The Explainable Attention-Guided Multimodal Ensemble Framework for Alzheimer's Disease Prediction and Severity Estimation

K. Laxmikantha
Department of CSE, KS Institute of Technology, Bengaluru, Visvesvaraya Technological University, Belagavi, Karnataka, India
klaxmikantha4@gmail.com (corresponding author)

P. S. Ashokkumar
Department of CSE, ACS College of Engineering, Bengaluru, Visvesvaraya Technological University, Belagavi, Karnataka, India
ashokdbit2017@gmail.com

Received: 19 February 2026 | Revised: 31 March 2026 and 13 April 2026 | Accepted: 26 April 2026

 DOI: https://doi.org/10.48084/etasr.18269

ABSTRACT

Globally, Alzheimer's Disease (AD) is a major cause of cognitive impairment. Despite the contribution made by machine-learning techniques to the evolution of diagnostic techniques, the existing systems rely mainly on one-modality inputs and are defined as opaque black-box models, thus restricting their clinical usefulness and predictive accuracy. In this paper, the Explainable Attention-Guided Feature-Weighted Multimodal Ensemble Framework (EAFW-MEF) framework is proposed for the evaluation of five types of AD and the severity estimation of the disease in continuous form. The framework integrates clinical data, psychological tests, and MRI-based volumetric biomarkers of the OASIS-3 dataset through a novel explainability-based fusion plan. In contrast with data-level fusion, the EAFW-MEF models global SHAP importance scores to guide an attentional score-driven weighting procedure through which cycling the multimodal features before the model is trained can be recalibrated. The classification is done in a hybrid stacked ensemble utilizing Random Forest, XGBoost, and LightGBM whose outputs are combined with an interpretable logistic regression meta-learner. In addition to classification, the study gives rise to the Alzheimer RiskSeverity Index, which provides a continuous estimate of disease severity at the current visit and can support clinical stratification. Experimental findings point to better performance, with a stratified 10-fold cross-validation accuracy of 99.12% and, therefore, exceeding the current state-of-the-art multimodal methods.

Keywords-Alzheimer's disease; multimodal learning; explainable artificial intelligence; SHAP; ensemble learning; disease severity estimation

I. INTRODUCTION

Alzheimer's Disease (AD) is a neurodegenerative, incurable condition resulting in progressive memory impairment, cognitive impairment, and decreased daily functioning. It is the leading cause of dementia [1]. Proper and early diagnosis is crucial in slowing down the progression of the disease and enhancing the patients' quality of life. Clinical diagnosis is, however, not easy since the symptoms of AD tend to be similar to those of other dementias and differ between different individuals. In this regard, Machine Learning (ML) and Artificial Intelligence (AI) are utilized in the automated prediction of AD through the use of neuroimaging, clinical, and cognitive data [2, 3]. Nonetheless, most of the current approaches utilize one modality, like structural MRI or cognitive scores, restricting robustness while being sensitive to noise or data imbalance. Moreover, numerous deep learning and ensemble models are black boxes, making them less interpretable to clinicians and less practical to trust [4].

Multimodal learning provides a more powerful method of AD prediction, as it combines complementary data on clinical measures, neuropsychological testing, and neuroimaging biomarkers [5]. Recent experiments indicate that the use of multimodal fusion tends to be more effective than unimodal schemes. The majority of the existing frameworks, however, are based on the basic concatenation of the features and do not take into consideration the unequal contribution of the various modalities [6]. This is a major limitation, because cognitive decline, impairment of functions, and structural changes of the brain affect diagnosis and the progression of the disease in different ways. Meanwhile, Explainable Artificial Intelligence (XAI) has become significant in medical AI, and tools like SHAP have been extensively used to determine influential

biomarkers in AD prediction. However, explainability is mostly used only after, but not during, model training [7, 8]. Consequently, explainability is not related to model optimization and clinical utility.

Recent AD research is largely concerned with discrete class prediction [9]. This method is not a reflection of the continuous nature of disease progression, although it is useful clinically. Practically, clinicians require not just diagnostic labels, but also effective severity estimations to plan treatment, and decide on long-term care [10]. To fill these gaps, this paper suggests a explainable, optimization-aware multimodal framework of AD severity prediction. The model uses attention-driven prioritization of features by SHAP to re-weight multimodal inputs and classify them, thus incorporating explainability into the learning process. The framework uses a stacked ensemble of Random Forest (RF), XGBoost (XGB), and LightGBM (LGBM) classifiers to enhance robustness and generalization.

Multimodal learning towards AD analysis has been heavily supported by the current research. Authors in [11] indicated that utilizing structural MRI, genetic markers, and clinical history fares better in early detection compared to unimodal schemes. Authors in [12] also highlighted the importance of cross-modal representation learning in effective multimodal fusion. Authors in [13] demonstrated that the combination of demographics, fluid biomarkers, and cognitive measurements could be used to predict the amyloid and tau PET results in biomarker-oriented settings. Corresponding biomedical applications have also demonstrated similar advantages of fusion, e.g. the EEG -ECoG signal synthesis to detect seizures [14]. Authors in [15, 16] mention the increasing importance of multimodal fusion in intelligent healthcare, combined with the necessity to have more valuable and clinically useful interpretations of fused information. Simultaneously, deep learning has demonstrated great potential in neuroimaging-based diagnosis [17, 18].

Ensemble learning is increasingly utilized to enhance prediction performance. Authors in [19] demonstrated that the variance of diverse learners diminishes when they are combined, and generalization is improved, whereas authors in [20, 21] reported that hybrid ensembles are effective in dealing with class imbalance and scarce medical data in AD detection. Meanwhile, explainable AI has become a necessity in the field of healthcare [22, 23]. SHAP and LIME are popular tools in the interpretation of model decisions in AD prediction [24]. The trade-off between interpretability and predictive accuracy of dementia-related AI systems has been explored in [25, 26], whereas authors in [27] utilized explainable deep learning in MRI-based AD staging. These results inspire the current research, which goes beyond the post-hoc explanation by incorporating explainability into feature fusion and severity modeling in multimodal classification.

II. PROPOSED METHODOLOGY

The proposed methodology incorporates explainability as a part of the learning pipeline to enhance robustness and clinical relevance. It comprises four key steps: multimodal data preparation, explainability-informed feature weighting and fusion, hybrid ensemble learning, and continuous AD risks-severity prediction. The overall architecture is shown in Figure 1. The framework incorporates heterogeneous inputs, such as clinical data, neuropsychological measures, and MRI-based volumetric biomarkers, to embody the complementary features of AD. Clinical data indicate impaired functional status, cognitive tests indicate mental deterioration, and MRI characteristics indicate neurodegeneration of the structure. K-Nearest Neighbors (KNN) algorithm is used to impute missing values and Z-score normalization is used to provide consistency across modalities. SMOTE is employed to eliminate class imbalance and enhance the learning between diagnostic groups. In contrast to traditional early-fusion models, the proposed method does the fusion after an explainability-based feature analysis, which gives the opportunity to weight features by their relative importance in disease prediction and progression.

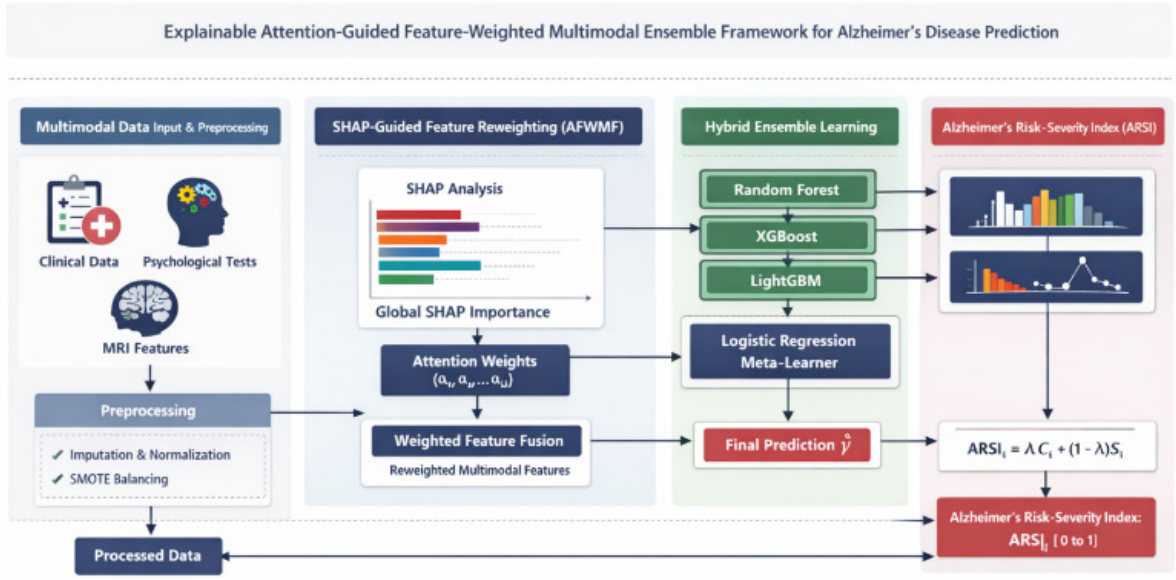

Fig. 1. Architecture of the proposed model.

AD progression is continuous in nature, although categorical diagnosis is clinically useful. To model this behavior, the proposed framework model characterizes the Alzheimer Risk-Severity Index (ARSI) as a continuous score (bounded) in the range [0,1] with larger values representing more disease burden.

For the *i*-th subject, the stacked ensemble outputs a posterior probability vector $\hat{p}_i = [\hat{p}_{i1}, \hat{p}_{i2}, \dots, \hat{p}_{iK}]$, where K denotes the total number of diagnostic classes and $\sum_{k=1}^{K} \hat{p}_{ik} = 1$. To quantify severity without assuming any artificial ordinal relationship among the diagnostic classes, a class-specific severity anchor $v_k \in [0,1]$ is derived from the training data using normalized clinical impairment measures. In this study, v_k represents the severity reference associated with class k.

$$v_k = \frac{1}{2}\left(\overline{\mathrm{CDR}_k}^{\mathrm{norm}} + \left(1 - \overline{\mathrm{MMSE}_k}^{\mathrm{norm}}\right)\right) \quad (1)$$

where $\mathrm{C\bar{D}R}_k^{\mathrm{norm}}$ and $\mathrm{M\bar{M}SE}_k^{\mathrm{norm}}$ denote the class-wise normalized mean values of CDR and MMSE, respectively, computed only from the training fold. Thus, v_k reflects the empirical severity level of each class. Using these anchors, the posterior severity expectation for subject i is computed by:

$$E_i = \sum_{i=1}^{n} \hat{p}_{ik} v_k \quad (2)$$

To account for predictive certainty, an ensemble confidence term is defined using normalized entropy:

$$C_i = 1 - \frac{-\sum_{k=1}^{K} \hat{p}_{ik} \log(\hat{p}_{ik})}{\log(K)}, \quad 0 \le C_i \le 1 \quad (3)$$

A large value of C_i represents a high degree of agreement of the ensemble, and a low value represents a diagnostic ambiguity. The severity component derived by the ensemble can then be expressed in the form of:

$$S_i = C_i \cdot E_i \quad (4)$$

Next, the SHAP-weighted biomarker burden score is constructed. Let B denote the set of biomarkers selected for severity modeling, and let ϕ_j denote the global SHAP importance of biomarker j. The corresponding normalized attention weight is represented by:

$$\alpha_j = \frac{|\emptyset_j|}{\sum_{l \in B} |\emptyset_l|} \quad (5)$$

Each biomarker value is first direction-aligned so that larger values always represent greater disease burden as represented in (6):

$$u_{ij} = \begin{cases} norm(x_{ij}), & \text{if biomarker } j \text{ increses with severity} \\ 1 - norm(x_{ij}), & \text{if biomarker } j \text{ decreses with severity} \end{cases} \quad (6)$$

where x_{ij} is the original value of biomarker $j$for subject i, and $\mathrm{norm}(\cdot)$ denotes min–max normalization to [0,1]. The SHAP-weighted biomarker burden is then given by:

$$B_i = \sum_{j \in B} \alpha_j u_{ij}, \quad 0 \le B_i \le 1 \quad (7)$$

Finally, the ARSI is defined as the convex combination of the ensemble-derived severity component and the SHAP-weighted biomarker burden:

$$ARSI_i = \lambda S_i + (1 - \lambda) B_i, \quad 0 \le ARSI_i \le 1 \quad (8)$$

where $\lambda \in [0,1]$ controls the trade-off between the ensemble-based severity evidence and the biomarker-based severity evidence. In this study, $\lambda = 0.6$, giving slightly higher emphasis (Table I) to the ensemble decision while preserving the contribution of clinically interpretable biomarkers. Table I indicates that the proposed ARSI is quite stable within a wide range of λ values, which proves that the index is not too sensitive to one parameter setting. The highest severity-estimation performance is noted at 0.6 which yields the highest correlation with the reference severity target and the lowest error. All quantities used in ARSI, including class severity anchors v_k, SHAP weights α_j, and normalization statistics, were computed using only the training partition in each cross-validation fold to prevent information leakage.

TABLE I. SENSITIVITY ANALYSIS OF THE ARSI PARAMETER λ

λ	**Spearman correlation with reference severity (ρ)**	**MAE of ARSI**	**Mean inter-class ARSI gap**
0.1	0.842	0.128	0.118
0.3	0.871	0.111	0.134
0.5	0.903	0.095	0.152
0.6	0.917	0.088	0.161
0.7	0.912	0.091	0.156
0.9	0.884	0.106	0.142

III. RESULTS AND DISCUSSION

A. Dataset and Pre-Processing

OASIS-3, a publicly available longitudinal dataset containing neuroimaging, clinical, cognitive, and biomarker data of cognitively normal participants and participants with different cognitive impairment levels and AD, was the subject of experiments [31]. The data consist of demographic and functional variables, including MMSE, APOE genotype, and CDR scores, neuropsychological variables, including logical memory, digit span, and Trail Making Tests, and MRI-based structural biomarkers, including intracranial volume, gray matter volume, supratentorial volume, and cortical white matter. The alignment of subjects across modalities was carried out with the help of unique identifiers and session metadata. OASIS, the data were categorized into five classes: Cognitively Normal (CN), AD, Non-AD Dementia, Uncertain Dementia, and Others. Table II gives a summary of the integrated multimodal dataset.

The training was conducted on a common preprocessing pipeline of all modalities. Missing values were filled in with KNN to maintain local feature correlations and feature scales across modalities were standardized by Z-score. In order to deal with class imbalance, SMOTE was used only with the training folds in order to avoid information leakage. Evaluation of model performance was done based on stratified 10-fold cross-validation where class distribution was maintained in every fold. The hybrid framework was made of three base learners and one meta-learner, and the hyperparameters were fixed in all experiments to ensure fair comparison. Table III lists the parameter settings.

TABLE II. SUMMARY OF THE DATASET USED

Modality	Description	No. of subjects	No. of sessions (samples)	No. of features
Clinical data	Demographic and functional assessments collected under ADRC protocols	1,098	6,200	13
Psychological Assessment	Standardized neuropsychological and cognitive performance tests	810	3,300	17
MRI segmentation	Structural brain volumetric measures derived with Free Surfer	1,050	2,050	10
Fused multimodal dataset	Subject-level integration of clinical, psychological, and MRI features	800	3,200	≈40

Despite the fact that OASIS-3 is longitudinal, this research creates predictions based on each visit and considers it a multimodal clinical snapshot. This decision was made due to the variation in the number of visits per subject, the inequality of the intervals between visits, and the unavailability of full multimodal data at all visits. A longitudinal design based on sequences would significantly decrease the usable cohort, create additional missingness, and exacerbate the imbalance in the classes. Thus, the proposed framework applies subject-level splitting to avoid identity leakage and makes predictions on visit-specific fused features.

TABLE III. SIMULATION PARAMETERS

Model	Parameter	Value used	Justification
RF	Number of trees (n_estimators)	100	Provides a balance between stability and computational efficiency
	Maximum depth (max_depth)	20	Prevents overfitting while capturing non-linear relations
	Minimum samples to split (min_samples_split)	2	Default setting to allow sufficient tree growth
	Bootstrap sampling	Enabled	Improves generalization
XGB	Number of estimators (n_estimators)	200	Enables learning of complex feature interactions
	Subsample ratio (subsample)	0.8	Reduces overfitting
LGBM	Number of estimators (n_estimators)	200	Ensures sufficient boosting rounds
	Number of leaves (num_leaves)	31	Recommended default for depth 6
	Feature fraction	0.8	Prevents feature dominance
	Bagging fraction	0.8	Reduces variance
	Maximum iterations	1,000	Ensures convergence
ARSI computation	Confidence–severity trade-off (λ)	0.6	Emphasizes model confidence while preserving biomarker influence

B. Overall Classification Performance

Table IV presents the average classification results from the stratified 10-fold cross-validation. The proposed EAFW-MEF hybrid ensemble consistently outperforms the individual base learners across all evaluated metrics. While standalone models demonstrate high performance, the ensemble architecture yields systematic improvements, validating the integration of multiple learners within an explainability-aware fusion framework.

TABLE IV. PERFORMANCE OF THE EAFW-MEF FRAMEWORK

Metric	RF	XGB	LGBM	Proposed EAFW-MEF (ensemble)
Accuracy (%)	98.84	98.96	98.91	99.12
Precision (%)	98.94	99.01	98.97	99.21
Recall (%)	98.79	98.88	98.85	99.05
F1-score (%)	98.75	98.94	98.9	99.13
AUC (%)	99.97	99.91	99.93	99.98

C. Class-Wise Analysis

Table V shows the diagnostic performance in the form of the classes of the proposed framework. The CN and AD classes have the highest precision and recall, showing that the well-defined clinical categories are strongly discriminated. The performance of Non-AD Dementia, Uncertain Dementia, and Others is also high albeit slightly lower than the performance in other classes due to their higher clinical heterogeneity, smaller sample representation, and overlapping features with neighboring classes. It is further analyzed that there are more ambiguous and transitional cases in Uncertain Dementia and Others, and they are more difficult than CN and AD to classify.

According to the normalized confusion matrix (Figure 2), the proposed model is able to distinguish between CN and AD, and most of the misclassifications belong in the Non-AD Dementia, Uncertain Dementia, and Others classes. This trend can be attributed to the reduced recall of such classes and indicates that the errors are primarily due to clinical heterogeneity and blurry class demarcation and not due to instability of the overall model. The pooled out-of-fold predictions on the 10 test folds were used to generate the confusion matrix; this is to guarantee that unseen subjects were used to obtain all the results.

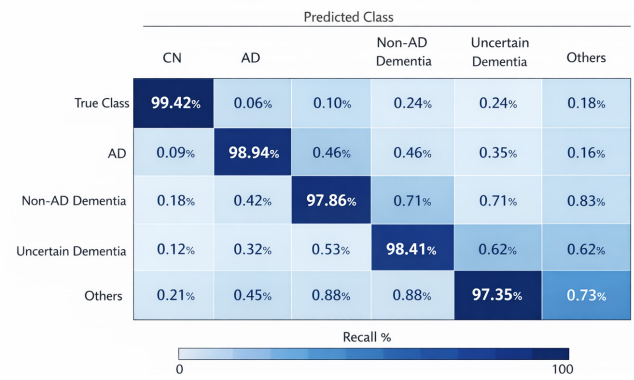

Fig. 2. Normalized confusion matrix.

TABLE V. CLASS-WISE PERFORMANCE OF EAFW-MEF

Class	Precision (%)	Recall (%)	F1-score (%)
CN	99.35	99.42	99.38
AD	99.08	98.94	99.01
Non-AD Dementia	98.21	97.86	98.03
Uncertain Dementia	98.67	98.41	98.54
Others	97.82	97.35	97.58

These errors, as indicated in Table VI, are clustered in the less distinct classes related to dementia and are not randomly dispersed. The most mixed-up classes are Other with Uncertain Dementia (1.11%) and Other with Non-AD Dementia (0.88%). Others (0.83) and Uncertain Dementia (0.71) are also misclassified as Others and Non-AD Dementia, respectively. These findings imply that the rest of the errors can be largely attributed to the overlap of symptoms and poor classification boundaries among the heterogeneous categories of dementia.

TABLE VI. DOMINANT MISCLASSIFICATION PATTERNS

True class	Dominant predicted class	Misclassification rate (%)
Non-AD Dementia	Others	0.83
Non-AD Dementia	Uncertain Dementia	0.71
Uncertain Dementia	Others	0.62
Uncertain Dementia	Non-AD Dementia	0.53
Others	Uncertain Dementia	1.11
Others	Non-AD Dementia	0.88

D. The Impact of Explainable Attention-Guided Fusion

Figure 3 shows the ablation experiment results on how the AFWMF component affects classification accuracy. The complete model with AFWMF has the best accuracy of 99.12%, which proves the efficacy of the explainability-based weighting of features. This demonstrates that AFWMF is an essential model discrimination contributor and not a supporting architectural element. The benefit also shows itself in comparison to the previous approaches, as can be clearly seen in Figure 3. These findings indicate that the traditional multimodal fusion algorithms are less efficient in utilizing heterogeneous features, especially when there are noisy or weakly informative dimensions.

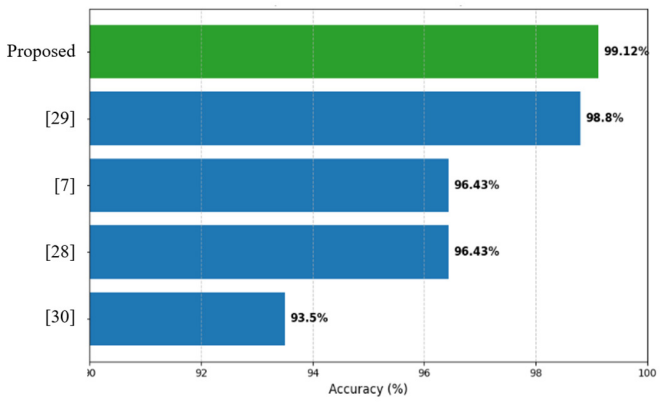

Fig. 3. Accuracy comparison.

E. Alzheimer's Risk–Severity Index (ARSI) Analysis

The ARSI analysis, illustrated in Figure 4, reveals a monotonic increase across diagnostic categories, effectively mapping the progressive nature of cognitive impairment. CN participants exhibit a low average ARSI of 0.15, signaling stable cognitive function. In contrast, the Uncertain Dementia group averages 0.38, reflecting the transitional characteristics and intermediate impairment levels typical of this stage.

Patients with Non-AD Dementia record a mean ARSI of 0.46, representing heightened functional decline while remaining distinct from the Alzheimer's profile. The highest scores are observed in the AD class, with a mean near 0.78, aligning with the severe structural degeneration and cognitive deficits associated with the disease. This systematic progression demonstrates that the ARSI is sensitive to both discrete categorical shifts and fine-grained gradations in clinical severity. By bridging the gap between rigid classification and continuous modeling, the index provides a more nuanced tool for tracking disease trajectory and prioritizing clinical interventions. Table VII gives the comparison with state-of-art methods.

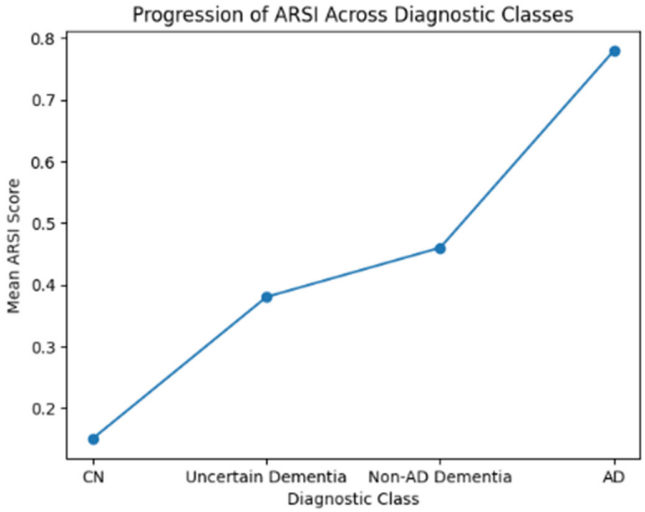

Fig. 4. Progression of ARSI across diagnostic classes.

For statistical significance analysis, the fold-wise accuracy values were calculated using the stratified 10-fold cross-validation to test whether the proposed EAFW-MEF framework significantly performed better than the baseline models. The paired tests were used since all the methods were evaluated using the same data partitions. The results of the proposed model demonstrated the mean accuracy improvements of 0.28%, 0.16%, and 0.21% over RF, XGB, and LGBM. The p-values of the three comparisons were 0.012, 0.021, and 0.015 when using the paired t-test, and 0.018, 0.027, and 0.019 when using the Wilcoxon signed-rank test. All the p-values were less than 0.05, and therefore, the performance gains of EAFW-MEF are statistically significant which confirms that the gains realized are not a result of random variation.

TABLE VII. COMPARISON WITH EXISTING METHODS

Reference	Dataset	Modality	Method	Explainability	Accuracy (%)
[7]	OASIS-3	MRI + Clinical	CNN + ML Hybrid	Post-hoc (SHAP)	74.89
[32]	OASIS-3	MRI + CT	CNN + RF Fusion	No	97.19
[33]	OASIS-3	MRI + DTI	Multimodal DL	No	96
[34]	OASIS-3	MRI + Clinical	Attention DL	Partial	94.1
[35]	OASIS + ADNI	MRI + PET + Clinical	Diffusion + Multimodal DL	No	~95+
[36]	OASIS-3	MRI + PET	CNN-based	No	~83
Proposed	OASIS-3	Clinical + Cognitive + MRI	SHAP-guided Attention + Ensemble	Integrated (pre-training)	99.12

One of the weaknesses of the current work is that even though it utilizes a longitudinal source dataset, the model is not learning any temporal dependencies among repeated visits. Thus, the suggested framework is to be seen as a visit-level multimodal prediction framework as opposed to a complete longitudinal evolution model. Here, the ARSI is the approximate disease burden during a visit and must not be viewed as a directly learned time course. Although this design can be of clinical use to make snapshot-based diagnosis and stratification, future research must expand the framework with sequence-conscious models to include visit sequence, non-uniform time spacing, and incomplete modality paths.

IV. CONCLUSION

In this work, the multimodal EAFW-MEF framework for the classification of Alzheimer's Disease (AD) with high accuracy and interpretability was presented. Unlike traditional multimodal methods, which are based on naive feature concatenation and post-hoc explainability, the proposed framework introduces explainability into the learning process through SHAP-guided attention-based features weighting. Such a design means that the model puts a priority on clinically significant things during fusion to achieve enhanced robustness, transparency and predictive performance. The hybrid ensemble architecture that is based on a combination of Random Forest, XGBoost, and LightGBM and a stacking interpretable meta-learner further improves the generalization through the use of complementary learning behaviours. Experimental assessment of the OASIS-3 data set showed that the presented framework gives a state-of-the-art result in terms of various metrics, and that the ablation experiment supports the fact that explainability-directed fusion has a tangible contribution. Besides, the proposed ARSI offers a continuous and clinically interpretable measure of the severity of the disease at the given visit that would complement categorical diagnosis and aid in more specific stratification of the patients. Nonetheless, the current framework has been visit-based and in future, longitudinal sequence modeling should be considered in future work to explicitly model temporal progression through repeated assessments.

DECLARATION OF COMPETING INTERESTS

The authors declare that there are no conflicts of interest related to this publication. No financial or personal relationships, affiliations, or other interests have influenced the content or the research presented in this manuscript. Furthermore, the authors confirm that the manuscript has not been previously published and is not under consideration for publication elsewhere.

ACKNOWLEDGMENT

This research received no external funding.

DATA AVAILABILITY

The data presented in this study are available on request.

AI USE AND DECLARATION OF GENERATIVE AI USE

The authors declare that no generative AI Tools were used in the preparation of this work

REFERENCES

[1] A. A. Soladoye, N. Aderinto, D. Osho, and D. B. Olawade, "Explainable machine learning models for early Alzheimer's disease detection using multimodal clinical data," *International Journal of Medical Informatics*, vol. 204, Dec. 2025, Art. no. 106093, https://doi.org/10.1016/j.ijmedinf.2025.106093.

[2] F. H. Al-bakri *et al.*, "A Feature-Augmented Explainable Artificial Intelligence Model for Diagnosing Alzheimer's Disease from Multimodal Clinical and Neuroimaging Data," *Diagnostics*, vol. 15, no. 16, Aug. 2025, Art. no. 2060, https://doi.org/10.3390/diagnostics15162060.

[3] M. L. Raza *et al.*, "Advancements in deep learning for early diagnosis of Alzheimer's disease using multimodal neuroimaging: challenges and future directions," *Frontiers in Neuroinformatics*, vol. 19, May 2025, https://doi.org/10.3389/fninf.2025.1557177.

[4] A. S. Alatrany, W. Khan, A. Hussain, H. Kolivand, and D. Al-Jumeily, "An explainable machine learning approach for Alzheimer's disease classification," *Scientific Reports*, vol. 14, no. 1, Feb. 2024, Art. no. 2637, https://doi.org/10.1038/s41598-024-51985-w.

[5] S. Muksimova, S. Umirzakova, J. Baltayev, and Y. I. Cho, "Multi-Modal Fusion and Longitudinal Analysis for Alzheimer's Disease Classification Using Deep Learning," *Diagnostics*, vol. 15, no. 6, Mar. 2025, https://doi.org/10.3390/diagnostics15060717.

[6] F. Leony and C. Lin, "Multimodal fusion architectures for Alzheimer's disease diagnosis: An experimental study," *Journal of Biomedical Informatics*, vol. 166, June 2025, Art. no. 104834, https://doi.org/10.1016/j.jbi.2025.104834.

[7] S. Jahan *et al.*, "Explainable AI-based Alzheimer's prediction and management using multimodal data," *PLOS ONE*, vol. 18, no. 11, 2023, Art. no. e0294253, https://doi.org/10.1371/journal.pone.0294253.

[8] V. Viswan, N. Shaffi, M. Mahmud, K. Subramanian, and F. Hajamohideen, "Explainable Artificial Intelligence in Alzheimer's Disease Classification: A Systematic Review," *Cognitive Computation*, vol. 16, no. 1, pp. 1–44, Jan. 2024, https://doi.org/10.1007/s12559-023-10192-x.

[9] L. G. Amato *et al.*, "Personalized modeling of Alzheimer's disease progression estimates neurodegeneration severity from EEG recordings," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 16, no. 1, 2024, Art. no. e12526, https://doi.org/10.1002/dad2.12526.

[10] G. H. Koksalmis, B. Soykan, L. J. Brattain, and H.-H. Huang, "Artificial Intelligence for Personalized Prediction of Alzheimer's Disease Progression: A Survey of Methods, Data Challenges, and Future Directions." arXiv, Apr. 29, 2025, https://doi.org/10.48550/arXiv.2504.21189.

[11] R. Zhang, J. Sheng, Q. Zhang, J. Wang, and B. Wang, "A review of multimodal fusion–based deep learning for Alzheimer's disease," *Neuroscience*, vol. 576, pp. 80–95, June 2025, https://doi.org/10.1016/j.neuroscience.2025.04.035.

[12] B. Huang, F. Yang, M. Yin, X. Mo, and C. Zhong, "A Review of Multimodal Medical Image Fusion Techniques," *Computational and Mathematical Methods in Medicine*, vol. 2020, no. 1, 2020, Art. no. 8279342, https://doi.org/10.1155/2020/8279342.

[13] T. Jo, K. Nho, and A. J. Saykin, "Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data," *Frontiers in Aging Neuroscience*, vol. 11, Aug. 2019, https://doi.org/10.3389/fnagi.2019.00220.

[14] S. Li and H. Tang, "Multimodal Alignment and Fusion: A Survey." arXiv, Oct. 11, 2025, https://doi.org/10.48550/arXiv.2411.17040.

[15] A. D. Arya *et al.*, "A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease," *Brain Informatics*, vol. 10, no. 1, July 2023, Art. no. 17, https://doi.org/10.1186/s40708-023-00195-7.

[16] T. Shaik, X. Tao, L. Li, H. Xie, and J. D. Velásquez, "A survey of multimodal information fusion for smart healthcare: Mapping the journey from data to wisdom," *Information Fusion*, vol. 102, Feb. 2024, Art. no. 102040, https://doi.org/10.1016/j.inffus.2023.102040.

[17] V. H. Jasodanand *et al.*, "AI-driven fusion of multimodal data for Alzheimer's disease biomarker assessment," *Nature Communications*, vol. 16, no. 1, Aug. 2025, Art. no. 7407, https://doi.org/10.1038/s41467-025-62590-4.

[18] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, "Ensemble Learning for Disease Prediction: A Review," *Healthcare*, vol. 11, no. 12, June 2023, https://doi.org/10.3390/healthcare11121808.

[19] A. A. Khan, O. Chaudhari, and R. Chandra, "A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation," *Expert Systems with Applications*, vol. 244, June 2024, Art. no. 122778, https://doi.org/10.1016/j.eswa.2023.122778.

[20] N. Hettikankanamage, N. Shafiabady, F. Chatteur, R. M. X. Wu, F. Ud Din, and J. Zhou, "eXplainable Artificial Intelligence (XAI): A Systematic Review for Unveiling the Black Box Models and Their Relevance to Biomedical Imaging and Sensing," *Sensors (Basel, Switzerland)*, vol. 25, no. 21, Oct. 2025, Art. no. 6649, https://doi.org/10.3390/s25216649.

[21] A. Prinston Pinto, Dattathreya, P. Prasada, and S. A. Rag, "Multimodal Coherence-Correlation Fusion of EEG and ECoG Signals for Enhanced Epileptic Seizure Detection Using Discrete Lyapunov Wavelet Transform," *IEEE Access*, vol. 13, pp. 163820–163831, 2025, https://doi.org/10.1109/ACCESS.2025.3608806.

[22] D. Muhammad and M. Bendechache, "Unveiling the black box: A systematic review of Explainable Artificial Intelligence in medical image analysis," *Computational and Structural Biotechnology Journal*, vol. 24, pp. 542–560, Dec. 2024, https://doi.org/10.1016/j.csbj.2024.08.005.

[23] V. Vimbi, N. Shaffi, and M. Mahmud, "Interpreting artificial intelligence models: a systematic review on the application of LIME and SHAP in Alzheimer's disease detection," *Brain Informatics*, vol. 11, no. 1, Apr. 2024, Art. no. 10, https://doi.org/10.1186/s40708-024-00222-1.

[24] Z. Sadeghi *et al.*, "A review of Explainable Artificial Intelligence in healthcare," *Computers and Electrical Engineering*, vol. 118, Aug. 2024, Art. no. 109370, https://doi.org/10.1016/j.compeleceng.2024.109370.

[25] S. A. Martin, F. J. Townend, F. Barkhof, and J. H. Cole, "Interpretable machine learning for dementia: A systematic review," *Alzheimer's & Dementia*, vol. 19, no. 5, pp. 2135–2149, 2023, https://doi.org/10.1002/alz.12948.

[26] M. Champendal, H. Müller, J. O. Prior, and C. S. dos Reis, "A scoping review of interpretability and explainability concerning artificial intelligence methods in medical imaging," *European Journal of Radiology*, vol. 169, Dec. 2023, Art. no. 111159, https://doi.org/10.1016/j.ejrad.2023.111159.

[27] H. C. Bharath, N. Pradeep, R. Shashidhar, P. Nooji, and J. Meghana, "An Explainable Deep Learning Model for Classification and Analysis of Alzheimer's Disease for Clinical Trust," *Engineering, Technology & Applied Science Research*, vol. 15, no. 6, pp. 28762–28768, Dec. 2025, https://doi.org/10.48084/etasr.12997.

[28] S. El-Sappagh, J. M. Alonso, S. M. R. Islam, A. M. Sultan, and K. S. Kwak, "A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease," *Scientific Reports*, vol. 11, no. 1, Jan. 2021, Art. no. 2660, https://doi.org/10.1038/s41598-021-82098-3.

[29] A. R. W. Sait and R. Nagaraj, "A Feature-Fusion Technique-Based Alzheimer's Disease Classification Using Magnetic Resonance Imaging," *Diagnostics*, vol. 14, no. 21, Oct. 2024, https://doi.org/10.3390/diagnostics14212363.

[30] F. Hajamohideen *et al.*, "Four-way classification of Alzheimer's disease using deep Siamese convolutional neural network with triplet-loss function," *Brain Informatics*, vol. 10, no. 5, Feb. 2023, https://doi.org/10.1186/s40708-023-00184-w.

[31] P. J. LaMontagne *et al.*, "OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease." medRxiv, Dec. 15, 2019, https://doi.org/10.1101/2019.12.13.19014902.

[32] Y. Mustafa and T. Luo, "Diagnosing Alzheimer's Disease using Early-Late Multimodal Data Fusion with Jacobian Maps," in *2023 IEEE International Conference on E-health Networking, Application & Services (Healthcom)*, Sept. 2023, pp. 49–55, https://doi.org/10.1109/Healthcom56612.2023.10472348.

[33] A. Massalimova and H. A. Varol, "Input Agnostic Deep Learning for Alzheimer's Disease Classification Using Multimodal MRI Images," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Aug. 2021, pp. 2875–2878, https://doi.org/10.1109/EMBC46164.2021.9629807.

[34] S. Rahman, M. M. Rahman, S. Bhatt, R. Sundararajan, and M. Faezipour, "NeuroNet-AD: A Multimodal Deep Learning Framework for Multiclass Alzheimer's Disease Diagnosis," *Bioengineering*, vol. 12, no. 10, Oct. 2025, https://doi.org/10.3390/bioengineering12101107.

[35] L. Han, "AD-Diff: enhancing Alzheimer's disease prediction accuracy through multimodal fusion," *Frontiers in Computational Neuroscience*, vol. 19, Mar. 2025, https://doi.org/10.3389/fncom.2025.1484540.

[36] G. Castellano, A. Esposito, E. Lella, G. Montanaro, and G. Vessio, "Automated detection of Alzheimer's disease: a multi-modal approach with 3D MRI and amyloid PET," *Scientific Reports*, vol. 14, no. 1, Mar. 2024, Art. no. 5210, https://doi.org/10.1038/s41598-024-56001-9.