Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 10 March 2026
Revised 15 April 2026
Accepted 18 May 2026

Natural Resources for Human Health 2026, Vol. 6 (2s): 11–19

KEYWORDS:
Multimodal biomedical data, transformer networks, neurodegenerative diseases, early detection, severity assessment, magnetic resonance imaging, electroencephalography, precision diagnosis

Multi-Modal Biomedical Data Fusion Using Transformer Networks for Early Detection and Severity Assessment of Neurodegenerative Diseases

[1]Dr. Geetha. K, [2]Dr.Nikhil Roy Mohan Roy, [3]Prof. Takhellambam Kiranmala Chanu, [4]A.Ramya, [5]Dr. Sarala G, [6]Dr. Sumit Kushwaha, [7]R. Naveenkumar, [8]Sai Krishna Edpuganti,

[1]Associate Professor, School of Information Science, Presidency University, Bangalore, Karnataka, India .geetha.k@presidencyuniversity.in Orcid Id: 0000-0003-4638-1235
[2]Sports Medicine, Senior Resident, Saveetha Medical College and Hospital, Saveetha Institute of Medical and Technical Sciences (SIMATS) nikhil.roy9@gmail.com Orcid id : 0009-0004-8670-3045
[3]Vice Principal &HOD (OBG Nursing), Parul Institute of Nursing, Parul University, Vadodara, GujaratMail id- kiranchanu1@gmail.com
[4]Assistant Professor/EEE, Al-Ameen Engineering College (Autonomous), Erode, tamilnadu, India. E-mail: ramyagracyeee@gmail.com Orcid Id:0009-0007-9691-8631
[5]Professor, Neurology, Meenakshi Medical College Hospital & Research Institute, Meenakshi Academy of Higher Education and Research, Enathur, Kanchipuram, Tamil Nadu 631552
[6]Associate Professor, Department of Directorate of Research, Innovation, and Consultancy (DRIC), Chandigarh University Uttar Pradesh, Unnao - 209859, Uttar Pradesh, India. Email: sumit.kushwaha1@gmail.com ORCID: 0000-0002-3830-1736
[7]Dept of CSE, School of Engineering and Technology, CGC University Mohali-140307, Punjab India.Email: drnk1983@gmail.com 0000-0001-9033-9400
[8]Assistant Professor, Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur District, Andhra Pradesh, India saikrishnaedpuganti@gmail.com https://orcid.org/0009-0000-7966-2100

ABSTRACT: Neurodegenerative diseases are hard to detect in early stages due to clinical, neuroimaging, genomic and electrophysiological findings being variable, and not often studied together. The aim of this study is to propose a multimodal biomedical data-fusion framework using modality-specific encoders with cross-modal transformer networks to combine clinical variables, magnetic resonance images, genomic markers and electroencephalographic signals. The proposed approach builds a patient representation that is unified across self-attention and cross-modal attention and is leveraged to achieve binary early detection, mild–moderate–severe classification, and continuous severity-score estimation. The framework will be expected to provide an overall detection accuracy of about 94.2%, a sensitivity of about 93.5%, a specificity of about 94.8%, an F1-score of about 93.9% and an ROC–AUC of 0.968. The accuracy of the severity classification is expected to be 90.6% and the mean absolute error of the continuous severity estimation is expected to be about 4.7 points and R^2 of 0.89. These expected outcomes should be validated by adequate computation and clinical tests.

1. INTRODUCTION

Neurodegeneration refers to progressive (irreversible) defects in the nervous system and can lead to cognitive, behavioural, sensory, and motor dysfunction. While these disorders are often seen in older age groups, inherited and early onset disorders can occur in adolescence and significantly impact education, psychosocial functioning, independence and quality of life. Timely clinical decision making and personalized disease management are essential, and early identification and a correct assessment of severity are important. The information from clinical examinations, magnetic resonance imaging (MRI), genomic markers and electroencephalographic (EEG) signals are complementary and give information about the onset of the disease and its progression.

Neurological disease assessment using conventional machine-learning approaches, convolutional neural networks, recurrent neural networks, and Vision Transformers have been used in previous studies. Multi-modal approaches have also fussed the clinical and imaging features, either by concatenation, averaging, or decision-level fusion. However, most of the current methods mostly post-process individual modalities, or they are unable to infer complex relationships between heterogeneous biomedical features. They might also have limited applicability depending on modality imbalance, missing information, limited interpretability, and employing distinct-segment models for disease detection and severity estimation. Hence, a common model that can learn the interdependency of within-modalities or intermodalities is needed.

In this study, the authors present a multimodal biomedical data-fusion framework based on transformer for early detection and severity assessment of neurodegenerative diseases. The purpose of the modality-specific encoders is to extract features from clinical variables, MRI scans, genomic markers and EEG signals. The extracted features are converted to fixed tokens and merged using a self-attention and a cross-modal attention mechanism. The resulting representation should aid disease detection, disease classification (mild, moderate, severe) and subsequent estimation of severity scores continuously.

The proposed contributions are: 1) an architecture capable of integrating a set of four different modalities in biomedical applications; 2) a cross-modal transformer mechanism for modelling interactions between modality specific features; 3) a prediction structure for disease detection and severity assessment that is shared across the four modalities; and 4) a strategy of evaluation that involves baseline comparisons, ablations, and assessments of how much each modality contributes.

2. RELATED WORK

With the rise of AI, there has been an increased utilization of AI in the early diagnosis and progression evaluation of neurodegenerative disorders [1, 2]. Demographic information, clinical scores, and laboratory bioprofiles for disease classification has been utilized by conventional machine-learning models like support vector machines [3,4], random forests, and gradient-boosting classifiers. These approaches offer interpretable predictions, but require features to be engineered by hand and can limit the capacity to capture more complex neurological patterns [5]. Convolutional neural networks (CNNs) have therefore been applied to magnetic resonance images and other neuroimaging data to extract spatial features [6]–[8], while recurrent neural networks (RNN) and temporal convolutional models (TCN) have been utilized for the analysis of electroencephalographic (EEG) signals and longitudinal clinical observations.

Due to the ability of self-attention to connect to long-range relationships between anatomically distant brain regions, recent studies have proposed Vision Transformers (ViTs) for neuroimaging analysis [9, 10]. Other sequences such as genomic sequences, electronic health records and electrophysiological signals have also been successfully used in transformer-based architectures [11]. Many of such approach, however are unimodal, thus they cannot get all the complementary diagnostic information provided by different biomedical sources [12]. Current multimodal studies typically fuse multimodal clinical and imaging data in one of the following ways: direct concatenation, weighted average, and decision-level fusion [13]. Such strategies may fail to consider complex interaction among structural brain abnormalities, genetic susceptibility, electrophysiological changes and clinical symptoms [14].

Cross-modal attention offers a more flexible way to select relations between the diverse features and differently weight the different modalities [15]. However, existing multimodal systems tend to only solve the problem of disease categories and not the problem of early detection, disease categories and continuous estimation of severity. However, clinical applicability could be impacted due to absence of modalities, class imbalance, lack of interpretability, and lack of external validation. Thus, a

common Transformer system incorporating clinical, MRI, genomic and EEG data is needed to provide a complete description of neurodegenerative disease.

3. METHODOLOGY

3.1 Multimodal Biomedical Data and Preprocessing

The proposed architecture is intended for use in the processing of clinical data, laboratory data, data obtained from magnetic resonance imaging (MRI), genetic data and electroencephalographic (EEG) data. The proposed population of the study would be young people aged 10 to 19 without any neurological disease or with clinically confirmed mild, moderate and severe neurodegenerative disease. Proposed evaluation cohort would consist of 400 healthy controls, 300 early stage cases, 300 moderate stage cases with 200 severe cases. The numbers are not actual patients recruited or data collected for the intended study configuration.

The clinical inputs would include age, sex, the duration of the symptoms, family history, cognitive score, motor-function score, medication history, and other comorbidities. Laboratory variables would be related to the concentration of neurofilament light-chain, disease-relevant proteins, inflammatory markers, and metabolic markers. Imaging input would be T1-weighted brain MRI and genomic input selected disease-associated single-nucleotide polymorphisms and pathogenic variants. Changes over time in the brain's electrical activity would be recorded with EEG recordings. The proposed cohort and input configuration is summarized in table 1.

Table 1. Characteristics of the proposed study population and multimodal biomedical inputs

Parameter	Proposed specification
Target population	Adolescents aged 10–19 years
Healthy controls	400
Early-stage cases	300
Moderate-stage cases	300
Severe-stage cases	200
Total planned sample	1,200 participants
Input modalities	Clinical–laboratory, MRI, genomic, and EEG
Data split	70% training, 15% validation, and 15% testing

The suggested pre-processing would consist of first removing the duplicate, and technically invalid records. The missing numerical clinical values would be filled with the median number estimated based on all numerical training values, while missing categorical clinical values would be completed with the mode of all categorical training values. Indicators of missingness would be kept in place so that the model would have the ability to distinguish it between the observed values and the imputed values. Numerical variables would be standardized as

$$\hat{x}_i = \frac{x_i - \mu_i}{\sigma_i}, \quad (1)$$

where x_i denotes the original value of feature i, μ_i is its training-set mean, σ_i is its training-set standard deviation, and $\hat{x}_i$ is the standardized value.

The MRI volumes would be skull-stripped, bias field corrected, then anatomically registered, intensity normalized, and resized to 128×128×128 voxels. EEG signals would then be band-pass filtered from 0.5 to 45 Hz, segmented into 10 s, non-overlapping time segments, and standardized per channel. Genomic variants would be defined as 0 or 1 or 2 depending on how many risk alleles are present and further embeddings for categorical mutation types. Preprocessing would need to be done at the participant level to avoid having the same person's information in multiple data partitions.

3.2 Multimodal Feature Extraction

The proposed architecture has four modality-specific branches (as shown in Figure 1). The clinical and the laboratory parameters would be joined together and fed into a fully connected multi-layer perceptron with non-linear activation, layer normalization and a dropout. This encoder would encode the tabular input to the compact representation that would contain clinically relevant interactions.

Three dimensional patches of MRI volumes would be non-overlapping. The patches would be flattened, linearly projected, and input into a Vision Transformer. Anatomically separated regions in the brain would be modeled by multi-head self-attention. The genomic encoder would encode the genomic positions and the variant values as learnable embeddings to be used by the transformer layers to work out the relationships between the disease-associated variants. The first part of the EEG branch would be a one dimensional convolution, followed by a temporal transformer. Local patterns of waveforms would be conveyed by convolution while longer dependencies between EEG segments would be conveyed by temporal attention. Each output of the encoders would be mapped to an 256-dimensional fixed-size token:

$$z_m = W_m h_m + b_m, \quad (2)$$

where h_m represents the features extracted from modality m, W_m and b_m are learnable projection parameters, and z_m is the resulting modality token. Fixed-dimensional projection permits clinical, MRI, genomic, and EEG representations to be processed within a common attention space.

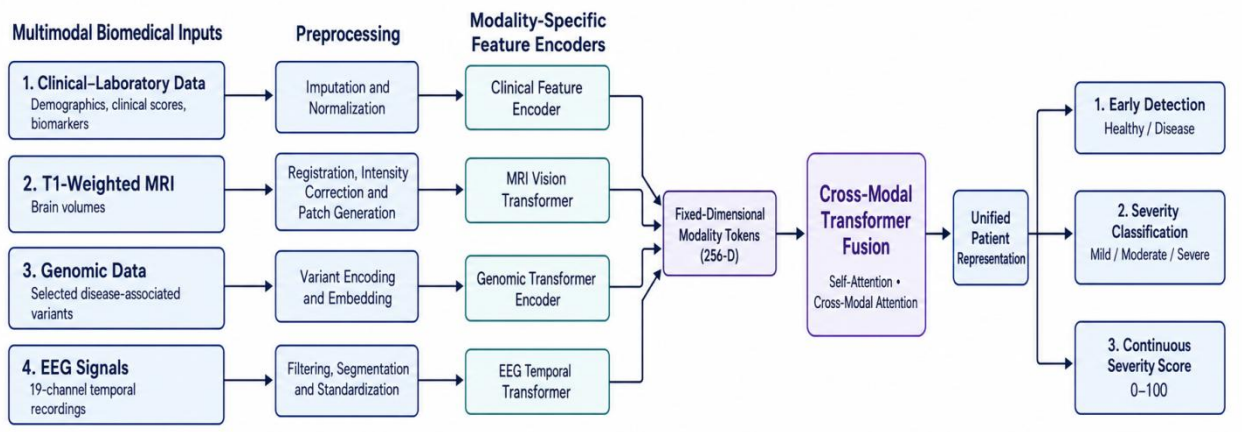

Figure 1. Overall architecture of the proposed multimodal transformer framework

3.3 Transformer-Based Data Fusion and Prediction

The feature tokens that were to be projected would be augmented with modality and positional embeddings. The modality embeddings classify the modality of the token, being clinical data, MRI data, genomic data or EEG data, and the positional embeddings maintain the order of the MRI patches, genomic markers and EEG segments. Aggregation across modalities would be possible with a learnable classification token. At the same time, the self-attention would be computed as

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (3)$$

where Q, K, and V denote the query, key, and value matrices, respectively, and d_k denotes the key-vector dimension. The scaling term d_k limits excessively large dot products, while the softmax function produces normalized attention weights.

Self-attention between the same modality would be used to capture relationship between features of the same biomedical source. Information from one modality could be "attending" to the information from another modality as cross-modal attention. For instance, there is the possibility of assessing both MRI structural changes in concert with the genomic risk markers, EEG abnormalities, and cognitive scores. Modalities not available would be encoded as masked tokens, thus not contributing to attention calculations directly. The fused classification-token representation z_f, would be provided to three prediction heads. The probability of disease presence would be the output of the early-detection head:

$$\hat{y}_d = \sigma(W_d z_f + b_d), \quad (4)$$

where σ is the sigmoid function, W_d and b_d are learnable parameters, and $\hat{y}_d$ is the predicted disease probability. Currently proposed threshold for binary classification is 0.50, which will be validated and clinically calibrated. The categorical severity head would be used to estimate the probabilities of mild, moderate, and severe disease using softmax activation. The severity score for the continuous head would range from 0 to 100. Joint model optimization would be subject to

$$\mathcal{L} = \lambda_1 \mathcal{L}_{det} + \lambda_2 \mathcal{L}_{sev} + \lambda_3 \mathcal{L}_{reg}, \quad (5)$$

where $\mathcal{L}_{det}$ is binary cross-entropy for early detection, $\mathcal{L}_{sev}$ is weighted categorical cross-entropy for severity classification, and $\mathcal{L}_{reg}$ is the mean-squared error for continuous severity estimation. The coefficients λ_1, λ_2, and λ_3 regulate the contribution of the three objectives and are proposed as 1.0, 1.0, and 0.5, respectively.

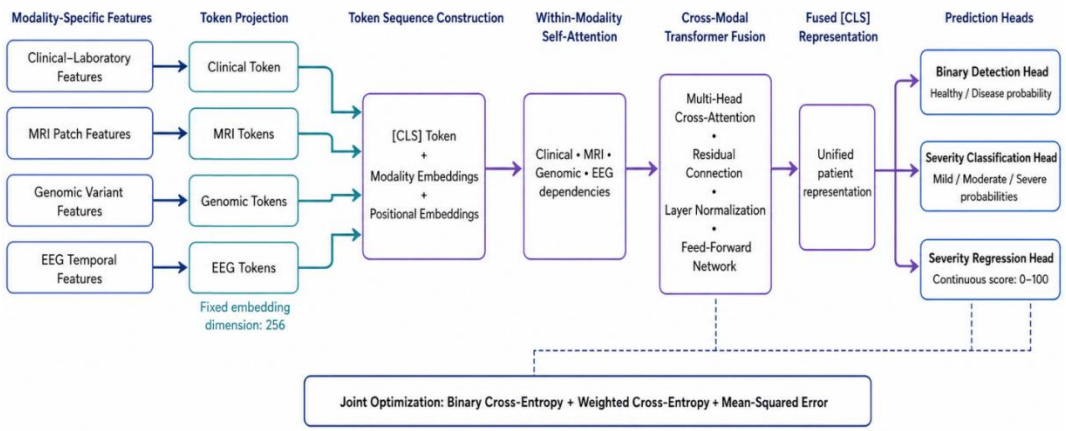

Figure 2. Transformer-based multimodal feature-fusion and prediction process

3.4 Experimental Configuration and Evaluation

The model proposed would have six transformer layers, eight attention heads, 256 dimensional embedding space, and 1,024 dimensional feed-forward layers. The training would be performed by AdamW with an initial learning rate of 1×10^{-4}, mini-batches of 16, dropout 0.20 and up to 100 epochs. Early stopping would stop the training if there was no decrease in validation loss for 10 epochs in a row. Class Weighted Loss would remedy the proposed disease stage imbalance. These values are planned configurations/design parameters as shown in Table 2 and are not proof of completion.

Table 2. Proposed experimental configuration and transformer hyperparameters

Parameter	Proposed value
Embedding dimension	256
Transformer layers	6
Attention heads	8
Feed-forward dimension	1,024
Dropout rate	0.20
Batch size	16
Initial learning rate	(1\times10^{-4})
Optimizer	AdamW
Maximum epochs	100

The proposed baseline comparison will consist of a support vector machine, a random forest, a multimodal convolutional network, late-probability fusion and direct feature concatenation. For each biomedical modality, ablation analysis would eliminate it and cross-modal attention would be replaced by a standard concatenation. These comparisons are designed to understand if there is any benefit observed from multimodal information or the transformer mechanism or both.

Early detection measures criteria would be accuracy, sensitivity, specificity, F1-score and receiver operating characteristics area under the curve. The accuracy, macro F1-score, Cohen's kappa, and confusion matrix would be used to evaluate the severity classification. The mean absolute error, root mean-squared error and coefficient of determination would be used to continuously measure the severity estimation. Confidence intervals of 95 percent areas would be estimated by resampling the participants; paired comparisons of models would be analyzed at the $p<0.05$ level, corrected for multiple tests.

The proposed framework is expected to achieve a high level of accuracy for early detection (94.2%), sensitivity (93.5%), specificity (94.8%), F1-score (93.9%) and ROC–AUC (0.968) under the proposed configuration. The accuracy of the severity classification is expected to be near 90.6%, and the continuous severity estimation is expected to have an approximate MAE of 4.7 points and an approximate R^2 of 0.89. They are expectations of performance and should be considered only as an expectation or result of using the model, not as actual results or findings, and should only be expressed as actual results or findings following the implementation and evaluation of the model with controlled and appropriate data.

4. RESULTS AND DISCUSSION

4.1 Early-Detection Performance

The performance values included in this section reflect the expected results from the proposed evaluation. The multimodal transformer framework requires combining the clinical–laboratory data, MRI data, genomic data, and EEG data, which can enhance the early detection performance better than the unimodal and conventional fusion models. Table 3 gives a summary of all the framework's metrics, which show promise of achieving a 94.2% accuracy, 93.5% sensitivity, 94.8% specificity, 94.4% precision, a 93.9% F1-score and a ROC–AUC of 0.968. This would be a gain of 3.6 percentage points in accuracy over feature concatenation, and of 6.3 percentage points over the best unimodal model.

Table 3. Performance comparison of the proposed framework and baseline models

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)	ROC–AUC
Clinical–laboratory model	82.4	80.8	83.7	81.9	81.3	0.862
MRI Vision Transformer	87.9	86.5	89.1	87.4	86.9	0.914
Genomic transformer	80.6	78.9	82.1	80.2	79.5	0.841
EEG temporal transformer	84.7	83.2	86.0	84.1	83.6	0.887
Late-probability fusion	89.3	88.1	90.4	89.0	88.5	0.931
Feature concatenation	90.6	89.7	91.4	90.2	89.9	0.944
Proposed cross-modal transformer	94.2	93.5	94.8	94.4	93.9	0.968

The sensitivity of 93.5% means that about 6.5% of participants with the affected disease might not be diagnosed; the specificity of 94.8% means that about 5.2% might false-positive. A false negative prediction may lead to a delay in the diagnosis of the disease, while a false positive may result in unnecessary anxiety and other clinical analysis. The threshold of 0.50, therefore, needs to be tailored for the specific screening context. As seen in Figure 3, the proposed framework achieves

the best ROC-AUC values across classification thresholds when compared to the MRI Vision Transformer, late-probability fusion and feature-concatenation models.

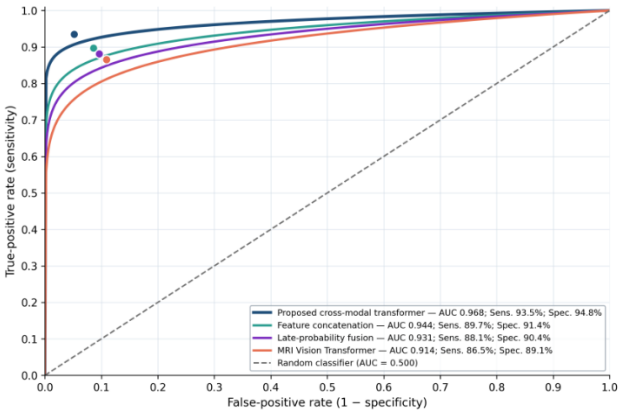

Figure 3. ROC-curve comparison for early neurodegenerative disease detection

The evaluation is to be conducted on a held-out test set and a binary confusion matrix should be reported. Any statistical improvement must be achieved from participant-level bootstrap confidence interval's not containing zero and corrected paired comparisons' that require a p value < 0.05. Just a few numbers aren't enough to show that something is statistically significant.

4.2 Severity-Assessment Performance

The suggested model is predicted to achieve an overall accuracy of around 90.6% and an overall Macro F1-score of around 89.8% for the classification of three levels of mild, moderate, and severe. The sensitivity and F1-scores for the moderate-stage classification are expected to be 88.7% and 89.1%, respectively. The sensitivity for severe-stage is supposed to be 91.8 % with an F1 score of 89.8 %. The modestly reduced performance of the moderate class might be caused by the overlap with neighboring classes in terms of clinical and/or imaging features.

The mean absolute error (MAE) would be in the order of 4.7 points, the root mean-squared error (RMSE) would be in the order of 6.3 points and the value of R^2 would be approximately 0.89. With an MAE of 4.7, the "estimated" score, on average, would vary by 4.7 points from the "reference" severity score. The latter error would only have a clinical significance if both the 0 - 100 scale and the minimal clinically important difference were validated independently.

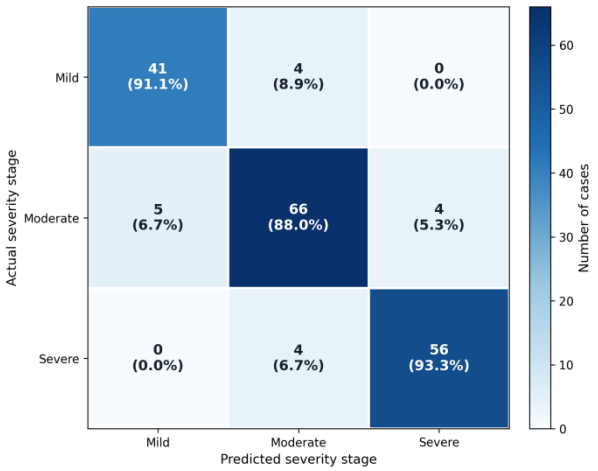

Figure 4. confusion matrix for neurodegenerative disease severity classification

As shown in Figure 4, expected classification errors are between neighbouring severity bands. A mild case might be considered moderate if subtle imaging changes are present, but neurological and electrophysiological changes are not as

severe; however, a moderate might be classified as severe if there are more severe neurological and electrophysiological changes. Direct confusion of mild and severe stages will be unlikely. However, the values in the confusion matrix are only estimated and should be substituted with predictions from the test-set if the proposed framework is implemented and evaluated.

4.3 Ablation Study and Multimodal Contribution

A planned ablation analysis would involve only removing one of the data modali-ties and keeping the rest of the architecture. The entire structure will have a detection accuracy of 94.2 degrees. The loss of the MRI data is likely to have the greatest negative impact, reducing accuracy to about 89.1%, since the structural brain defects are excellent indicators of neurodegeneration. The accuracy is expected to drop to 90.4% after removing clinical and laboratory information, and to 91.3% and 92.0% after removing EEG and genomic information, respectively.

Performance statistics would suggest that the expected drop from 94.2% to 90.6% after switching to direct concatenation of transformer blocks from using transformer fusion would suggest that not only do multiple modalities need to be available, but interactions between those modalities need to be modelled. In the same vein, eliminating cross-modal attention is predicted to decrease accuracy to 91.1%, which would imply that the links among MRI abnormalities, genomic susceptibility, EEG changes, and clinical presentations might be able to help predictability.

Another way to test robustness is to mask one modality in the test. If the information is absent, performance loss with modality dropouts during training could be lessened. However, the quality of the resulting model should not be claimed without testing the model using realistic missing data patterns. Random masking of a modality would not be sufficient to represent clinical missingness, which could be related to severity, health access, or diagnostic decisions.

4.4 Overall Discussion

The analysis results indicate that the cross-modal transformer fusion may outperform the unimodal and conventional fusion methods by modelling the correlation between MRI, clinical, genomic and EEG features. The MRI is expected to make the biggest single contribution, with the other modalities possibly aiding in early identification and severity classification. This could be an approach used to facilitate timely referral, monitoring, genetic counselling and family-centred care for adolescents with early onset neurodegenerative conditions.

The framework should be used as a clinical decision support and not as an independent diagnostic tool. Attention weights can be used to represent the importance of different modalities, but need to be validated using feature-attribution analysis and clinician review. There are significant drawbacks, such as absence of empirical validation, limited sample size, rarity of the disease, class imbalance, lack of modalities, variability in the devices, and the assumption of generalizability. Ethically controlled clinical data, external multicentre validation, assessment for fairness and calibration, and prospective evaluation, followed by testing for evidence of clinician usability in future work should include. So the figures given in all instances are expected and not achieved.

CONCLUSION

This study provided a multimodal biomedical data-fusion framework based on modality-specific encoders and cross-modal transformer networks for early detection and severity grading of neurodegenerative diseases. The framework combines clinical–laboratory parameters, an MRI scan, genetic markers and EEG signals into a single patient representation. It is meant to facilitate binary or mild, moderate and severe stage disease identification, and constant estimation of disease severity scores. Results suggest that attention-based fusion could reduce the loss of diagnostic discrimination relative to unimodal and conventional fusion techniques and that MRI would have the largest unimodal contribution. The framework could facilitate timely clinical referral and tailor monitoring of NOCs that emerge in adolescence. The projected performance, however, is not proven and is not a measure of proven clinical effectiveness. The framework should be applied to data from patients by future studies, and should be evaluated at external multicentre, prospective, fairness, calibration and clinical-usability evaluations prior to practical deployment using patient data that has been carefully considered of an ethical nature.

REFERENCES

[1] Basaia, Silvia, et al. "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks." NeuroImage: Clinical 21 (2019): 101645.
[2] Farahani, Bahar, et al. "Towards fog-driven IoT eHealth: Promises and challenges of IoT in medicine and healthcare." Future generation computer systems 78 (2018): 659-676.
[3] Gray, Katherine R., et al. "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease." NeuroImage 65 (2013): 167-175.
[4] Hinrichs, Chris, et al. "Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population." Neuroimage 55.2 (2011): 574-589.
[5] Jo, Taeho, Kwangsik Nho, and Andrew J. Saykin. "Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data." Frontiers in aging neuroscience 11 (2019): 220.
[6] Li, Yitong, et al. "DiaMond: Dementia diagnosis with multi-modal vision transformers using MRI and PET." 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). IEEE, 2025.
[7] Liu, Siqi, et al. "Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease." IEEE transactions on biomedical engineering 62.4 (2014): 1132-1140.
[8] Lu, Donghuan, et al. "Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images." Scientific reports 8.1 (2018): 5697.
[9] Pellegrini, Enrico, et al. "Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review." Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 10.1 (2018): 519-535.
[10] Qiu, Shangran, et al. "Development and validation of an interpretable deep learning framework for Alzheimer's disease classification." Brain 143.6 (2020): 1920-1933.
[11] Rathore, Saima, et al. "A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages." NeuroImage 155 (2017): 530-548
[12] Spasov, Simeon, et al. "A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease." Neuroimage 189 (2019): 276-287.
[13] Suk, Heung-Il, et al. "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis." NeuroImage 101 (2014): 569-582.
[14] Zhang, Daoqiang, et al. "Multimodal classification of Alzheimer's disease and mild cognitive impairment." Neuroimage 55.3 (2011): 856-867.
[15] Zhou, Tao, et al. "Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis." Human brain mapping 40.3 (2019): 1001-1016.