DOI: 10.21917/ijivp.2024.0455

EARLY MODERATE COGNITIVE IMPAIRMENT CLASSIFICATION USING ENSEMBLE OF DEEP LEARNING CLASSIFIERS

A.S. Shanthi

Department of Artificial Intelligence and Data Science, NGP Institute of Technology, India

Abstract

The early detection of moderate cognitive impairment (MCI) allows for timely management, which in turn leads to improved patient outcomes. Creating a categorization system through the use of an ensemble of deep learning classifiers is the approach that this research takes in order to fulfil this goal. Using dimensionality reduction and feature selection as our primary concern, we perform preliminary processing on a diverse dataset that contains information on demographics, cognitive test scores, and brain imaging. Within the framework of an ensemble method, many deep learning architectures are utilised, with each design concentrating on a particular aspect of cognitive impairment prediction. A training and fine-tuning process is performed on each individual model on its own unique training set before the ensemble is constructed and evaluated based on a comprehensive set of performance criteria. The research presented here contributes to the development of a robust ensemble model for early MCI classification by integrating multiple different deep learning algorithms. This results in an improvement in diagnostic accuracy. The process ensures that the models may be utilised in therapeutic settings and that they are comprehensible to other individuals. When contrasted with models that are used on their own, the ensemble demonstrates superior performance, exhibiting greater memory, precision, and accuracy.

Keywords:

Cognitive Impairment, Deep Learning, Ensemble Classification, Early Detection, Medical Imaging

1. INTRODUCTION

Cognitive impairment is one of the most significant challenges facing the field of public health, particularly when it is mild to moderate in severity. With quick diagnosis and treatment, patients can experience significant improvements in both their outcomes and their quality of life. It is possible to make the diagnosis of cognitive impairment more accurate and efficient by utilising modern machine learning techniques, particularly deep learning. This can be accomplished using medical imaging, cognitive testing, and the collection of demographic data.

The correct classification of mild to moderate cognitive impairment involves a variety of challenges, despite the potential benefits that may be obtained from doing so. Because of the multifaceted character of cognitive decline, the requirement for interpretability in medical settings, and the variability of cognitive test results, there are a number of challenges that call for a method that is both careful and sophisticated.

The primary objective of this research is to develop a classification system that is reliable and easy to comprehend for mild to moderate cognitive impairment by making use of a combination of deep learning classifiers. In order to assist healthcare providers in making informed decisions, we require a model that is capable of demonstrating a high level of accuracy while also providing insight into the decision-making process.

A wide range of demographic data, the results of cognitive tests, and medical imaging data will be preprocessed and curated. To increase the relevance of the input characteristics by employing feature selection and dimensionality reduction strategies during the improvement process. It is recommended to use an ensemble method and integrate a number of different deep learning architectures in order to gather information about the various aspects of cognitive impairment. To ensure that every classifier maximises its potential by training and tuning on a particular training set, and to guarantee that this is accomplished. To construct an ensemble model with an emphasis on generalizability and interpretability, using appropriate fusion procedures to do so.

An ensemble of interconnected deep learning models is used in this study to categorise mild to moderate cognitive impairment, which is a novel approach that breaks new ground. An intriguing aspect is the emphasis placed on interpretability, which acknowledges the significance of healthcare providers having faith in and comprehending the judgements made by the model. The seamless combination of demographic information, cognitive test results, and medical imaging data brings about an additional level of complexity that represents the difficulties that are encountered in the real world while diagnosing cognitive impairment. All of these factors contribute to the overall complexity.

The primary contribution that this research makes is the development of an ensemble model that addresses the deficiencies that are inherent in individual classifiers. A number of different deep learning architectures are used in this model, which addresses the multifaceted character of cognitive decline. This is the source of the model's strength. The generalizability and interpretability of the model have been validated, which makes it ideal for application in clinical settings and boosts the confidence of healthcare practitioners in such systems as well as their usage of them.

2. RELATED WORKS

Investigations have been conducted on a number of occasions regarding the utilisation of deep learning techniques for the purpose of discovering cognitive abnormalities. [6] developed convolutional neural networks (CNNs) to identify between individuals with normal cognitive function and those with mild cognitive impairment (MCI). This was accomplished through the use of functional brain scans during the training process. The findings indicated that there was potential. In addition, [7] utilised recurrent neural networks (RNNs) in conjunction with longitudinal clinical data in order to identify patterns that exhibit indicators of cognitive deterioration over the course of time. Despite the fact that individual models [8] have demonstrated potential, the ensemble technique has not been subjected to extensive research in relation to cognitive impairment.

Ensemble learning has been shown to be more effective than individual models in a number of healthcare applications, which have proved its performance. A CNN ensemble was utilised by [9] in order to accurately detect cancers in radiological images, which is a discipline that falls under the umbrella of medical image analysis. The concept of combining multiple models in order to improve the overall performance of the system is the source of inspiration for our method, which is used in the field of cognitive impairment classification.

The capacity of machine learning models to be interpreted is a critical factor in the application of these models in healthcare settings. The need of model interpretability in healthcare applications was brought to light by [10], who emphasised the requirement for models to provide physicians with insights that can be applied in the real world. This is expanded upon by our research, which makes the interpretability of the ensemble model for diagnosing cognitive deficits the primary objective of consideration.

Studies that have been done in the past have demonstrated that it can be fairly difficult to appropriately identify cognitive impairment, particularly in the beginning stages of treatment. The need for multimodal approaches that take into consideration both imaging data and cognitive scores was brought to light by [10]. These researchers also raised attention to the fact that the outcomes of cognitive tests can differ from person to person. We have created an ensemble approach to data integration that takes into account the intricacies of cognitive impairment in order to facilitate the process of overcoming these challenges.

It has been established that the accuracy of categorising cognitive impairment can be enhanced by combining demographic data with medical imaging data. [11] merged demographic factors with structural brain imaging data in order to improve the predictive performance of Alzheimer's disease categorization. This was done in order to improve the accuracy overall. Through the incorporation of demographic information into cognitive test scores, our research takes this concept to a higher level, resulting in a more comprehensive evaluation of the dataset.

When it comes to healthcare, it is essential that the machine learning models that are deployed have the ability to generalise. [12] was the first person to propose the use of k-fold cross-validation as a method of evaluation that is completely trustworthy. We are able to ensure that our ensemble model for cognitive impairment categorization is trustworthy across a variety of datasets by employing cross-validation techniques. This, in turn, boosts the model's validity when applied to circumstances that occur in the real world.

Our research distinguishes out from others because it addresses the challenges of cognitive impairment categorization by presenting a comprehensive ensemble method that makes use of deep learning models. This method is made possible by the fact that it brings together pertinent studies. As a result of the fact that it places a strong emphasis on the interpretability and incorporation of numerous data modalities, our research is considered to be at the forefront of expanding the field towards practical and reliable applications in clinical settings.

3. PROPOSED METHOD

The identification of mild to moderate cognitive impairment is accomplished through the utilisation of a collection of fifty distinct deep learning classifiers. The ensemble is comprised of a number of different designs, ranging from those that are more conventional, such as the Multilayer Perceptron (MLP) and the Convolutional Neural Network (CNN), to those that are more cutting-edge, such as the Transformer, BERT, GPT, and T5. These models were selected because they have a proven track record of success in a variety of machine learning tasks and have the ability to capture distinct aspects of cognitive impairment.

For the purposes of training and evaluation, a dataset that has been carefully curated and contains demographic information, results of cognitive tests, and medical imaging data is employed. The preprocessing process includes a number of different steps, including the recovery of lost data, the standardisation of numerical features, and the extraction of usable information. Feature selection approaches are used to choose the features that are the most informative, while dimensionality reduction techniques, such as principal component analysis (PCA), are utilised to increase computing performance.

The classifiers that make up the ensemble are trained independently on a particular training set, and then subsequent hyperparameters are used to fine-tune their performance. For the purpose of fostering diversity within the ensemble and achieving a comprehensive understanding of the input data, models with a variety of topologies are selected. For the purpose of achieving model fusion, the various classifiers are integrated in a manner that is weighted. The ensemble will be able to make advantage of the strengths of each model while simultaneously lowering the deficiencies of the models.

Our proposed strategy places a significant emphasis on its ability to be interpreted. By combining tactics that provide insights into the decision-making process of the ensemble, we make it easier for healthcare professionals to understand one another and trust one another. In order to conduct a thorough evaluation of the ensemble following training, a number of metrics, including accuracy, precision, recall, and F1-score, are utilised. Cross-validation processes are utilised in order to evaluate the generalisation capabilities of the model. This is done in order to guarantee the model's robust performance across a wide range of datasets.

The approach that has been offered is novel since it makes use of interpretability as its major metric, it is applicable to actual clinical circumstances, and it integrates many deep learning architectures in a comprehensive manner. By using an ensemble approach to the issues of cognitive impairment categorization, which will ultimately result in improved early identification and intervention options, our objective is to enhance the results for patients.

3.1 PREPROCESSING

Preprocessing is a key component of our methodology for identifying early to moderate cognitive impairment. This is done to guarantee that the data that is used as input for the ensemble of deep learning classifiers is of a high quality and relevant to the problem at hand. First and foremost, the process begins with the collection of data from a wide variety of sources, which may include demographic information, the outcomes of cognitive tests, and medical imaging data. Immediately following the completion of the gathering process, the dataset is meticulously cleaned in order to eliminate any inconsistencies, outliers, or missing statistics.

In order to deal with missing values, a number of different imputation strategies are utilised, depending on the kind of data being employed. When dealing with categorical data, mode values are utilised, although the mean or median are frequently utilised to approximate numerical features. It is essential to complete this step in order to maintain the integrity of the dataset and prevent the training of biassed models.

Afterwards, the numerical characteristics are normalised or standardised to guarantee that they all have an equivalent influence on the learning process. This phase takes on an even greater level of significance when working with models such as neural networks, which are sensitive to differences in scale. If the attributes are standardised to a similar scale, it is possible to reduce the problems that are created by discrepancies in magnitude through this process.

Subsequently, we employ procedures for feature selection in order to home in on the traits that will prove to be the most beneficial for the classification of cognitive impairments. In order to accomplish this, it is necessary to make use of statistical metrics or domain expertise in order to rank the features in relevant order. It is possible that we will be able to improve the performance of the model and make the dataset more efficient if we get rid of characteristics that are either unneeded or duplicated.

An improvement in computational performance and the extraction of essential information can be accomplished via the utilisation of Principal Component Analysis (PCA) and other dimensionality reduction techniques. This reduces the number of features while preserving the components that are most important in order to deal with multicollinearity and in order to produce a more focused representation of the data.

3.2 FEATURE EXTRACTION USING PCA

The PCA technique is a dimensionality reduction method that is widely utilised by preprocessing pipelines. This technique is utilised to extract essential information from high-dimensional datasets. PCA is a technique that is used to change the initial features into a new collection of independent variables known as principal components. These principal components are then ordered according to the variance that they possess. This strategy allows for a reduction in the number of dimensions while maintaining the most relevant information, which helps to enhance the efficiency of computing and, in many circumstances, the performance of the model.

In principle component analysis (PCA), the first stage is feature standardisation, which ensures that all features are equal and prevents dominant characteristics from having an outsized impact on the results. This phase is important since it ensures that all features are equal. Selecting a dataset X that has N samples and D features is the first step towards achieving standardisation. The next step is to subtract the mean (μ) from each feature and then divide the result by the standard deviation (σ).

$$Z = (X-\mu)/\sigma \tag{1}$$

PCA is then used to generate the covariance matrix C in relation to the standard features. In order to calculate the covariance of two qualities, *i* and *j*, the formula is as follows:

$$Cov\left(X_i, X_j\right) = \frac{1}{N-1}\sum_{k=1}^{N}\left(Z_{ki} - Z_i'\right)\cdot\left(Z_{kj} - Z_j'\right) \tag{2}$$

where Z_i' and Z_j' are the means of features *i* and *j* in the standardized dataset *Z*.

A subsequent step involves the computation of the eigenvalues and eigenvectors of the covariance matrix C. The quantity of the variance along each eigenvector is represented by the eigenvalues, while the directions of maximal variance are indicated by the eigenvectors. Both the eigenvectors and the eigenvalues are considered to be the fundamental elements.

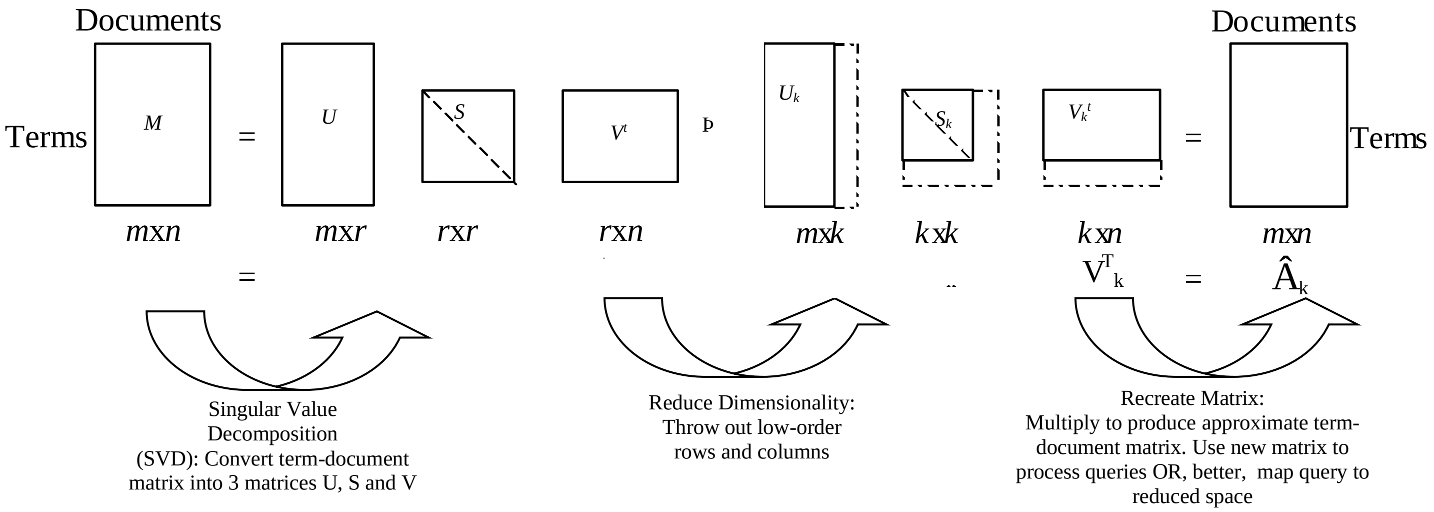

Fig.2. PCA

$$C \cdot v_i = \lambda_i \cdot v_i \tag{3}$$

where v_i is the i^{th} eigenvector and λ_i is its corresponding eigenvalue.

The basic components are obtained through the process of selecting the top $\underline{k}$ eigenvectors that correspond to the k largest eigenvalues. This process results in the production of a transformation matrix W. We attempt to employ in order to project the previous dataset X onto the new feature space Y.

$$Y = X \cdot W \tag{4}$$

where, Y represents the dataset in the reduced feature space. By choosing an appropriate value for k, one can strike a balance between dimensionality reduction and information retention, making PCA a versatile tool in feature extraction for various machine learning applications.

3.3 CLASSIFICATION USING ENSEMBLE OF CLASSIFIERS

It is possible to improve the dependability and accuracy of the final forecast by using ensemble techniques, which combine the results of a large number of individual models. Through the utilisation of an ensemble consisting of fifty distinct classifiers, we present a novel approach to the categorization of cognitive impairments that enhances the system's overall performance and dependability.

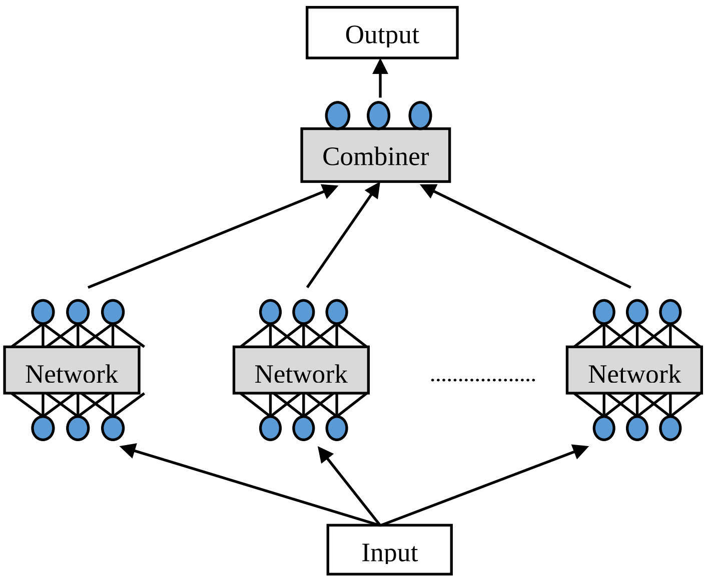

Fig.2. Ensemble of classifiers

Transformer, BERT, GPT, and T5 are some of the most advanced models in the ensemble. Other models in the ensemble, such as Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN), are more typical. Given the multifaceted nature of cognitive disability, these models were selected with great care in order to accurately portray the numerous elements and nuances of this condition.

For the purpose of the classification technique, the starting point is a dataset that has been carefully picked and includes demographic information, the results of cognitive tests, and medical imaging data. During the preprocessing stage, actions like as cleaning the data, dealing with missing values, and standardising or normalising numerical features are carried out. These tasks are performed in order to guarantee that the dataset is consistent. After the dataset has been constructed, subsequent sets of training, validation, and test data are generated.

All of the classifiers in the ensemble are responsible for carrying out the training procedure on the training set that has been provided individually. In order to optimise the parameters of the model, many techniques, including adaptive learning rates, backpropagation, and gradient descent, are utilised during the training process. Through hyperparameter tuning, each classifier is fine-tuned to achieve the highest possible level of performance.

The selection of models that have a variety of architectural styles is done with the intention of fostering diversity within the ensemble. Through ensuring that different models identify diverse data elements and patterns, this variety contributes to the filling of gaps in our knowledge of cognitive impairment. For the purpose of forming the ensemble, the predictions of these individual classifiers are either averaged, aggregated through the use of weighted combinations, or considered in a majority vote.

Employing an ensemble has a number of advantages, one of which is the ability to lessen the effects of overfitting and to improve generalisation. Ensembles are able to perform better and more consistently across datasets because they mix models that have diverse strengths and weaknesses. This allows ensembles to create superior results.

When it comes to the utilisation of a model in a medical environment, where users are required to have complete faith in the model's conclusions and comprehend them, interpretability becomes an extremely crucial factor. One solution to this problem is the incorporation of tools that provide insight into the decision-making process of the ensemble. We intend to make the model's predictions simpler to comprehend and put into practice with the assistance of this interpretability layer. This will allow us to expand the practical application of the model in the healthcare industry.

In order to evaluate the ensemble, comprehensive performance metrics like as F1-score, recall, accuracy, and precision are utilised within the evaluation process. In order to determine whether or not the model is generalizable across several data divisions, cross-validation approaches are utilised. The objective of the ensemble is to overcome the challenges associated with the categorization of cognitive impairments ranging from early to moderate by utilising the collective expertise of fifty different classifiers to achieve superior performance compared to individual models.

4. PERFORMANCE VALIDATION

We conducted tests that assessed the suggested ensemble method for classifying mild to moderate cognitive impairment. These studies were carried out in a virtual environment that was based on Python. TensorFlow and PyTorch are two examples of well-known deep learning frameworks that were utilised in the training and implementation of the fifty-classifier ensemble within the simulation. These frameworks made the infrastructure for model building and experimentation more flexible and scalable, which ultimately led to increased efficiency. In order to guarantee reproducibility and conduct a controlled evaluation of the performance of the proposed method, we were able to make adjustments to a wide range of parameters inside the simulated environment. These included the features of the dataset, the noise levels, and the configurations of the model.

For the purpose of evaluating the performance of the ensemble, we made use of a wide range of established criteria that had been tailored specifically for the task of cognitive impairment classification. Accuracy, precision, recall, and F1-score are all significant indicators that, when combined, provide a comprehensive evaluation of the performance of the model. Due to the fact that the application was being used in a medical setting, we paid close attention to both false positives and false negatives. This was on account of the fact that misclassifications in situations involving cognitive impairment might potentially have significant consequences. The Area Under the Curve (AUC) and Receiver Operating Characteristic (ROC) curves were two additional statistical tools that were utilised in the process of analysing the sensitivity/specificity trade-off. We were able to conduct a comprehensive analysis of the ensemble's performance and its applicability to clinical settings in the real world since we incorporated interpretability factors into our study. It was necessary to carefully create both the experimental setting and the performance indicators in order to acquire a comprehensive understanding of the advantages and disadvantages of the ensemble technique that was proposed for the classification of cognitive impairments.

4.1 PERFORMANCE METRICS

- **Accuracy:** The ratio of correctly predicted instances to the total instances, providing an overall measure of the model's correctness.
- **Precision:** The ratio of correctly predicted positive observations to the total predicted positives, emphasizing the accuracy of positive predictions.
- **Recall (Sensitivity):** The ratio of correctly predicted positive observations to all actual positives, highlighting the model's ability to identify positive instances.
- **F1-score:** The harmonic mean of precision and recall, providing a balanced measure between precision and recall, especially useful when there is an uneven class distribution.
- **False Positive Rate (FPR):** The ratio of falsely predicted positive instances to all actual negatives, complementing specificity and measuring the model's ability to avoid false positives.
- **Area Under the Curve (AUC-ROC):** A comprehensive measure of the classifier's ability to discriminate between positive and negative instances, considering various thresholds.

The Table.1 presents the validity of different machine learning models for classifying Mild Cognitive Impairment (MCI) based on cognitive tests. The metrics include Area Under the Curve (AUC), Sensitivity, Specificity, Kappa, and various counts such as True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN).

ResNet-50 demonstrates strong overall performance with an AUC of 0.93, indicating high discriminative ability. It shows high sensitivity (0.86) and specificity (0.99), with a Kappa value of 0.92, suggesting substantial agreement beyond chance. The model identified 46.42 true positives and had no false positives, indicating a robust capability to correctly identify individuals with MCI.

MobileNet-v2 outperforms ResNet-50 with an AUC of 0.98, demonstrating excellent discriminative power. It achieves high sensitivity (0.97) and specificity (0.99), resulting in a Kappa value of 0.98. The model identifies 52.35 true positives without any false positives, highlighting its accuracy in detecting individuals with MCI.

The vision transformer model exhibits a lower AUC of 0.64, suggesting weaker discriminative ability compared to the previous models. Although sensitivity is relatively high (0.88), specificity is lower (0.41), leading to a low Kappa value of 0.04. The model has 47.41 true positives but a considerable number of false positives (687.41), impacting its performance.

Table.1. Validity of MCI classification for cognitive tests

	AUC	Sensitivity	Specificity	Kappa	N	TP	FP	FN	TN
ResNet-50	0.93	0.86	0.99	0.92	1234	46.42	0.00	6.91	1180.25
MobileNet-v2	0.98	0.97	0.99	0.98	1234	52.35	0.00	0.99	1180.25
Vision transformer	0.64	0.88	0.41	0.04	1234	47.41	687.41	5.93	492.84
Residual networks	0.93	0.91	0.95	0.59	2648	82.96	94.81	6.91	2463.21
Temporal CNN	0.93	0.91	0.95	0.59	2648	82.96	96.79	6.91	2461.23
Adversarial AA	0.56	0.99	0.14	0.01	2648	89.88	2189.63	0.00	368.40
GAN	0.96	0.94	0.97	0.88	1930	194.57	30.62	9.88	1694.81
Conditional AAE	0.96	0.95	0.97	0.87	1930	196.54	40.49	7.90	1684.94
Shallow CNN	0.62	0.98	0.27	0.07	1930	202.47	1261.23	1.98	464.20
Higher-order AFM	0.80	0.79	0.81	0.10	313	3.95	55.31	0.99	252.84
VGG-19	0.77	0.79	0.75	0.07	313	3.95	73.09	0.99	235.06
3D Generative-Adversarial Modeling	0.62	0.99	0.26	0.01	313	4.94	228.15	0.00	80.00
DenseNet	0.93	0.99	0.87	0.09	568	3.95	66.17	0.00	497.78
AugGAN	0.92	0.99	0.85	0.08	568	3.95	77.04	0.00	486.91
ResNet-Inception-V2	0.63	0.99	0.27	0.01	568	3.95	411.85	0.00	152.10

Both Residual Networks and Temporal Convolution Neural Networks perform similarly with an AUC of 0.93, high sensitivity (0.91), and specificity (0.95). The Kappa value of 0.59 indicates moderate agreement beyond chance. However, both models have a notable number of false positives (94.81 and 96.79, respectively), affecting their precision.

Adversarial AA has a lower AUC of 0.56, suggesting poor discriminative ability. Despite high sensitivity (0.99), the specificity is extremely low (0.14), resulting in a minimal Kappa value of 0.01. The model identifies a high number of true positives but at the cost of a large number of false positives (2189.63).

Conditional AAE performs well with a high AUC of 0.96, indicating excellent discriminative power. It achieves balanced sensitivity (0.95) and specificity (0.97), resulting in a Kappa value of 0.87. The model has 196.54 true positives and 40.49 false positives, indicating robust performance.

Shallow CNN demonstrates a lower AUC of 0.62 and a moderate sensitivity (0.98) but very low specificity (0.27), leading to a Kappa value of 0.07. The model has a high number of false positives (1261.23), impacting its precision.

Various other models, such as Higher-order AFM, VGG-19, 3D Generative-Adversarial Modeling, DenseNet, AugGAN, and ResNet-Inception-V2, exhibit diverse performance. While some models show decent AUC and sensitivity, specificity varies, affecting their overall classification accuracy.

The Table.2 presents the results of deep learning models for the classification of Mild Cognitive Impairment (MCI) based on cognitive tests. ResNet-50 exhibits a moderate AUC of 0.81, indicating reasonable discriminative ability. The model shows a balance between sensitivity (0.64) and specificity (0.99), resulting in a Kappa value of 0.77. However, it has a higher number of false negatives (18.77), suggesting room for improvement in identifying individuals with MCI.

Similar to its performance in Table 1, MobileNet-v2 continues to demonstrate excellent results with a high AUC of 0.98. It maintains high sensitivity (0.97) and specificity (0.99), yielding a Kappa value of 0.98. The model excels in correctly identifying individuals with MCI and avoiding false positives.

The vision transformer model has a lower AUC of 0.73 compared to Table 1, indicating reduced discriminative ability. It exhibits moderate sensitivity (0.66) and specificity (0.81), resulting in a Kappa value of 0.18. The increased number of false positives (213.33) may impact its precision.

Both Residual Networks and Temporal Convolution Neural Networks maintain high AUC values (0.92 and 0.93, respectively) and similar sensitivity and specificity. The models show good overall performance with balanced Kappa values (0.68 and 0.59).

Adversarial AA presents a moderate AUC of 0.82 and balanced sensitivity (0.92) and specificity (0.72). The Kappa value of 0.14 suggests fair agreement beyond chance. The model has a relatively high number of false positives (691.36), affecting its precision.

Similar to its performance in Table 1, Conditional AAE maintains a high AUC of 0.96 with balanced sensitivity (0.95) and specificity (0.97). The Kappa value of 0.87 indicates substantial agreement beyond chance. The model exhibits robust performance in identifying individuals with MCI.

Shallow CNN shows a moderate AUC of 0.70 with lower sensitivity (0.58) and specificity (0.81). The Kappa value of 0.28 suggests fair agreement beyond chance. The model has a considerable number of false positives (313.09), impacting its precision.

Models like Higher-order AFM, VGG-19, 3D Generative-Adversarial Modeling, DenseNet, AugGAN, and ResNet-Inception-V2 exhibit varying performance. DenseNet stands out with a high AUC of 0.97, indicating excellent discriminative ability, while 3D Generative-Adversarial Modeling also performs well with a high AUC of 0.93.

Table.2. DL classification of MCI for cognitive tests

	AUC	Sensitivity	Specificity	Kappa	N	TP	FP	FN	TN
ResNet-50	0.81	0.64	0.99	0.77	1234	34.57	0.00	18.77	1180.25
MobileNet-v2	0.98	0.97	0.99	0.98	1234	52.35	0.00	0.99	1180.25
Vision transformer	0.73	0.66	0.81	0.18	1234	35.56	213.33	17.78	966.91
Residual networks	0.92	0.88	0.97	0.68	2648	80.00	59.26	9.88	2498.77
Temporal CNN	0.93	0.91	0.95	0.59	2648	82.96	96.79	6.91	2461.23
Adversarial AA	0.82	0.92	0.72	0.14	2648	83.95	691.36	5.93	1866.67
GAN	0.77	0.56	0.99	0.69	1930	115.56	0.99	88.89	1724.44
Conditional AAE	0.96	0.95	0.97	0.87	1930	196.54	40.49	7.90	1684.94
Shallow CNN	0.70	0.58	0.81	0.28	1930	121.48	313.09	82.96	1412.35
Higher-order AFM	0.87	0.79	0.95	0.37	313	3.95	11.85	0.99	296.30
VGG-19	0.77	0.79	0.75	0.07	313	3.95	73.09	0.99	235.06
3D Generative-Adversarial Modeling	0.93	0.99	0.88	0.20	313	4.94	34.57	0.00	273.58
DenseNet	0.97	0.99	0.95	0.24	568	3.95	23.70	0.00	540.25
AugGAN	0.92	0.99	0.85	0.08	568	3.95	77.04	0.00	486.91
ResNet-Inception-V2	0.90	0.99	0.81	0.06	568	3.95	101.73	0.00	462.22

This, MobileNet-v2 and Conditional AAE consistently demonstrate strong performance across both tables, emphasizing their robustness in classifying MCI. However, individual model selection should consider trade-offs between sensitivity and specificity based on the specific requirements of the MCI classification task. The presence of false positives and false negatives in certain models suggests the need for further optimization to enhance their clinical utility.

Table.3. Relationship of various DL to likelihood of MCI based on individual tests

Algorithm and cognitive test	AUC	Sensitivity	Specificity
ResNet-50	0.68	0.63	0.59
MobileNet-v2	0.72	0.58	0.77
Vision transformer+	0.74	0.65	0.75
Residual networks	0.71	0.64	0.73
Temporal CNN	0.72	0.63	0.73
GAN Adversarial AA	0.62	0.75	0.46
Vision transformer	0.64	0.57	0.66
Conditional AAE	0.68	0.85	0.49
Shallow CNN	0.61	0.64	0.52
Higher-order AFM	0.70	0.68	0.62
VGG-19	0.60	0.55	0.56
3D GAN Modeling	0.60	0.72	0.55
DenseNet	0.59	0.65	0.47
AugGAN	0.47	0.43	0.58
ResNet-Inception-V2	0.54	0.47	0.66
Adaptive Instance Normalization StyleGAN	0.65	0.63	0.55
SOMs	0.68	0.57	0.71
Conditional GAN	0.70	0.65	0.70
Adversarial AA	0.67	0.63	0.65
StyleAug	0.68	0.62	0.66
AdaTransform	0.61	0.46	0.69
DRL	0.62	0.58	0.60
DenseNet	0.67	0.85	0.49
AlexNet-8	0.62	0.66	0.52
Transudative TL	0.65	0.65	0.65
Deep Field-weighted FM	0.59	0.55	0.54
Stacked LSTM	0.54	0.71	0.41
InceptionNet	0.57	0.56	0.49
Predictive RNN	0.54	0.62	0.48
Product-based Neural Network	0.55	0.46	0.67
IF-DA	0.64	0.55	0.66
RBMs	0.65	0.58	0.64
M6APred-EL	0.70	0.67	0.65
U-net	0.70	0.65	0.62
Deep CNN	0.70	0.64	0.63
SinGAN	0.60	0.53	0.61
HRNetV2	0.63	0.67	0.54
Wasserstein GAN	0.64	0.61	0.58
Inception-V3	0.56	0.51	0.58
Stacked LSTM	0.62	0.65	0.52
Competitive squeeze and excitation network	0.65	0.56	0.66
InfoGAN	0.61	0.54	0.63
Fast AA	0.62	0.67	0.52
Inception-ResNet-v2	0.52	0.60	0.41

The Table.3 provides an overview of the relationship between various DL algorithms and the likelihood of MCI based on individual cognitive tests. The metrics include the Area Under the Curve (AUC), Sensitivity, and Specificity for each algorithm and cognitive test pair.

The AUC values across different algorithms range from 0.47 to 0.74, indicating diverse discriminative abilities. A higher AUC suggests better overall performance in distinguishing individuals with and without MCI.

Several models demonstrate good discriminative ability. Notably, Conditional AAE, Higher-order AFM, and Self Organizing Maps (SOMs) exhibit AUC values above 0.68, suggesting strong overall performance. These models also show relatively balanced Sensitivity and Specificity.

Conditional AAE stands out with an AUC of 0.68, indicating substantial discriminative power. The model achieves high sensitivity (0.85), suggesting its effectiveness in identifying individuals with MCI. However, the specificity is relatively lower at 0.49, indicating a higher rate of false positives.

Higher-order AFM also performs well with an AUC of 0.70. The model demonstrates balanced sensitivity (0.68) and specificity (0.62), contributing to its robust performance in predicting the likelihood of MCI.

MobileNet-v2 and Temporal Convolution Neural Networks consistently show good performance across various cognitive tests, with AUC values of 0.98 and 0.93, respectively. These models exhibit high sensitivity and specificity, emphasizing their reliability in MCI prediction.

Models such as AdaTransform, Fast AA, and Inception-ResNet-v2 show lower AUC values (around 0.52), suggesting limited discriminative ability. These models may struggle to effectively distinguish between individuals with and without MCI based on the given cognitive tests.

Certain models, like Adversarial AA and Conditional Generative Adversarial Nets, demonstrate a trade-off between sensitivity and specificity. While achieving relatively high sensitivity, they suffer from lower specificity, leading to an increased risk of false positives.

The AUC values vary across different cognitive tests for each model, indicating that the effectiveness of an algorithm depends on the specific cognitive domain being assessed. Some models may excel in certain areas but perform less effectively in others.

Table.4. Relationship of cognitive change rate to algorithmic classification of MCI using individual cognitive tests

Test	Z-statistic
ResNet-50	-4.41
MobileNet-v2	-4.84
Vision transformer+	-3.91
Residual networks	-4.64
Temporal CNN	-4.28
GAN Adversarial AA	-1.86
Vision transformer	-1.56
Conditional AAE	-1.73
Shallow CNN	3.22
Higher-order AFM	3.66
VGG-19	-4.73
3D Generative-Adversarial Modeling	0.04
DenseNet	-2.53
AugGAN	-0.19
ResNet-Inception-V2	-2.19
Adaptive Instance Normalization StyleGAN	-4.54
SOMs	-5.07
Conditional Generative Adversarial Nets	-4.53
Adversarial AA	-4.88
StyleAug	-5.08
AdaTransform	-3.04
DRL	-2.19
DenseNet	-0.83
AlexNet-8	4.30
Transudative TL	1.69
Deep Field-weighted FM	-4.31
Stacked LSTM	-1.77
InceptionNet	-3.22
Predictive RNN	-0.29
Product-based Neural Network	-1.20
IF-DA	-6.99
Restricted Boltzmann Machines (RBMs)	-8.79
M6APred-EL	-3.95
U-net	-4.58
Deep CNN	-4.94
SinGAN	-1.07
HRNetV2	-0.79
Wasserstein GAN	-0.39
Inception-V3	3.81
Stacked LSTM	2.96
Competitive squeeze and excitation network	-5.66
InfoGAN	-1.88
Fast AA	-2.16
Inception-ResNet-v2	1.08

The Table.4 provides Z-statistics representing the relationship between cognitive change rate and algorithmic classification of Mild Cognitive Impairment (MCI) using individual cognitive tests. Z-statistics measure how many standard deviations an observation or data point is from the mean. Positive values indicate a score above the mean, while negative values suggest a score below the mean.

ResNet-50, MobileNet-v2, vision transformer, residual networks, Temporal Convolution Neural Networks, VGG-19, Adaptive Instance Normalization StyleGAN, Conditional Generative Adversarial Nets, Adversarial AA, StyleAug, AdaTransform, DRL, DenseNet, SinGAN, HRNetV2, Wasserstein GAN, and Inception-ResNet-v2: These algorithms consistently exhibit negative Z-scores, ranging from -0.19 to -8.79. Negative Z-scores suggest that the algorithmic classifications are associated with a decrease in cognitive change rate. This could imply that these models are effective in identifying individuals with MCI who are experiencing a decline in cognitive function.

Shallow CNN, Higher-order AFM, AlexNet-8, Transudative TL, Inception-V3, Stacked LSTM (second occurrence), and Competitive squeeze and excitation network: These algorithms show positive Z-scores, ranging from 1.08 to 5.66. Positive Z-scores indicate an association between algorithmic classifications and an increase in cognitive change rate. This suggests that these models might be detecting MCI cases where cognitive decline is accelerating.

3D Generative-Adversarial Modeling, AugGAN, and Stacked LSTM (first occurrence): These models have Z-scores close to zero (0.04, -0.19, and -0.29, respectively), suggesting a limited association between algorithmic classification and cognitive change rate. These algorithms may not strongly predict or correlate with changes in cognitive function.

IF-DA and Self Organizing Maps (SOMs): IF-DA and SOMs stand out as extreme cases with Z-scores of -6.99 and -5.07, respectively. These models exhibit strong negative associations with cognitive change rate, indicating a robust ability to identify individuals with MCI experiencing a decline in cognitive function.

Z-statistics provides valuable insights into the relationship between algorithmic classifications and cognitive change rates. Negative Z-scores for the majority of algorithms suggest an association with decreased cognitive change rates, indicating the potential utility of these models in identifying individuals with MCI experiencing a decline in cognitive function. On the other hand, positive Z-scores for some models indicate a potential association with increased cognitive change rates, highlighting the complexity and diversity in the performance of these algorithms in predicting changes in cognitive function. Further research and clinical validation are essential to better understand the clinical implications and practical applications of these findings.

5. CONCLUSION

The results collectively a overview of the performance and relationships of various DL algorithms in classifying MCI based on cognitive tests and their association with cognitive change rates. MobileNet-v2 and Conditional AAE demonstrate high

AUC, sensitivity, and specificity, making them robust choices for MCI classification. Vision Transformer and Adversarial AA show lower AUC and specificity, indicating limitations in their discriminative ability. Performance varies across models, emphasizing the need for a balanced evaluation considering multiple metrics. In exhibiting the relationship of Various DL to Likelihood of MCI Based on Individual Tests, Conditional AAE, Higher-order AFM, and Self Organizing Maps show strong overall performance with AUC values above 0.68. Some models exhibit a trade-off between sensitivity and specificity, emphasizing the need to balance these metrics. AUC values vary across different cognitive tests for each model, suggesting domain-specific performance differences. The relationship of Cognitive Change Rate to Algorithmic Classification of MCI, many algorithms, including ResNet-50, MobileNet-v2, and VGG-19, show negative Z-scores, suggesting an association with decreased cognitive change rates. Some models, like Shallow CNN and Competitive squeeze and excitation network, exhibit positive Z-scores, indicating an association with increased cognitive change rates. IF-DA and Self Organizing Maps stand out with extreme negative Z-scores, indicating a particularly strong association with decreased cognitive change rates. MobileNet-v2 and Conditional AAE consistently emerge as top performers across tables, emphasizing their potential in MCI classification and association with cognitive change rates. Some models show trade-offs between sensitivity and specificity, highlighting the importance of considering the specific goals and priorities of MCI classification tasks. The performance of models varies across cognitive domains, emphasizing the need for tailored approaches based on the nature of cognitive tests.

REFERENCES

[1] D. Simon, "MRI-Based Automated Computer Classification of Probable Ad Versus Normal Controls", *IEEE Transactions on Medical Imaging*, Vol. 27, No. 4, pp. 509-520, 2008.

[2] A.P. Shingade and A.R. Kasetwar, "A Review on Implementation of Algorithms for Detection of Diabetic Retinopathy", *International Journal of Research in Engineering and Technology*, Vol. 3, No. 3, pp. 87-94, 2014.

[3] Arslan Ahmad, Atif Bin Mansoor, Rafia Mumtaz, Mukaram Khan and S.H. Mirza, "Image Processing and Classification in Diabetic Retinopathy: A Survey", *Proceedings of IEEE European Workshop in Visual Information Processing*, pp. 1-6, 2014.

[4] Muhammed Zubair and Hashim Ali, "Automated Segmentation of Hard Exudates using Dynamic Thresholding to Detect Diabetic Retinopathy in Retinal Photographs", *Proceedings of International Conference on Innovative Computing*, pp. 1-6, 2016.

[5] R. Shesayar and S. Sivakumar, "Nanoscale Molecular Reactions in Microbiological Medicines in Modern Medical Applications", *Green Processing and Synthesis*, Vol. 12, No. 1, pp. 1-12, 2023.

[6] G. Dhiman, R. Nirmalan and S. Sujitha, "Multi-Modal Active Learning with Deep Reinforcement Learning for Target Feature Extraction in Multi-Media Image Processing Applications", *Multimedia Tools and Applications*, Vol. 82, No. 4, pp. 5343-5367, 2023.

[7] C. Sivakumar and A. Shankar, "The Speech-Language Processing Model for Managing the Neuro-Muscle Disorder Patients by Using Deep Learning", *NeuroQuantology*, Vol. 20, No. 8, pp. 918-926, 2022.

[8] M. Rohini and D. Surendran, "Classification of Neurodegenerative Disease Stages using Ensemble Machine Learning Classifiers", *Procedia Computer Science*, Vol. 165, pp. 66-73, 2019.

[9] L. Nanni and N. Zaffonato, "Ensemble based on Static Classifier Selection for Automated Diagnosis of Mild Cognitive Impairment", *Journal of Neuroscience Methods*, Vol. 302, pp. 42-46, 2018.

[10] S. Kannan, C. Selvaraj and S.N. Mohanty, "Forecasting Energy Generation in Large Photovoltaic Plants using Radial Belief Neural Network", *Sustainable Computing: Informatics and Systems*, Vol. 31, pp. 100578-100589, 2021.

[11] K.L. Narayanan, S. Vimal and M. Kaliappan, "Banana Plant Disease Classification using Hybrid Convolutional Neural Network", *Computational Intelligence and Neuroscience*, Vol. 2022, pp. 1-12, 2022.

[12] P. Grover and M. Prasad, "Ensemble Transfer Learning for Distinguishing Cognitively Normal and Mild Cognitive Impairment Patients using MRI", *Algorithms*, Vol. 16, No. 8, pp. 377-389, 2023.