. Available: https://doi.org/10.19113/sdufenbed.453462.

[9] H. Kalkan, U. M. Akkaya, G. Inal-Gültekin, and A. M. Sanchez-Perez, "Prediction of Alzheimer's disease by a novel image-based representation of gene expression," *Genes*, vol. 13, no. 8, 1406, 2022. [Online]. Available: https://doi.org/10.3390/genes13081406.

[10] L. Lannfelt, C. Möller, H. Basun, G. Osswald, D. Sehlin, A. Satlin, V. Logovinsky, and P. Gellerfors, "Perspectives on future Alzheimer therapies: Amyloid-B protofibrils - A new target for immunotherapy with BAN2401 in Alzheimer's disease," *Alzheimer's Research & Therapy*, vol. 6, no. 2, 16, 2014. [Online]. Available: https://doi.org/10.1186/alzrt246.

[11] T. Lee and H. Lee, "Prediction of Alzheimer's Disease Using Blood Gene Expression Data," *Scientific Reports*, vol. 10, no. 1, 2020. [Online]. Available: https://doi.org/10.1038/s41598-020-60595-1.

[12] D. Nachun, E. M. Ramos, A. Karydas, D. Dokuru, F. Gao, Z. Yang, V. Van Berlo, R. Sears, J. Kramer, A. L. Boxer, H. Rosen, B. L. Miller, and G. Coppola, "Systems-level analysis of peripheral blood gene expression in dementia patients reveals an innate immune response shared across multiple disorders," *BioRxiv*, 2019. [Online]. Available: https://doi.org/10.1101/2019.12.13.875112.

[13] J. Sevigny et al., "The antibody aducanumab reduces AB plaques in Alzheimer's disease," *Nature*, vol. 537, no. 7618, pp. 50–56, 2016. [Online]. Available: https://doi.org/10.1038/nature19323.

[14] A. Sharma, E. Vans, D. Shigemizu, K. A. Boroevich, and T. Tsunoda, "DeepInsight: A methodology to transform a non-image data to an image for Convolution Neural Network Architecture," *Nature Scientific Reports*, vol. 9, no. 1, 2019. [Online]. Available: https://doi.org/10.1038/s41598-019-47765-6.

[15] S. Sood et al., "A novel multi-tissue RNA diagnostic of healthy ageing relates to cognitive health status," *Genome Biology*, vol. 16, no. 1, 2015. [Online]. Available: https://doi.org/10.1186/s13059-015-0750-x.

[16] C. H. Van Dyck et al., "Lecanemab in early Alzheimer's disease," *New England Journal of Medicine*, 2022. [Online]. Available: https://doi.org/10.1056/nejmoa2212948.

[17] A. Vaswani et al., "Attention is all you need," *arXiv.org*, 2017. [Online]. Available: https://doi.org/10.48550/arXiv.1706.03762.

[18] J. Weller and A. Budson, "Current understanding of Alzheimer's disease diagnosis and treatment," *F1000Research*, vol. 7, 1161, 2018. [Online]. Available: https://doi.org/10.12688/f1000research.14506.1.

[19] World Health Organization, "Dementia," WHO, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dementia.