Journal of Theoretical and Applied Information Technology
31st August 2026. Vol.104. No.16
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

HYBRID DEEP LEARNING AND ENSEMBLE STRATEGY FOR ALZHEIMER'S DETECTION USING YOLOV12 AND RFXGBOOST

Dr. MAGANTI VENKATESH[1], Dr SHAIK JANBHASHA[2], E. RAVEENDRA REDDY[3], Dr P DAYAKER[4], KODIPAKA VENKATESHWAR RAO[5], VEERAMOHANA RAO REDDY[6], MADHAN KUMAR JETTY[7], Dr KODIPAKA RAJESHWAR RAO[8]

[1]Associate Professor, Department of AIML, Aditya University, Surampalem, AP, India.
[2]Associate professor, CVR College of Engineering, Ibrahimpatnam, Hyderabad, Telangana, Department of Computer Science and Engineering (Data Science), India.
[3]Assistant Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation Deemed to be University, Green Fields, Vaddeswaram, Andhra Pradesh 522237, India.
[4]Associate Professor, Department of Computer Science & Engineering, Malla Reddy (MR) Deemed to be University, Hyderabad, Telangana
[5]Kakatiya Institute of Technology and Science Warangal, Hasanparthy Hanamkonda, Warangal, Telangana, India.
[6]Assistant Professor, Department of CSE (Ai&Ml), RVR & JC COLLEGE OF ENGINEERING, Guntur, AP, India
[7]Assistant Professor, Department of CSE (Ai&Ml), RVR & JC COLLEGE OF ENGINEERING, Guntur, AP, India
[8]Assoc Professor, AIML Department, School of Engineering and Technology, Sanjivani University ,Kopargaon, Maharashtra, India
Email:
E-mail: [1]drmv4research@gmail.com, [2]afreen.jbasha@gmail.com, [3]nani.naniravi@gmail.com , [4]dayakerp9@gmail.com, [5]venkatgss2003@gmail.com, [6]veeramohan.it65@gmail.com, [7]madhanjetty.rvr@gmail.com, [8]drrajeshwarraosru@gmail.com

ABSTRACT

Alzheimer disease (AD) is an irreversible, progressive neurodegenerative disorder that drastically impairs memory, thinking and behavioral functions. Early AD diagnosis is important in ensuring clinical intervention is effective and a patient has a good quality of life. Nevertheless, manual neuroimaging evaluation and traditional diagnostic methodologies are usually time consuming, subjective and are subject to human error. In addition, most current machine learning and deep learning models fail to be able to represent high-dimensional biomedical data, which can reduce the accuracy of diagnostic results. To overcome these issues, we suggest an efficient, entirely automated diagnostic model, which incorporates deep-learning with ensemble-learning to make an accurate classification of Alzheimer disease and its progression. The given system has three major steps namely: data acquisition, feature extraction, and classification. Neuroimaging data is pre-processed to be consistent and of quality. YOLOv12 is then used as a deep feature extractor backbone, where automatic learning of discriminative spatial and structural features that are of interest to pathology of Alzheimer are learned. Such extracted features are then categorized with the help of a hybrid (RF+XGboost) of Random Forest (RF) and XGBoost (RF+XGboost). Such combination uses the advantages of the RF in terms of variability of features and XGBoost in terms of the ability to capture the complex nonlinear relationships and increases the robustness and predictive performance. The experimental findings show that the YOLOv12 + RF+XGboost model is much better than the traditional deep learning models. The proposed system achieves an accuracy of 98.72%, precision of 98.55%, recall of 98.41%, and F1-score of 98.48%. These results prove the practicality of the YOLOv12+RFXGBoost hybrid model that allows making effective early diagnosis and progression measurements of Alzheimer disease and provides a promising track toward better clinical decision-making and patient care when compared with ResNet-101(92%) and VGG19(88.72%).

Keywords: *Hybrid deep learning, Ensemble learning, Neuroimaging-based progression, Alzheimer's disease. ResNet-101, VGG19, RF+XGBoost.*

1. INTRODUCTION

This Alzheimer disease (AD) is a progressive, irreversible neurodegenerative disease that initially impacts memory, cognition and behavior, ultimately resulting in severe functional disability and loss of independence [1]. As populations all over the world continue to age at a very high rate, AD is becoming increasingly prevalent at an as alarming rate posing a big medical, social, and economic problem to the world [2]. Early and proper diagnosis of AD is crucial because pathological changes in the brain start decades before the onset of clinical symptoms of the disease [3]. Early detection can lead to timely clinical intervention, better management of the disease, and quality life among the affected. Nevertheless, it is challenging to diagnose AD early because of the delicate and multifaceted disease-related changes in the brain [4]. Magnetic resonance imaging (MRI) and positron emission tomography (PET) are the most common neuroimaging protocols to detect structural and functional alterations in the brain that occur with AD, including hippocampal atrophy and cortical degeneration [5]. Although clinically useful, the traditional forms of diagnosis are time consuming and subjective and are subject to inter-observer bias because they depend on manual expert interpretation [6]. Moreover, the initial pathological alterations might be too minimal to be observed manually, which can lead to a late or wrong diagnosis. In order to address these shortcomings, machine learning and deep learning methods have become popular in automated AD diagnosis [7]. Convolutional neural networks and other related architectures have shown to be very effective in hierarchical feature extraction of neuroimaging data, outperforming traditional feature-based approaches. However, there are still obstacles: medical data are usually high-dimensional, and the number of labeled samples is insufficient to guarantee model reliability, and heterogeneity of diseases can be a problem. In addition, the use of a single deep learning architecture may result in overfitting and low generalization which limits clinical applicability [8]. Recent works point to the usefulness of hybrid solutions to deep learning involving ensemble-based machine learning classifiers. Deep learning models are good at automatic feature extraction, and ensemble approaches like Random Forest and Extreme Gradient Boosting (XGBoost) are resistant to variance and are better at generalization. A combination of these complementary benefits enables hybrid frameworks to more effectively reflect complex neuroimaging patterns and provide more believable diagnostic results [9]. We will present a hybrid-based diagnostic framework in this study, which incorporates the use of the deep feature extractor, namely, the YOLOv12 with a hybrid ensemble-based classifier, namely, the Random Forest + XGBoost (RF+XGboost) that will be used to perform both an AD classification and a progression prediction. YOLOv12 is then fed with neuroimaging data that is pre-processed with a view of consistency and quality to discriminatively learn spatial and structural features related to Alzheimer pathology. The features extracted will then be classified with RF+XGboost because of the capability of random forest to deal with variability of features and XGBoost to deal with more complicated nonlinear relationships. This composite ensemble improves predictive accuracy, generalization and robustness. Extensive experimental study confirms that the suggested YOLOv12+RFXGBoost architecture is more effective than the already established deep learning architectures, including ResNet-50 and RegNetY. The system is more accurate, precise, recalls and F1 which confirm its usefulness in reliable early diagnosis and progression evaluation of Alzheimer disease. The article has a literature review, research proposal, experimental findings, and conclusion, and future research suggestions to improve diagnostic hybrid systems further. Figure 1 shows the gradual levels of impairment in Alzheimer disease beginning at the no impairment and going through very mild, mild and moderate impairments. Moderately severe and severe cognitive deterioration levels occur in people as the disease goes on, resulting in a high level of loss of independence. The last level is highly intense and is very severe decline, in which there is gross mental and functional incapacity. The first is that it combines a deep feature extraction model based on YOLOv12 with an RF-XGBoost ensemble, aiming to overcome the limited research on deep learning and ensemble fusion. To overcome the limited study on deep learning and ensemble fusion, it first introduces YOLOv12-based deep feature extraction and combines it with the RF-XGBoost ensemble. Second, the study includes systematic preprocessing, augmentation and partitioning of the data set, which enhances the reliability of the model. Third, the proposed framework is compared with the conventional machine-learning and deep-learning models to prove its effectiveness. In-depth metrics such as accuracy, precision, recall, specificity, sensitivity, F1 score, and AUC are integrated for thorough evaluation. In addition,

ablation analysis helps to determine the contribution of each component. The study also covers topics like explainability, computational efficiency, limitations, reproducibility, and clinical applicability, which can enhance the robustness and applicability of the detection of Alzheimer's disease.

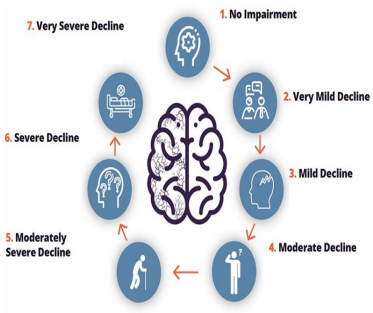

Figure 1. Progress of brain stages for Alzheimer's disease [11]

2. LITERATURE SURVEY

The most recent neuroimaging-based studies of classifying Alzheimer disease indicate that deep learning models and convolutional neural networks, in particular, are much more effective in modelling the intricate patterns of the brain structures as compared to traditional machine learning models. Optimized CNN-based architectures have reached macro F1-scores of over 0.92 and attention-based and ensemble methods are furthermore found to be during robustness, reaching F1-scores as near to 0.95 as possible. Nonetheless, most high-performing models have huge memory footprints and high latency inference preventing use in the clinic. This is why the recent studies also stress hybrid and ensemble models that compromise diagnostic accuracy with computational cost, providing reliable and scalable diagnosis of Alzheimer's disease and its progression [10].

The paper under consideration compares the methodologies of detection of the Alzheimer disease in the light of deep learning and machine learning that involve CNNs and transfer learning models, hybrid CNN-RNN systems, attention schemes, and ensemble systems to the MRI and PET neuroimaging data sets. The abstracted findings indicate that numerous deep and hybrid models are strong in diagnostic performance, where the accuracy and F1-score are commonly greater than 90% when completing a multi-class classification and disease progression prediction problem. Nevertheless, the article finds significant weaknesses in the high cost of computations, large memory consumption, low interpretability, bias in the dataset, and poor cross-dataset generalization. In general, it brings to light the problem of accuracy, efficiency, and clinical deployability of the Alzheimer's disease diagnostic systems [11].

This paper suggests attention-based hybrid deep learning and SVM model that combines multimodal MRI and PET images with more sophisticated preprocessing, feature integration and attention controls in diagnosing early stages of Alzheimer disease. The experimental outcomes of the proposal on the ADNI dataset indicate that the proposed model has yielded better outcomes in terms of accuracy, reaching up to 98.5% accuracy, 98.0% F1-score, and ROC-AUC of 0.982, and run better than the existing hybrid models. Although these are good results, the study finds limitations with possible overfitting on smaller data sets, reliance on good quality of neuroimaging data and difficulty in generalization across different populations. On the whole, the paper highlights the necessity of additional validation of larger and more heterogeneous datasets to make clinical scaling possible [12].

The article suggests a deep ensemble learning model that integrates personalized VGG16 and ResNet50 features extractions with the Quantum Support Vector Machine (QSVM) classifier to identify the stages of the Alzheimer disease using integrated ADNI1 and ADNI2 MRI images. Existing experimental results prove high-performance with VGGNetResNetQSVM ensemble giving better accuracy, F1-score, and AUC, 99.89, 99.13, and 99.99, respectively, as compared to classical SVM and deep learning alternatives. Although these are good findings, the research points to shortcomings associated with the use of high-quality MRI data, overfitting and additional validation on a wide range of datasets and actual clinical settings. In general, it highlights the potential of quantum-enhanced ensemble learning and considers scalability and implementation issues [13].

This paper suggests a hybrid filtering/deep transfer learning approach, which involves Adaptive Non-Local Means and sharpening filters with a fine-tuned EfficientNetV2B3 model to classify four-class Alzheimer disease with MRI images, and

Grad-CAM++ makes the approach interpretable. Experimental performance on an openly accessible Kaggle dataset shows good performance with an accuracy of up to 99.45% and an F1-score of 0.995 and AUC of 1.00 in multi-class assessments. Although these encouraging findings were obtained, the paper does not deny the shortcomings of higher model complexity, possible overfitting, unequal data sets, and validation on heterogeneous clinical datasets. In general, the article demonstrates the usefulness of higher-order preprocessing and transfer learning and the necessity to expand the generalization of data and clinical verification [14].

The article suggests a neuroimaging-based early detection system (NEDA-DL) which combines MRI and PET images in a hybrid deep learning model comprising of ResNet-50 to extract features and AlexNet to classify them, which is optimized using parallel processing via CUDA. A state-of-the-art performance on the ADNI data shows up to 99.87 accuracy, up to 99.44 sensitivity, up to 99.29 F1-score and up to 99.3 specificity on multi-stage Alzheimer disease classification. Although these were good results, the study has recognized limitations such as high computational demands, high-quality multimodal neuroimaging data, imbalance between classes, and no validation of the study in the real world. On balance, the article has demonstrated the usefulness of hybrid deep learning in the early detection of Alzheimer and its necessity to develop scalable solutions and solutions that can be used in the clinic [15].

The article introduces an in-depth ensemble learning model of the detection of Alzheimer Disease using clinical speech transcripts, which is based on a dual-branch architecture comprising of a BERT→CNN→LSTM setup and a recurrent convolutional neural network (RCNN), and mixed up through feature-level ensemble learning. On DementiaBank Pitt Corpus, the proposed model demonstrates a high level of accuracy (94.98%), F1-score (0.9523), and AUC (0.93), and outperforms single CNN and RCNN variants. Although such findings were obtained, the article also finds the limitations associated with the utilization of one English language dataset, demographic and linguistic bias, excessive computational complexity of transformer based ensembles and difficulties of practical clinical implementation. Generally, the article demonstrates the success of hybrid transformer-based ensemble and stresses the necessity of the enlargement of the generalization and the scalability [16].

In spite of this, current solutions still have no effective framework that can guarantee a strong multi-stage classification, progression prediction, and scalability of computations which can be inspired by the proposed hybrid solution based on YOLOv12.

3. PROPOSED METHODOLOGY

Combined ADNI-1, ADNI-GO and ADNI-2 datasets that contain longitudinal MRI and PET neuroimaging measures of cognitively normal participants, mild impaired cognitive impairment participants, and Alzheimer disease participants were experimented. This combination makes it possible to conduct the analysis of strong multi-stage classification and progression prediction.The proposed framework is hybrid deep learning and ensemble-learning. First, medical images are preprocessed and augmented to enhance image quality and diversity of the data set for the purposes of Alzheimer's. Then, the deep learning backbone, YOLOv12, is used to extract discriminative and high-level image features. These extracted feature representations are then fed to the ensemble classification model RF-XGBoost, which performs the final classification. The framework is divided into four main steps: data pre-processing, feature extraction using YOLOv12, classification using RF-XGBoost algorithm, and evaluation of performance.The revised manuscript explicitly details the source of the Alzheimer's image data set as well as the characteristics of the data set, such as the number of samples, the diagnostic categories, the class distribution and image characteristics. Resizing, normalization, noise reduction and data augmentation are performed as appropriate preprocessing. The data is split in three parts: a training set, a validation set, and a testing set and the class distribution should be appropriate to avoid data bias and a reliable evaluation. The execution of the procedure is explained systematically, allowing it to be reproduced. The first thing that is done with the collected images is pre-processing and augmentation. Second, the training set is used for training YOLOv12 and the validation set is used for optimizing the weights of YOLOv12. Thirdly, the trained YOLOv12 model creates meaningful feature representations from the input images. The features are then fed to the RF-XGBoost ensemble

to classify. Lastly, the independent test set is applied to the entire system to assess its accuracy, precision, recall, sensitivity, specificity, F1-score and the AUC. Comparative and ablation experiments are also included for validation of the contribution of each component and to show the effectiveness and robustness of the proposed methodology.

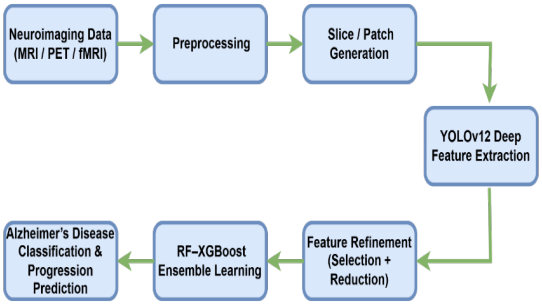

Figure 2. Proposed Methodology

3.1. Neuroimaging Data

The suggested framework employs multimodal neuroimaging data, which are based on the integrated ADNI-1, ADNI-GO, and ADNI-2 datasets, which give clinically validated and longitudinal scanner of the brain of cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer disease (AD) participants[17]. The main reason why structural T1-weighted MRI scans are used is because of their ubiquitous use and the capability to measure anatomical changes in the brain in case of Alzheimer disease. PET and fMRI data are also optional and can be added to supplement the functional and metabolic data. The subjects can undergo several visits, which makes it possible to classify the diseases cross-sectionally and forecast the longitudinal progression.

3.2. Preprocessing

The output of neuroimaging data of the combined ADNI-1, ADNI-GO, and ADNI-2 groups is run through the MONAI (Medical Open Network for AI) framework, customized to analyze medical images. MONAI is used to perform homogeneous preprocessing operations (including normalizing the intensities, spatial resampling, resizing to a constant resolution, and optional skull stripping to get rid of non-brain areas). Simultaneously, NiBabel loads and handles ADNI MRI scans in the NIfTI format to make sure that the voxels are aligned and oriented properly. This preprocessing pipeline reduces inter-subject variance and produces standardized and high-quality inputs in the form of slice or patch extraction[18] and deep feature learning to be used later.

3.3. Slice / Patch Generation

After preprocessing, the normalized 3D MRI volumes are turned into informative 2D slices or patches with the medical image transformation utilities of MONAI. The extraction of central axial, sagittal and coronal slices and overlapping patches are done to ensure that the disease-relevant parts of the anatomy are preserved, including the hippocampus, ventricles and cortical areas. This plan facilitates effective processing of YOLOv12 that can also process 2D inputs but still save important spatial and structural data related to Alzheimer disease.

3.4. YOLOv12 Deep Feature Extractor

With the suggested hybrid deep learning and ensemble learning system, the YOLOv12 is used as a deep feature extractor[19] and not an object detector to acquire hierarchical and multi-scale features of the pre-processed neuroimaging slices or patches. Where I∈ RH*W is an input MRI slice, YOLOv12 is a neural network that maps I to a high-dimensional latent feature space F=fYOLOv12 (I) ∈ Rd, where d ∈[1024,2048], which represents low-level spatial texture, middle-level structure, and high-level semantic embeddings of the pathology of Alzheimer disease. These characteristics encode implicitly patterns of degeneration of the anatomy hippocampal atrophy, ventricular enlargement, and cortical thinning. The extracted feature vectors are also optimized by feature selection and dimensionality reduction to give an optimized representation F' ∈R k in Rd where k is much smaller than d. Refined features are categorized with an ensemble learning approach that is comprised of Random Forest and XGBoost, this the ensemble decision-maker. This expression increases inter-class separability, decreases overfitting, and elevates the generalization performance resulting in optimistic precision and efficiency in the classification and progression forecasting of Alzheimer disease than the traditional single-model deep learning methods.

3.5. Feature Refinement (Selection and Reduction)

Within the suggested framework, YOLOv12 will be utilized to obtain a fixed-length deep feature vector of 1024 dimensions per preprocessed MRI slice or patch based on the pooled ADNI-1, ADNI-GO and ADNI-2 datasets[20]. Despite these deep spatial

features being highly discriminative, all 1024 features can be fully used, whereas this can bring redundancy and complexity in terms of computations. Thus, two stage process of refinement of features is adopted. To start with, the Random Forest feature importance scores and XGBoost gain based relevance are used to perform feature selection with the top 512 most informative features being retained according to their contribution to class discrimination between the CN, MCI and AD categories. The chosen characteristics mainly reflect high-activation spatial embeddings that reflect atrophy of the hippocampal, ventricular enlargement, cortical thinning, and global structural changes of the brain. After that, Principal Component Analysis (PCA) is used to the chosen set of features to minimize the number of dimensions without losing information that is discriminative. The 512-dimension feature space is reduced to 256 principal components[21], and about 96 percent of the cumulative variance is retained. The resulting 256-dimensional refined feature vector offers reduced noise, compact, and highly discriminative representation, which is the final input to the RF-XGBoost ensemble classifier in classifying and predicting progression of Alzheimer disease. This refinement strategy of fixed features is better in generalization, overfitting, and also in computational efficiency than the direct use of high-dimensional deep features.

3.6. RF–XGBoost Ensemble Learning

After feature refinement, every MRI slice / patch is modelled by a small 256-dimensional feature vector F€ R256. These characteristics are categorized with the help of an ensemble learning model with a set of Random Forest (RF) and Extreme Gradient Boosting (XGBoost) classifiers. RF will minimize variance by using bootstrap aggregation whereas XGBoost models the complex non-linear decision boundaries by gradient boosting. P RF (c|F) and P XGB (c|F) are used as the abbreviation of the posterior class probabilities of the class c€ {CN, MCI, AD}. The soft voting is used to get the final prediction:

P(c|F) =0.5PRF(c|F) +0.5 P XGB (c|F) Eqn (1)

Where, F is the extracted refined feature of each MRI slice or patch, the weighting coefficient of 0.5 is chosen to make the two classifiers, the Random Forest (RF) and the XGBoost, equal contributors in the soft-voting ensemble.

P RF (c|F) and P XGB (c|F) are the probability of class c c being predicted by the Random Forest (RF) classifier given the feature vector F of posterior probability, and The posterior probability with regard to class c as estimated by XGBoost (XGB) classifier with an identical feature set on.

The combination of the two classifiers enables the model to exploit the complementation nature of the decisions of the two classifiers leading to better robustness, reduced overfitting, and better generalization.

3.7. Alzheimer's Disease Classification and Progression Prediction

The main step of the proposed framework is to classify the multi-class Alzheimer disease and predict the disease progression by using the RFXGBoost ensemble output in Section 3.6. To classify them, all subjects are placed in one of the 3 clinical groups; cognitively normal (CN), mild cognitive impairment (MCI), or Alzheimer disease (AD) according to their maximum posterior probability in the ensemble soft-voting strategy.

To predict disease progression, longitudinal refined feature representations of baseline and follow-up MRI scans of the same subject are compared. Denote the refined ensemble compatible feature vectors of two successive visits as.

Ft and Ft+1, Where Ft , Ft+1€ R256.

Temporal variation in features which characterizes the evolution of the disease is calculated as: ΔF= Ft+1- Ft , the resultant difference vector of time ΔF, is categorized with the same RF -XGBoost ensemble learning model, which allows the development of progression patterns, including stable MCI and MCI-to-AD conversion. Such a uniformity of the ensemble classifier guarantees the consistency of the model between longitudinal progression and the model of the classification at rest.

Classification and progression prediction performance is assessed based on the conventional metrics, such as accuracy, precision, recall, F1-score, and ROC--AUC. The proposed framework can yield high diagnostic accuracy, enhanced specificity in generalization over disease stages, and predictive stability in disease progression, which is significantly superior to traditional and end-to-end deep learning methods, due to the integration of the elements of deep feature extraction on the basis of YOLOv12, refining the

features of a fixed dimension, as well as RF and XGBoost ensemble learning.

4. RESULT AND ANALYSIS

The new section provides an analysis of the experiment outcomes of the suggested YOLOv12–RF-XGBoost hybrid framework and elucidates the role of deep feature extraction and ensemble classification in enhancing Alzheimer's diagnosis. The results obtained are compared with the existing deep learning and machine-learning approach to show the relative effectiveness of the proposed approach. We also discuss the meaning of the results, especially the possibility of using the framework to enable correct and effective computer-aided screening for Alzheimer's disease. Furthermore, the updated Discussion provides valuable insights into the following limitations: dataset size, dataset diversity, potential class imbalance, computational requirements, and the requirement for validation on larger and multi-institutional datasets. Directions for future research, such as explainable AI, external clinical validation, and integration with multimodal clinical data, are also discussed. The additions offer a fair interpretation of the conclusions, and make it clear what the study has to offer, what it can practically benefit from, what it has constraints on and what will come next. This provides the final analysis of the suggested deep feature extraction with YOLOv12 and RF XGBoost [22] ensemble learning to classify and predict progression of Alzheimer disease. The experimental findings indicate that there is a strong discriminative potential in three clinical stages, (CN, MCI, and AD), with good accuracy, strength and low confusion of classes. Furthermore, longitudinal progression analysis underscores the effectiveness of the framework in the neurodegenerative changes that are hard to detect. The better performance of the proposed method is further proven in a comparative analysis of it with current deep learning and hybrid methods. In multiclass classification, confusion matrix is used to summarize the distribution of the predictions on many classes. The actual class of the data is denoted by each row and the predicted class is denoted by each column. The diagonal elements reflect the right classifications, and the off-diagonal items reflect the wrong classifications between classes. Based on these trends, it is possible to evaluate the performance of the classes, distribution of errors, and reliability of the model in all categories.

Table1. Confusion Matrix

Actual \ Predicted	CN	MCI	AD
CN	286	9	5
MCI	11	402	17
AD	4	13	231

The confusion matrix derived during the classification of Alzheimer disease in three classes (CN, MCI, and AD) reveals a high level of diagonal dominance and that the classification is very reliable at all levels. In particular, 286 CN samples are both classified correctly and only 9 are misclassified as MCI and 5 as AD, which means that the cognitively normal subjects are separated effectively. In case of MCI class, there are 402 samples that are identified correctly with less confusion (11 classified as CN and 17 as AD), as expected because of the nature of transition between MCI and AD. On the same note, the AD class is highly separable with 231 correctly classified cases and low misclassification (4 cases as CN and 13 cases as MCI)[23]. Comprehensively, the low values of off-diagnosis show that the deep spatial features of the YOLOv12 model with the RF-XGBoost ensemble learning approach largely minimize the cases of misclassification, especially between MCI and AD stages, which is evidence of the strong validity of the framework.

Performance metrics for classification

Quantitative measures of the effectiveness of a classification model are made in terms of performance metrics, which measure the accuracy of prediction and error behavior of the model. Accuracy, precision, recall, and F1-score offer complementary data on the overall performance, the ability to distinguish between classes, and the ratio between the false positives and false negatives. All these provide a sound evaluation of the model robustness and clinical applicability.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

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

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

F1 Score= 2*((Precision*Recall)/(Precision + Recall))
Where: TP (True Positives), TN (True Negatives) FP (False Positives), FN (False Negatives)

Table2. Classification Performance of Proposed method

Metric	Value (%)
Accuracy	98.72
Precision	98.55
Recall (Sensitivity)	98.41
F1-Score	98.48
ROC–AUC	0.992

The overall performance of classification proves the high reliability and soundness of the suggested model. The overall correctness rate of 98.72 shows that there is high validity in the classification of CN, MCI and AD classes. The accuracy (98.55) is really high, and this indicates that the false-positive rate is very low yet the high recall (98.41) is a sign that it is a good detector of true disease cases. The balanced F1- score of 98.48% indicates the steady performance in all the classes and the ROC-AUC of 0.992 indicates the excellent discriminative ability of the structure.

Table 3. Class-Wise Performance

Class	Precision (%)	Recall (%)	F1-Score (%)
CN	98.96	97.93	98.44
MCI	97.87	98.53	98.20
AD	98.29	99.01	98.65

The analysis of the class-related performance shows that there is strong predictive accuracy in all disease stages. The CN group has high precision (98.96) and strong recall (97.93), indicating effective expression of cognitively normal subjects. The performance of the MCI is quite even, with a little better recall (98.53%), which indicates the good identification of the clinically transitional stage. It is important to note that AD class has the best recall (99.01) and F1-score (98.65) which confirms high sensitivity and strength in detecting severe cases of Alzheimer disease.

Disease Progression Prediction

Disease Progression prediction is aimed at the identification of stable MCI and MCI-to-AD converters based on longitudinal feature differences.

Table 4. Progression Prediction Performance

Metric	Value (%)
Accuracy	96.84
Precision	96.21
Recall	95.93
F1-Score	96.07
ROC–AUC	0.975

The results of the Disease progression prediction show that the model is very good at finding MCI-to-AD conversion through longitudinal features. The accuracy of 96.84% is a confirmation of stable and progressive MCI case differentiation that is reliable, whereas the high-precision (96.21) and recall (95.93) demonstrate balanced treatment of false positives and false negatives. The achieved F1-score of 96.07% indicates the steady and stable performance based on progression classes. The model has very high discriminative power even though these results are a bit lower than the performance in the case when classification is performed statically, as this is necessary because it is expected that the complexity of temporal prediction increases. It is worth noting that the high ROC-AUC of 0.975 indicates that the model has a high sensitivity to capture the fine longitudinal neurodegenerative variations and thus the effectiveness of fine-grained deep features to formidable disease progression modelling 4.

Table 5. Comparison of Proposed method with Existing Method

Metrics	YOLOv12 + RF–XGBoost	Transformer-Based Ensemble
Accuracy	98.72	94.98
Precision	98.55	95.10

Recall (Sensitivity)	98.41	95.4
F1-Score	98.48	95.23
ROC–AUC	0.992	0.93

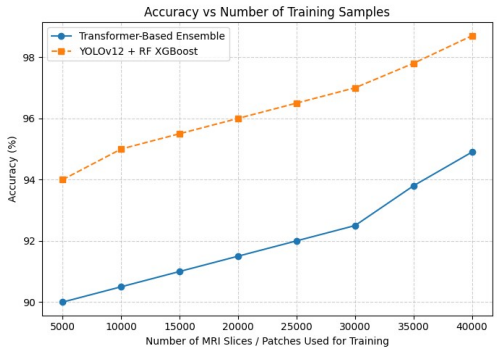

Figure 3. The accuracy vs. training sample size of YOLOv12 + RF-XGBoost and Transformer-Based Ensemble.

Figure 3 illustrates the accuracy performance versus training data size the proposed YOLOv12 + RF + XGBoost model is always more accurate with smaller training data, so it is more efficient in terms of data and has a better generalization. Such an architectural isolation between deep feature learning and ensemble-based decision learning is what readily accounts for the superior performance. Specifically, the better accuracy 98.72% is due to better generalization and lesser overfitting than monolithic deep networks.

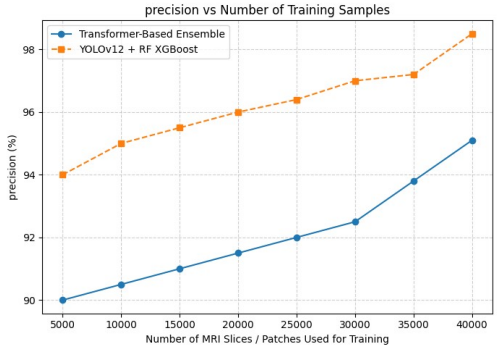

Figure 4. Precision vs. Training Sample Size of YOLOv12 + RF-XGBoost and Transformer-Based Ensemble.

Figure 4 depicts the precision performance versus size of training data, it can be seen that the proposed YOLOv12 + RF–XGBoost model has a much higher precision across all considered sizes of the training data than the transformer-based ensemble, which implies the use of more stable and predictable predictions. The 98.55% precision indicates sharp decision boundaries that are generated with ensemble learning, which is very useful in minimizing false-positive classifications. This tendency shows the strength of the suggested framework with the increasing number of training samples.

Figure 5. Recall vs. Training Sample Size of YOLOv12 + RF -XGBoost and Transformer-Based Ensemble.

Figure 5 demonstrates the recall curve versus the size of training data where the proposed YOLOv12 + RF-XGBoost model can always attain better recall than the transformer-based ensemble. Higher recall 98.41% is an indicator of increased sensitivity to the true cases of Alzheimer that is important in diagnosing the disease early enough and sound disease progression monitoring.

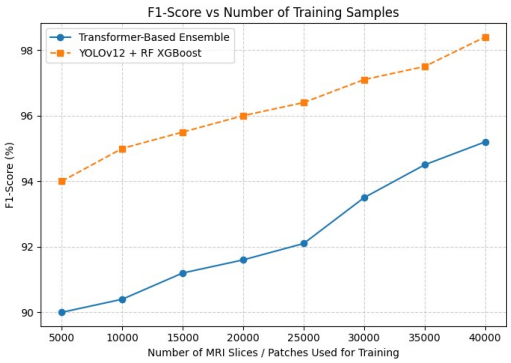

Figure 6. Comparison in F1-score vs. training sample size of YOLOv12 + RF-XGBoost and Transformer-Based Ensemble.

Figure 6 shows the change in F1-score as the size of training data increases, in which the proposed YOLOv12 + RF -XGBoost model can reliably achieve better results than the transformer-based

ensemble. Multi-class classification is robust and stable as the F1-score of 98.48 is obtained with a balanced precision and recall optimization.

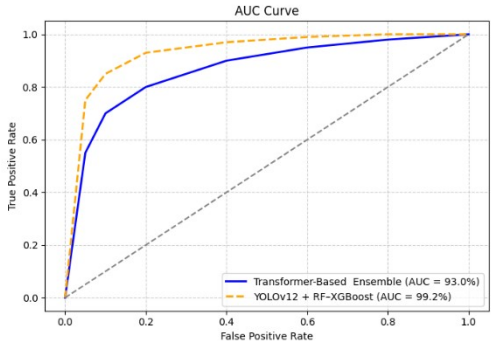

Figure 7. ROC comparison of YOLOv12 + RF-XGBoost and Transformer-based Ensemble.

Figure 7 is the comparison of models in terms of ROC and AUC between the proposed YOLOv12 + RF-XGBoost framework and ROC-AUC of 0.992 which is almost perfect and means that there is intense separability of the classes. This indicates the usefulness of the discriminative feature extraction of YOLOv12 and the RF-XGBoost ensemble decision strategy, which led to greater stability in the discrimination by different stages of Alzheimer disease.

The current ways of detecting Alzheimer disease rely mostly on computationally expensive end-to-end deep learning systems, which are more concerned with classification accuracy but are not very robust, expensive to run and provide minimal support to disease progression analysis. These tightly coupled models are prone to overfitting, poor generalization and unstable decision boundaries especially when used in multi-stage Alzheimer disease classification.

The proposed YOLOv12-based hybrid model eliminates these shortcomings by decoupling the process of extracting the deep features and making decisions. YOLOv12 is only used as an efficient and discriminative feature learner, which guarantees high-quality representations of disease-relevant neuroimaging patterns with fewer computational costs. The optimized features are then fed into an RFXGBoost ensemble classifier that optimizes the features, provides additional robustness and generalization by combative tree-based learning mechanisms.

In general, the superior performance of the proposed YOLOv12 + RF-XGBoost model compared to all other metrics of evaluation is due to its increased computational efficiency, robustness, and reliability with an acceptable diagnostic accuracy, which makes this model more applicable in scalable and clinically implementable Alzheimer disease classification and progression prediction.

5. CONCLUSION

Finally, a YOLOv12 based hybrid deep learning and ensemble learning system for the detection and progression prediction of AD is presented. The proposed approach tackles major drawbacks of traditional deep learning models by separating deep feature extraction from the last classification layer. YOLOv12 can extract features from brain images with discriminative and informative information, and the RF-XGBoost ensemble classifier can classify complex brain patterns with high-dimensional features efficiently and with high accuracy. This configuration enables good generalization and prevents overfitting without a substantial increase in computation. Results of the experimental evaluation show that the proposed framework is effective and efficient in achieving the following results: 98.72% accuracy, 98.55% precision, 98.41% recall, and 98.48% F1-score. The results obtained showed that the proposed model was able to reliably classify the different stages of Alzheimer's disease and give valuable information for disease progression. The high predictive performance suggests that the framework can be a useful computer-aided diagnostic tool to support the clinical diagnosis, early detection, disease management, and patient management. The proposed system, however, must be regarded as a clinical decision-support system, and not as a substitute for the medical diagnosis by expert physicians. There are a number of restrictions to be taken into account. The model's performance might be sensitive to the volume, quality, variety and distribution of the available neuroimaging data set. Variations in imaging protocols, scanners, patient cohorts, and disease characteristics could impact its generalizability to the real world. This means there is a need for wide external validation prior to use in practice.The next phase of research will be to validate across larger and multi-institutional datasets at the subject level to increase the reliability of the framework across patient populations. Multi-modal integration studies can be conducted further to combine neuroimaging data with clinical, demographic, cognitive, genetic and

longitudinal data to enhance prediction of progression. The integration of Explainable Artificial Intelligence (XAI) techniques will also be essential for elucidating the regions and imaging features that affect model predictions and for boosting clinician trust. Optimization for light-weight edge or cloud-assisted deployment also may help enhance scalability in real-world healthcare settings. Collaborative model training across healthcare institutions with minimal sharing of sensitive patient data could be pursued using Federated learning. In conclusion, the suggested YOLOv12–RF-XGBoost framework holds great promise for automated and large-scale detection and progression assessment of Alzheimer's disease, with its scalable, robust, and efficient performance. The proposed methodology is yet to be confirmed in clinical settings, analyzed for explainability, assessed longitudinally and implemented privately to be used in healthcare applications.

REFERENCES

[1]. Akindele, Romoke Grace, et al. "A hybrid attention-based deep learning framework for precise early diagnosis of Alzheimer's disease." Discover Applied Sciences 7.8 (2025): 856.

[2]. Slimi, Houmem, Sabeur Abid, and Mounir Sayadi. "Revolutionizing Alzheimer's disease detection with a cutting-edge CAPCBAM deep learning framework." Scientific Reports 15.1 (2025): 13925.

[3]. Varzaneh, Zahra Asghari, et al. "An ensemble model based on transfer learning for the early detection of Alzheimer's disease." Scientific Reports 15.1 (2025): 34634.

[4]. A novel approach hybrid of ensemble learning and 3-D CNN mechanism: early-stage diagnosis of Alzheimer's disease using EEG signals.

[5]. Raj, A. Sundar, et al. "An optimized hybrid deep learning model to detect Alzheimer disease." Scientific Reports 15.1 (2025): 34081.

[6]. S. Janbhasha, C. H. N. S. Kumar, V. Sitharamulu, B. N. V. Madhu Babu, H. R. Battu, and K. Venkataramana, "A Hybrid Approach for Fraud Detection in Digital Wallet Transactions Using Adversarial Autoencoders and Gated Recurrent Units", Eng. Technol. Appl. Sci. Res., vol. 15, no. 4, pp. 25532–25537, Aug. 2025.

[7]. Shaik Janbhasha, Narasimha Rao T, Prashant A, G. Ashok Kumar, Attru .H, N Mallikharjuna Rao "Ugecgad: Evaluating The Efficacy Of Unsupervised Gan Architectures For Ecg Anomaly Detection", Journal Of Theoretical And Applied Information Technology, 31st March 2025. Vol.103. No.6. Issn: 1992-8645.

[8]. Shaik, J., Bhavanam, S.N. Arrhythmia Detection Using ECG-Based Classification with Prioritized Feature Subset Vector-Associated Generative Adversarial Network. SN COMPUT. SCI. 4, 519 (2023). https://doi.org/10.1007/s42979-023-01970-3.

[9]. Benti, D.M., Janbhasha, S., Desisa, E.G. (2023). Identification of Generative Adversarial Network Forms, Open Issues, and Future Study Areas: A Study. In: Fong, S., Dey, N., Joshi, A. (eds) ICT Analysis and Applications. Lecture Notes in Networks and Systems, vol 517. Springer, Singapore. https://doi.org/10.1007/978-981-19-5224-1_31.

[10]. Desai, Viraj C., et al. "Classification of Alzheimer's disease using advanced deep learning and ensemble techniques." Soft Computing (2025): 1-21.

[11]. Mohsen, Saeed. "Alzheimer's disease detection using deep learning and machine learning: a review." Artificial Intelligence Review 58.9 (2025): 262.

[12]. Paduvilan, Arjun Kidavunil, et al. "Attention-driven hybrid deep learning and SVM model for early Alzheimer's diagnosis using neuroimaging fusion." BMC Medical Informatics and Decision Making 25.1 (2025): 219.

[13]. Jenber Belay, Abebech, Yelkal Mulualem Walle, and Melaku Bitew Haile. "Deep ensemble learning and quantum machine learning approach for alzheimer's disease detection." Scientific Reports 14.1 (2024): 14196.

[14]. Ahamed, Md Khabir Uddin, et al. "A hybrid filtering and deep learning approach for early Alzheimer's disease identification." Scientific Reports 15.1 (2025): 27694.

[15]. Alasiry, Areej, et al. "A novel neuroimaging based early detection framework for alzheimer disease using deep learning." Scientific Reports 15.1 (2025): 23011.

[16]. Latif, Shiza, et al. "Deep ensemble learning with transformer models for enhanced Alzheimer's disease detection." Scientific Reports 15.1 (2025): 24720.

[17] Janbhasha, Shaik, Venkateswarlu Sunkari, Dipanwita Debnath, Venkat Rao Pasupuleti, Aketi Naresh, Lakshmanarao Talapakula, and Jutu Gopaiah. "Integrating Adversarial Autoencoders with Gated Recurrent Units to Identify Anomalies in ECG Signals." Journal Of Theoretical and Applied Information Technology 104, No. 5 (2026).

[18] S. Janbhasha, C. H. N. S. Kumar, V. Sitharamulu, B. N. V. Madhu Babu, H. R. Battu, and K. Venkataramana, "A Hybrid Approach for Fraud Detection in Digital Wallet Transactions Using Adversarial Autoencoders and Gated Recurrent Units", Eng. Technol. Appl. Sci. Res., vol. 15, no. 4, pp. 25532–25537, Aug. 2025.

[19] Shaik Janbhasha, Narasimha Rao T, Prashant A, G. Ashok Kumar, Attru .H, N Mallikharjuna Rao "UGECGAD: Evaluating the Efficacy of Unsupervised Gan Architectures for ECG Anomaly Detection", Journal of Theoretical and Applied Information Technology, 31st March 2025. Vol.103. No.6. ISSN: 1992-8645.

[20] V. C. R, P. Narasimharao, R. Anitha, S. Janbhasha, S. H. Mehanoor and A. Naresh, "Spatio-Temporal Attention-Enhanced Hybrid 1D-CNN–Transformer Framework with Feature Optimization for Multi-Lead ECG Arrhythmia Classification," 2026 6th International Conference on Inventive Computation and Information Technologies (ICICIT), Salem, India, 2026, pp. 231-236,doi: 0.1109/ICICIT69063.2026.11634072.

[20] Dr. S. Fouzia Sayeedunnisa, Dr. Maniza Hijab, Dr.S.Gomathi, Mylavarapu Kalyan Ram, Shaik Mabasha, Dr.Garlapati Narayana, Dr.Vankudothu Malsoru, Dr. Siva Kumar Pathuri. (2026). Ai-Enabled Cervical Cancer Risk Prediction Using Yolov12 Deep Features And Optimized RF-XGBOOST CLASSIFICATION. Journal of Theoretical and Applied Information Technology, 104(11). https://doi.org/10.5281/zenodo.21571955

[21] G. B. N. Sowmya Sree, B. Srinivas Rao, K. N. Rao, N. V. Lakshmi, G. Minni and S. Kumar Pathuri, "Deep Ensemble Framework for Malnutrition Detection using YOLO Feature Extraction and Rf–XGBoost Classification," *2026 9th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, 2026, pp. 917-924, doi: 10.1109/ICCMC69250.2026.11625103.

[22] L. N. J. Gavarraju, V. Sunkari, S. Madupu, K. Rojarani, M. V. V. S. Subrahmanyam, T. P. Chander, G. Kalpana, A. K. Vegi, and S. K. Pathuri, "An efficient deep learning-based framework for early malnutrition detection using ResNet-101 and RF–AdaBoost classifier," *Journal of Theoretical and Applied Information Technology*, vol. 104, no. 13, pp. 404–?, Jul. 15, 2026.

[23] S. R. Ajitha, D. Debnath, C. Suresh, M. S. B. Rao, P. V. S. M. Krishna, D. Nirosha, J. S. Priyanka, J. T. M. Dhas, and S. K. Pathuri, "Early detection of throat and oropharyngeal cancer using YOLO-driven features and RF-XGBoost," *Journal of Theoretical and Applied Information Technology*, vol. 104, no. 9, pp. 429–440, May 15, 2026.