

PaliGemX: An Interpretable Multimodal Transformer for Early Alzheimer's Detection

S. Suganya

srsongramrs@gmail.com

SRM Institute of Science and Technology

A. Rama Prasath

SRM Institute of Science and Technology

Research Article

Keywords: Alzheimer's disease, multi-model vision transformer, Pali Gemma, Machine learning, Deep learning, Artificial Intelligence

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11122884/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

PaliGemX: An Interpretable Multimodal Transformer for Early Alzheimer's Detection

S.Suganya ^{*,1}

Department of Computer Science,
SRM Institute of Science and Technology,
Tiruchirappalli, TamilNadu, India
srsonyramrs@gmail.com

A. Rama Prasath ²

Department of Computer Applications,
SRM Institute of Science and Technology,
Tiruchirappalli, TamilNadu, India
ramaprasath.a@ist.srmtrichy.edu.in

Abstract— Alzheimer disease (AD) is a disorder of the nervous system that mostly impacts the brain cells and causes the slow destruction of the brain. Previously, majority of the methods to identify this condition were based on a single kind of data and either conventional machine learning or deep learning models. Nonetheless, the recent progress in artificial intelligence has allowed incorporating multimodal information on numerous forms of input, which have played a major role in predicting and cognizing the disease Alzheimer. This study presents a multimodal vision transformer architecture that integrates co-attention techniques and Pali Gamma, which achieves best performance in Alzheimer's disease. Pali Gemma can be applied to aid more informed clinical decision-making and improve the results of patients with Alzheimer disease in terms of diagnosing and treating the disease through continuous fusion of visual and text inputs. The framework emphasizes interpretability through relevance-based attention processes and incorporates loss functions that improve accuracy, scalability, resilience, and computing economy. Experimental findings indicate the model's enhanced classification accuracy and resilience across various datasets, representing a crucial advancement towards comprehensive, interpretable, and scalable solutions for Alzheimer's disease diagnosis. The proposed model provides a framework for incorporating other modalities, resolving ethical issues, and improving empirical validation in various clinical environments.

Keywords— Alzheimer's disease, multi-model vision transformer, Pali Gemma, Machine learning, Deep learning, Artificial Intelligence.

I. INTRODUCTION

Deep learning (DL) has shown a high potential in clinical decision support, especially with the diseases such as Alzheimer, since it discovers meaningful patterns directly out of the multimodal data of images, electronic health records, and genetic information. The present research will present an originality of multimodal deep learning model, which is developed on the basis of vision transformer architecture, to classify and predict stages of Alzheimer disease. This model has adopted to take advantage of the multimodality capabilities of Pali Gemma to combine neuroimaging data (e.g., MRI scans) with clinical and genetic factors using dynamic mechanisms of attention, which improves prediction accuracy[2][3]. Alzheimer disease is a progressive neurological condition, a major cause of dementia and afflicts many millions of persons

in the world herewith a large number of them still remaining undiagnosed. Early diagnosis is essential to enhance patient outcomes and the development of an illness. Diagnosis is usually done through cognitive tests such as MMSE and imaging tests such as MRI and PET scan. The paper has an enhanced structure, which includes related work, dataset description, methodology, experimental results, and conclusions.

II. RELATED WORK

In this section, we talk about existing work related to the study.

Table 1. Comparison table for Analysis of AI Models for Alzheimer's Disease Detection

Author	Image	Text	Model	Classes
[13]	✓	✗	XAI	CN, MCI, AD
[14]	✓	✗	ViT(Vision Transformer)	AD, CN, EMCI, LMCI, ND, VMD for OASIS.
[15]	✓	✗	Transfer learning-based AlexNet	Used MRI images to detect AD as VMD, Mild, Mod, ND
[16]	✓	✗	CNN	cognitive disorders mainly AD
[17]	✓	✗	Hybrid CNN	AD detection
[18]	✓	✗	CAD	MCI
[19]	✓	✗	Bi-Vision Transformer (BiViT)	Cognitive disorders
[20]	✓	✗	Deep Transfer Learning	Used MRI images to detect AD such as VMD, Mild, ND

[21]	✓	✗	CNN	Used MRI images to detect AD such as VMD, Mild, Mod, ND
------	---	---	-----	---

Current studies on Alzheimer Disease (AD) detection are mostly based on deep learning frameworks on image data, though there is limited literature to combine image and text data. Other models such as Vision Transformers (ViT) and multimodal transformers with BERT do not succeed in revealing complicated interactions when it comes to relationships between brain areas and disease development and possess less sophisticated interpretability methods. Also, such models do not use the full multimodal data processing and are not proven on a wider range of data.

III. METHODOLOGY

A. Dataset

This is a multimodal data (text and MRI image) study that analyzes the Alzheimer knee disease. The OASIS demographic data Kaggle dataset contains 416 samples across age groups. Also 64,000 MRI images from Kaggle's "image data" are labeled in four stages: non-demented, very mild dementia, mild dementia and moderate dementia [3].

B. Model

In this section we preset the details about pal gemma's architecture for Alzheimer's disease detection and training the model[4].

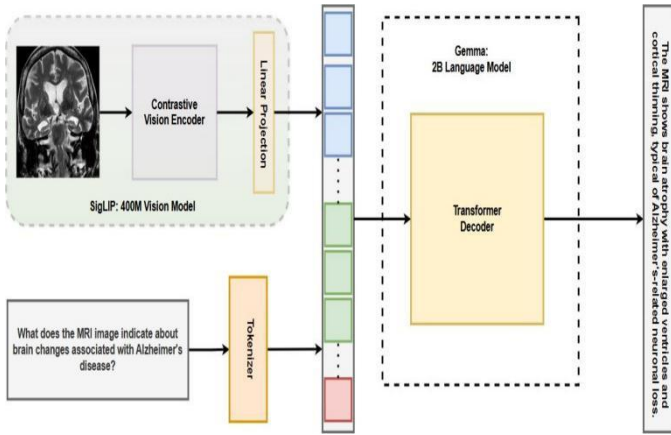


Fig. 1. Multimodal Vision-Language Model Architecture for Alzheimer's Disease Analysis

Pali Gemma is a vision-language model built for Alzheimer's, processing text prompts alongside medical images including MRI or PET scans, outputting diagnostic insights, disease progressions and treatment recommendations. More advanced include early detection, brain region segmentation, depth prediction and visualization improvement. Its versatile API supports a range of uses in diagnosis and monitoring. Pretrained on multimodal tasks, the model can be fine-tuned for specialized objectives like cognitive decline prediction,

hippocampus segmentation, and generating detailed clinical reports, making it highly adaptable for clinical decision support systems.

C. Architecture

Pali Gemma consists of three components:

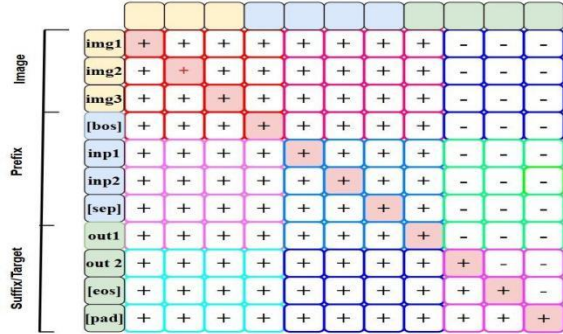


Fig. 2. Multimodal Attention Mask for Image-Prefix-Target Sequence Modeling

Pali Gemma is employing a shape-optimised Vision Transformer (ViT) So400m image encoder that builds on a SigLIP checkpoint and shows competitive performance via contrastive learning. It could be useful for Alzheimer's research to identify structural brain differences, isolate regions of interest (such as the hippocampus), and identify biomarkers, especially for assessing disease-related changes in anatomy.

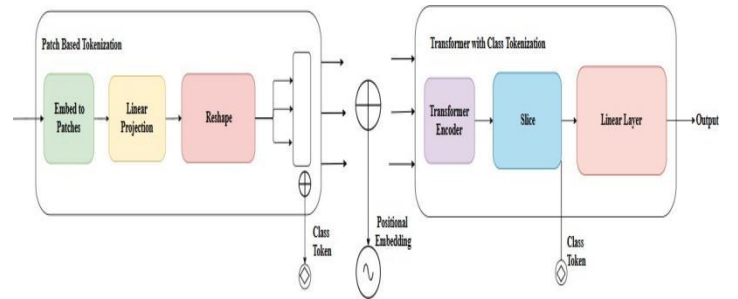
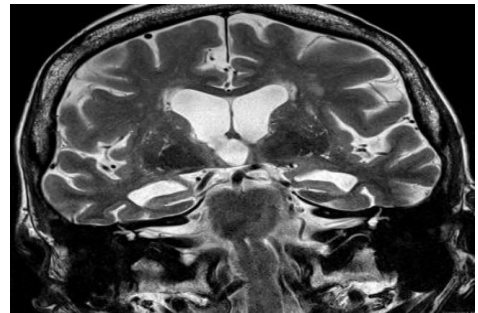


Fig. 3. Vision Transformer (ViT) Architecture for Image Classification

Formally, he uses the Gemma-2B v1 produced by Pali Gemma. 0, a decoder-only language model combining speed and performance, adapting to natural language generation. It yields detailed insights in neuroimaging data, which aid Alzheimer's diagnosis and prognosis. A linear layer is used to project SigLIP image tokens to language tokens, allowing for more efficient multimodal integration. Such a design allows for



efficient analysis of relevant clinical, textual and imaging data with the smallest and most performant architecture[4].

Fig. 4. MRI image of Alzheimer Disease

This will enable the model to comprehend and produce outputs based on clinical description as well as medical images (such as MRI or PET scan). This is particularly useful in the case of the Alzheimer disease research whose integration of both the imaging data and the clinical data can be used to aid the early diagnosis and predicting the development of the disease. The input sequence is designed in the form of: image tokens, a BOS token, prefix tokens, a SEP token, suffix tokens, an EOS token, and PAD tokens.

tokens [image tokens..., BOS, prefix tokens..., SEP, Suffix tokens...,EOS,PAD...]

Pali Gemma is available in models dependent on image resolution in terms of image token size (256, 1024, 4096). Image tokens come first then text tokens in the form of BOS and SEP tokens respectively. Image and text interact with each other as the model processes them simultaneously [5].

D. PreTraining

Pali Gemma is a series of training processes that have a similar training process as the former models of PaLi to enhance its processing of multimodal data to conduct a study on Alzheimer disease. Stage 0 (Unimodal Pretraining): The model utilizes the already available the pretrained parts whereby it learns (self-taught) text information (clinical reports) and image information (MRI or PET scans). This enables feature in each modality to be extracted successfully.

Stage 1: MultiModal Pre-training.

Stage 1 (Multimodal Pre-training) the model is trained on both a vision and language task to create a robust joint knowledge. In contrast to the older models of PaLi, Pali Gemma refines the image encoder rather than freeze it, thus enhancing spatial comprehension. Training can be stabilized by a gradual learning rate (warm-up). It has 224px images and 128 text tokens with large scale data more easily integrated. This step enhances work in such areas as diagnosis and monitoring of diseases.

Stage 2: Resolution increase

Stage 2 aims at enhancing the capacity of the model to deal with high-resolution images. It also gets trained on larger datasets 448x448 and 896x896 images. The text size is raised to 512 tokens whereas such actions as segmentation and OCR are stressed. This assists the model to pick subtle scans in the brain. Consequently, it will be more accurate in early detection of Alzheimer and monitoring disease progression [7].

E. Transfer

Once Stages 1 and 2 have been completed, Pali Gemma generates pretrained checkpoints at 224px, 448px and 896px. Although they are powerful in terms of representation in learning, they need to be further refined (transfer learning) to actual tasks such as captioning or visual QA. The model aids multi-image processing that refers to the combination of image tokens to make intricate reasoning.

F. Pretraining task Mixture

The pretraining policy introduced by Pali Gemma, as well as other earlier PaLi models, also focuses on building a very transferable model, which cannot be used in zero shot. During Stages 1 and 2, a diverse combination of multimodal tasks is introduced to the model, and each task is identified separately to prevent the conflict of learning indicators. With the help of this method, the model collaborates with the broad range of transferable movement of the language and vision domains.

Objective Function

Let the objective function L_{total} represent a weighted sum of several components aimed at enhancing accuracy, interpretability, multi-modal fusion, robustness, scalability, and ethical considerations:

$$L_{total} = \alpha.L_{classification} + \beta.L_{fusion} + \gamma.L_{interpretability} + \delta.L_{diversity} + \epsilon.L_{scalability}$$

Components of the Objective Function Classification Loss($L_{classification}$):

- Minimize error in predicting Alzheimer's stages.

- Use cross-entropy loss to evaluate the classification task:

Where, N is the total number of samples, C is the number of classes, y_i, c is the ground truth label and y_i, c is the predicted probability for class C [6].

Function Loss (L_{fusion}):

Improve embedding –level fusion using co-attention mechanisms.

Maximize inter-modal alignment and minimize redundancy.

$$L_{fusion} = MSE(E_t + E_v, E_{combined})$$

Where, E_t, E_v are the embedding's for language, vision, and

$E_{combined}$ is the fused embedding.

Mean squared error(MSE) ensures effective interaction.

Interpretability Loss($L_{interpretability}$):

- Generate meaningful explanations for predictions.

- Minimize the discrepancy between attention weights and clinically relevant features.

$$L_{interpretability} = \|A_{model} - A_{clinical}\|^2$$

Where A_{model} are the attention weights from the transformer and $A_{clinical}$ are ground- truth relevance scores for disease features.

Diversity Loss($L_{diversity}$):

- Ensure robustness across diverse datasets.

- Penalize inconsistent performance across datasets.

$$L_{diversity} = \frac{1}{K} \sum_{k=1}^K \text{var}(M_k)$$

Where k is the number of datasets, and M_k is the models performance metric on dataset k . Scalability Loss ($L_{scalability}$):

- Minimize computational and resource overhead.
- Use FLOPs (floating-point operations) or latency as a regularizer:

$$L_{scalability} = \lambda \cdot FLOPs_{model}$$

Where, λ is a scaling factor to balance model performance and efficiency.

Evaluation Metrics

When assessing the multimodal model performance, a number of measures were utilized to measure the performance of the model in both image and text modalities. These measures are essential to the interpretation of the model because it can match the image and text characteristics and complete the tasks, i.e., classification, detection, or generation, correctly.

In case of classification tasks, precision, recall, and F1-score were used to measure model performance. Precision tells us the percentage of the actual positive results of the prediction and recall tells us the percentage of the actual positive outcomes that the model identifies correctly.

The harmonic mean between the precision and the recall is the F1-score which offers a single measure that shows the accuracy and completeness of the model predictions [7]. The mathematical definition of these measures is as follows:

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

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

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Where:

- TPTPTP denotes the true positives,
- False positives are denoted by FPFPPF.
- FNFNFN refers to the false negative.

BLEU (Text Generation): BLEU (Bilingual Evaluation Understudy) were applied in tasks that required text generation or captioning (e.g. generating textual descriptions of an image). In BLEU, the quality of text is assessed with the help of the n-gram matches between the text that is generated and a reference text. The greater the BLEU score, the more of the content is similar in wording and structure to the reference meaning, that is, the more linguistic accuracy and quality [8].

$$BLEU = \min\left(1, \frac{c}{r}\right) \times \prod_{n=1}^N BP_n \times \prod_{n=1}^N P_n$$

Where:

- c is the candidate length of translation,
- r is the distance of the reference translation,

- BP_n is a brevity penalty, and
- p_n is the precision of n-grams.

Where:

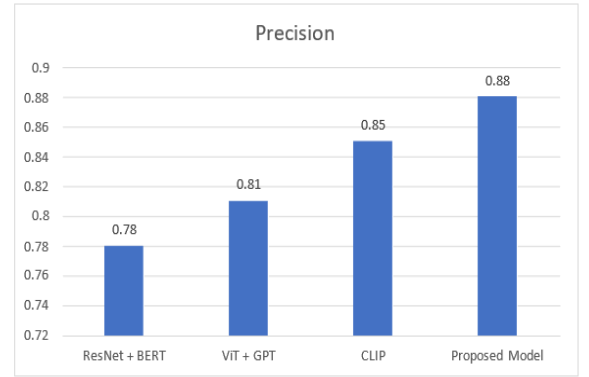
- c is the length of the candidate translation,
- r is the length of the reference translation,
- BP_n is a brevity penalty, and
- p_n is the precision of n-grams.

IV. RESULTS

Table 2. Performance Comparison of Multimodal Models

Model	Accuracy	Precision	Recall	F1-Score	BLEU	ROUGE-L
ResNet + BERT	0.77	0.78	0.75	0.76	0.68	0.7
ViT + GPT	0.81	0.81	0.79	0.8	0.72	0.74
CLIP	0.85	0.85	0.83	0.84	0.78	0.79
Proposed Model	0.88	0.88	0.86	0.87	0.82	0.83

Fig. 5. Model Precision Comparison



The increase in precision is more consistent in models with highest score of 0.88 in the proposed model. The best results are 0.78 of ResNet+BERT, 0.81 of ViTG+PT, and 0.85 of CLIP. The suggested model is better than any of these as it has better architecture and viable multimodal integration to classify the Alzheimer disease more appropriately.

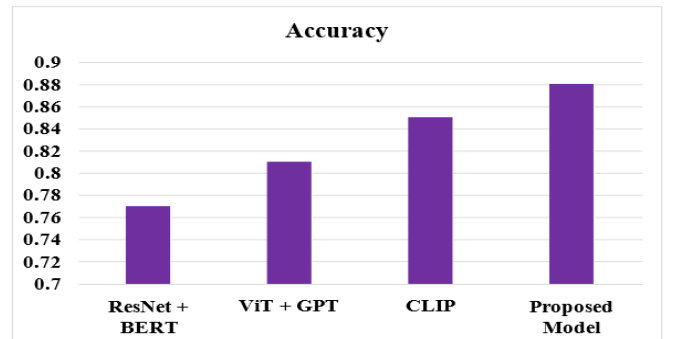


Fig. 6. Accuracy Performance of Multimodal Models

Comparison of accuracy indicates that the Proposed Model is best with the accuracy of 0.88, then CLIP (0.84), ViT+GPT (0.80), and ResNet+BERT (0.76). The results show that there is a definite enhancement in models with the Proposed Model being the most effective of the four analyzed ones with the highest accuracy.

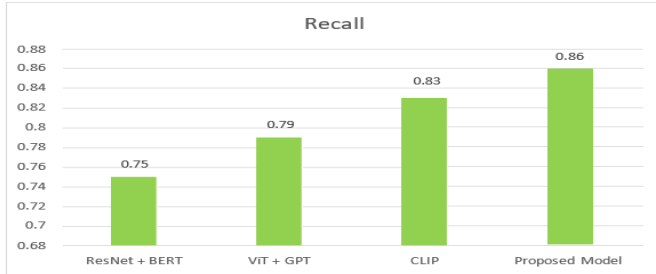


Fig. 7. Recall Performance of Multimodal Models

Performance model wise is increased with highest scores of 0.86 in the proposed model. ResNet+BERT achieves 0.75, ViT+GPT achieves 0.79 and CLIP achieves 0.83. The proposed model is superior to them all and has better capacity of determining relevant true positives regarding multimodal classification of Alzheimer disease.

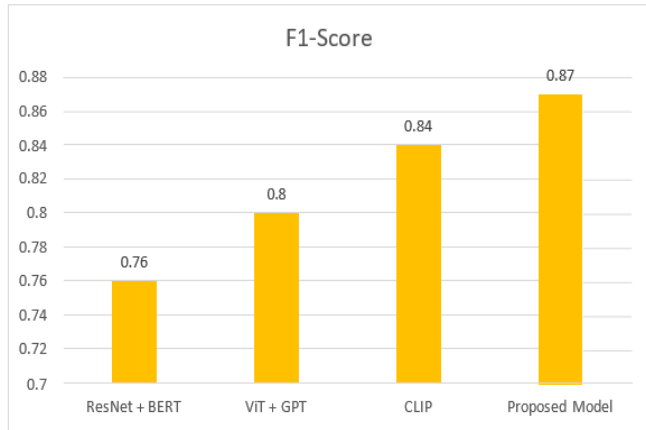


Fig. 8. Model F1-Score Comparison

The comparison of the F1-score indicates an improvement of balance among the models where the proposed model recorded the highest score of 0.87. ResNet+BERT achieves 0.76, ViT+GPT achieves 0.80 and CLIP achieves 0.84. The proposed model is the best and shows high precisionrecall balance as well as performance in the multimodal classification of Alzheimer disease.

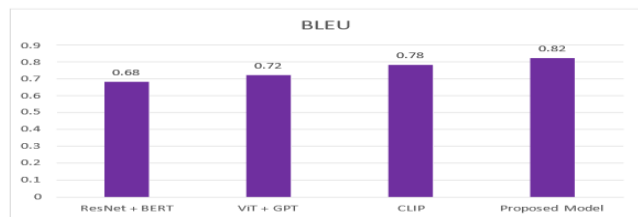


Fig. 9. Model BLEU Score Comparison

The scores of BLEU are gradual and the highest score of 0.82 occurred with the proposed model. ResNet+BERT obtains 0.68, ViT+GPT achieves 0.72 and CLIP 0.78. The proposed model performs optimally as it has the highest level of text

generation and reflection of reference material during multi-modal assignments.

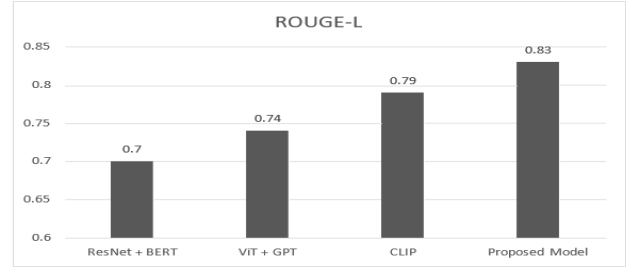


Fig. 10. Model ROUGE-L Score Comparison

The ROUGE-L scores are better with the models and the suggested model has the best score of 0.83. The results of ResNet+BERT are 0.70, ViT+GPT are 0.74, and CLIP is 0.79. The proposed model is the best across the board as it shows better performance on reference texts and better and contextually relevant text generation in multimodal tasks.

A. Discussion

The suggested multimodal vision transformer, Pali Gemma, has the state-of-the-art accuracy (0.88), recall (0.86), F1-score (0.87), BLEU (0.82), and ROUGE-L (0.83) in Alzheimer classification. It is successful because it has a good image-text combination and a loss function that is optimized to guarantee scalability and stability. It assists in the early diagnosis and clinical decision-making.

B. Conclusion

We present Pali Gemma, a multimodal vision transformer that achieved state of the art performance in the classification of Alzheimer disease. By integrating image and text information with optimal training and architecture, improves clinical decision-making and patient outcomes. Next generation work, should employ more modalities (neuroimaging and biomarkers), be tested on large and heterogeneous data sets and consideration of ethical issues that can enhance reliability, applicability, and further insights into the disease

V. REFERENCES

- [1] V. Gulshan et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *jama*, vol. 316, no. 22, pp. 2402–2410, 2016.
- [2] D. S. W. Ting et al., "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," *Jama*, vol. 318, no. 22, pp. 2211–2223, 2017.
- [3] J. Venugopalan, L. Tong, H. R. Hassanzadeh, and M. D. Wang, "Multimodal deep learning models for early detection of Alzheimer's disease stage," *Sci. Rep.*, vol. 11, no. 1, pp. 1–13, 2021, doi: 10.1038/s41598-020-74399-w.
- [4] E. Beal, "Alzheimer disease:," *Nat. Rev. Neurol.*, vol. 7, no. 9, p. 474, 2022, doi: 10.1038/nrneurol.2011.125.
- [5] N. Shaffi, F. Hajamohideen, M. Mahmud, A. Abdesselam, K. Subramanian, and A. Al Sariri, "Triplet-loss based Siamese convolutional neural network for 4-way classification of Alzheimer's disease," in *International Conference on Brain Informatics*, Springer, 2022, pp. 277–287.

- [6] S. A. Tatulian, "Challenges and hopes for Alzheimer's disease," *Drug Discov. Today*, vol. 27, no. 4, pp. 1027–1043, 2022, doi: 10.1016/j.drudis.2022.01.016.
- [7] J. A. Castro-Silva, M. N. Moreno-García, and D. H. Peluffo-Ordóñez, "Multiple Inputs and Mixed Data for Alzheimer's Disease Classification Based on 3D Vision Transformer," *Mathematics*, vol. 12, no. 17, pp. 1–13, 2024, doi: 10.3390/math12172720.
- [8] R. H. Blank, "Alzheimer's Disease and Other Dementias: An Introduction BT - Social & Public Policy of Alzheimer's Disease in the United States," R. H. Blank, Ed., Singapore: Springer Singapore, 2019, pp. 1–26. doi: 10.1007/978-981-13-0656-3_1.
- [9] A. Association, "2015 Alzheimer's disease facts and figures," *Alzheimer's Dement.*, vol. 11, no. 3, pp. 332–384, 2015.
- [10] S. MANGIATERRA, "The human rights perspective on dementia: from prevention to social care priority in national agenda".
- [11] M. A. Ebrahimighahnavieh, S. Luo, and R. Chiong, "Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review," *Comput. Methods Programs Biomed.*, vol. 187, p. 105242, 2020.
- [12] L. L., "Machine learning classifiers based on dimensionality reduction techniques for the early diagnosis of Alzheimer's disease using magnetic resonance imaging and positron emission tomography brain data.," 2021.
- [13] L. A. De Santi, E. Pasini, M. F. Santarelli, D. Genovesi, and V. Positano, "An explainable convolutional neural network for the early diagnosis of Alzheimer's disease from 18F-FDG PET," *J. Digit. Imaging*, vol. 36, no. 1, pp. 189–203, 2023.
- [14] M. Odusami, R. Damasevicius, E. Milieskaite-Belousoviene, and R. Maskeliunas, "Multimodal Neuroimaging Fusion for Alzheimer's Disease: An Image Colorization Approach With Mobile Vision Transformer," *Int. J. Imaging Syst. Technol.*, vol. 34, no. 5, pp. 1–13, 2024, doi: 10.1002/ima.23158.
- [15] T. M. Ghazal and G. Issa, "Alzheimer disease detection empowered with transfer learning," *Comput. Mater. Contin.*, vol. 70, no. 3, pp. 5005–5019, 2022.
- [16] H. B. Baydargil, J. Park, and I. F. Ince, "Anomaly-based Alzheimer's disease detection using entropy-based probability Positron Emission Tomography images," *ETRI J.*, vol. 46, no. 3, pp. 513–525, 2024, doi: 10.4218/etrij.2023-0123.
- [17] M. Yildirim and A. Cinar, "Classification of Alzheimer's disease MRI images with CNN based hybrid method," *Ing. des Syst. d'Information*, vol. 25, no. 4, pp. 413–418, 2020, doi: 10.18280/isi.250402.
- [18] I. Beheshti, H. Demirel, and H. Matsuda, "Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm," *Comput. Biol. Med.*, vol. 83, no. November2016, pp. 109–119, 2017, doi: 10.1016/j.compbiomed.2017.02.011.
- [19] S. M. A. H. Shah, M. Q. Khan, A. Rizwan, S. U. Jan, N. A. Samee, and M. M. Jamjoom, "Computer-aided diagnosis of Alzheimer's disease and neurocognitive disorders with multimodal Bi-Vision Transformer (BiViT)," vol. 27, no. 3. Springer London, 2024. doi: 10.1007/s10044-024-01297-6.
- [20] P. S. Sisodia, G. K. Ameta, Y. Kumar, and N. Chaplot, "A Review of Deep Transfer Learning Approaches for Class-Wise Prediction of Alzheimer's Disease Using MRI Images," *Arch. Comput. Methods Eng.*, vol. 30, no. 4, pp. 2409–2429, 2023, doi: 10.1007/s11831-022-09870-0.
- [21] M. Alshammari, S. Arabia, M. Mezher, and S. Arabia, "A Modified Convolutional Neural Networks For MRI-based Images For Detection and Stage," 2021.
- [22] H. T. Gorji and N. Kaabouch, "A deep learning approach for diagnosis of mild cognitive impairment based on mri images," *Brain Sci.*, vol. 9, no. 9, 2019, doi: 10.3390/brainsci9090217.
- [23] M. A. Bravo-Ortiz et al., "A systematic review of vision transformers and convolutional neural networks for Alzheimer's disease classification using 3D MRI images," *Neural Comput. Appl.*, vol. 0123456789, pp. 21985–22012, 2024, doi: 10.1007/s00521-024-10420-x.
- [24] M. S. Kamal, "Interpretable Transformers for Alzheimer Disease Diagnosis on Multi- modal Data," 2024.
- [25] C. L. Y. Cui, "Trans-ResNet: Integrating Transformers and CNNs for Alzheimer's Disease classification," 2022, [Online]. Available: 10.1109/ISBI52829.2022.9761549
- [26] Z. Hu, Y. Li, Z. Wang, S. Zhang, and W. Hou, "Conv-Swinformer: Integration of CNN and shift window attention for Alzheimer's disease classification," *Comput. Biol. Med.*, vol. 164, 2023, doi: 10.1016/j.compbiomed.2023.107304.
- [27] N. Shaffi, K. Subramanian, V. Vimbi, F. Hajamohideen, A. Abdesselam, and M. Mahmud, "Performance Evaluation of Deep, Shallow and Ensemble Machine Learning Methods for the Automated Classification of Alzheimer's Disease," *Int. J. Neural Syst.*, vol. 34, no. 7, 2024, doi: 10.1142/S0129065724500291.
- [28] N. Shaffi, V. Viswan, and M. Mahmud, "Ensemble of vision transformer architectures for efficient Alzheimer's Disease classification," *Brain Informatics*, vol. 11, no. 1, p. 25, 2024.
- [29] M. A. Mahmud Joy, S. Nasrin, A. Siddiqua, and D. M. Farid, "ViTAD: Leveraging modified vision transformer for Alzheimer's disease multi-stage classification from brain MRI scans," *Brain Res.*, vol. 1847, p. 149302, 2024, doi: <https://doi.org/10.1016/j.brainres.2024.149302>.
- [30] L. L. Wenjie Kang, "Multi-model and multi-slice ensemble learning architecture based on 2D convolutional neural networks for Alzheimer's disease diagnosis1," 2021, [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0010482521004728>, 2021.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TablesPaper3.docx](#)