Received: 22 June 2022 | Revised: 1 February 2023 | Accepted: 26 June 2023

DOI: 10.1111/coin.12594

ORIGINAL ARTICLE

Computational Intelligence WILEY

An intelligent hierarchical residual attention learning-based conjoined twin neural network for Alzheimer's stage detection and prediction

Venkatesh Gauri Shankar[1,2] | **Dilip Singh Sisodia**[1] | **Preeti Chandrakar**[1]

[1]Department of Computer Science and Engineering, National Institute of Technology Raipur, Raipur, India

[2]Department of Information Technology, Manipal University Jaipur, Jaipur, India

Correspondence
Venkatesh Gauri Shankar, Department of Computer Science and Engineering, National Institute of Technology Raipur, Raipur, India.
Email: vgshankar.phd2018.cse@nitrr.ac.in

Abstract

Alzheimer's disorder (AD) causes permanent impairment in the brain's memory of the cellular system, leading to the initiation of dementia. Earlier detection of Alzheimer's disease in the initial stages is challenging for researchers. Deep learning and machine learning-based techniques can help resolve many issues associated with brain imaging exploration. Brain MR Images (Brain-MRI) are used to detect Alzheimer's in computable research work. To correctly categorize the stages of Alzheimer's disease, discriminative features need to be extracted from the MR images. Recently, many studies have used deep learning methods for the early detection of this disorder. However, overfitting degrades the deep learning method's performance because the dataset's selection images are smaller and imbalanced. Some studies could not reach more discriminative and effectual attention-aware features for Alzheimer's stage classification to increase the model performance. In this paper, we develop a novel hierarchical residual attention learning-inspired multistage conjoined twin network (HRAL-CTNN) to classify the stages of Alzheimer's. We used augmentation approaches to scale insufficient and imbalanced data. The HRAL-CTNN is efficiently overcoming the issues of

not obtaining efficient attention-aware and generative features for Alzheimer's stage classification. The proposed model solved the problem of redundant features by extracting attentive discriminant features, and scaling imbalance data by data augmentation, after that training and validation using HRAL-CTNN. The execution of this proposed work has been performed on the ADNI MRI dataset. This work achieved outstanding accuracy of $99.97 \pm 0.01\%$ and F1 score of $99.30 \pm 0.02\%$ for Alzheimer's stage classification. This model proposed by our group outperformed the existing related studies in terms of the model's performance score.

KEYWORDS

Alzheimer's disorder, attention learning, conjoined neural network, convolutional neural network, deep learning

1 | INTRODUCTION

Alzheimer's disease (AD) damages brain tissue and kills neurons, causing memory loss and impairment of daily tasks, including reading, speaking, and writing.[1,2] Numerous studies have shown that neurodegenerative changes associated with Alzheimer's disease begin decades before symptomatic disease.[2,3] Patients with mild cognitive impairment behave aggressively, while end-stage AD leads to death due to heart failure and respiratory dysfunction.[1,2,4] All indicators of Alzheimer's disorder start slowly, but over time, when an individual's brain disorders begin, they are severely affected. Therefore, early diagnosis and treatment of Alzheimer's can improve patient outcomes.[5–7] Many people suffer from this disease every year. It is expected that 1 in 85 people worldwide will suffer from AD by 2050.[8]

Alzheimer's disease reduces the size of the brain's cerebral cortex and hippocampus while expanding the size of the ventricles. As a part of the central nervous system (CNS), perfection in memory and logical abilities are destroyed along with neuronal disorders in the rest of the brain, eventually leading to a person's death.[3,9] Alzheimer's disease has five primary stages: cognitive normal (CN), significant memory concern (SMC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and Alzheimer's disease (AD).[5,7] Researchers have developed various computer-aided diagnostic systems (CADSs) to accurately find and classify the extracted features from MR images and design a detection cum prediction model.[5,9,10]

Alzheimer's disease is still not well diagnosed, and there is no cure. Instead, a few ways to treat the disease slow its spread.[1] Therefore, early detection is an important part of improving the lives of people and their families from the disease. This is done with cognitive, psychological, or clinical tests with computer-aided support. MRI (magnetic resonance imaging) of the brain is some of the most important in clinical trials.[6] This is because these images show how the shape of the brain has changed over time and has a strong link to the brain's structure. MRI images are primarily used to see how the shape of the brain changes over time.[11,12] In MRI images, the areas

where cells are dying because of the disease have very low intensities, which makes them look darker than healthy areas.[13,14]

Most of the ways images are used to help diagnose AD are related to computer vision,[15] but recently, some machine learning[16,17] techniques have been added. For example, Stonnington et al.[18] used regression analysis with likelihood estimation to look for and track the disease, while Li et al.[19] used support vector machines (SVMs) to help with diagnoses. Researchers often use the ML method to find image patterns automatically.[20,21] MRI has been used to diagnose AD in its early stages using classical machine learning algorithms such as the SVM algorithm[22] and linear analysis.[23] To classify the MRI, it was recently suggested to use a feed-forward neural network[21] that used a dual-tree complex wavelet transform to extract the features. They compared their model to other popular methods and reported higher performance. In Reference 24, a study with four Alzheimer's stages was suggested. The study looked at the diagnosis of AD, early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and healthy controls (HCs). Multicore SVM[25] and weighted random SVM[26] have also been employed for the AD classification, and the proposed model performance has increased. An approach[27] of machine learning that is based on random forests has been used to find the many methylation regions that have the capability to be employed as biomarkers for Alzheimer's disease.

Rather than machine learning (ML), researchers are also using deep learning (DL) to describe AD in MRI images by making computational models with multiple processing layers.[28,29] DL is a subdomain of ML, but it automatically learns from the images and extracts the features.[28–30] In a traditional machine learning model, an expert label the data, which can be slightly subjective. DL eliminates the need for that expert.[31–34] DL technologies and deep networks are increasingly used in medical imaging.[35,36]

Convolutional neural networks (CNNs) are DL models that analyze multidimensional data, such as time series and photographs. Convolutional neural networks can analyze these types of data more effectively than traditional ML methods.[37,38] Convolutional neural networks can analyze various data that are more accurately than conventional approaches. They build successive feature maps by extracting simple characteristics of the data (e.g., vertices and edges in images) in the first layers and then grouping those simple qualities into more complex patterns using the information obtained from those simple features (e.g., shapes, area, volume, etc.).[39] The creation of these feature maps involves the application of convolutional procedures that have trainable kernels.[39,40] The feature maps are generated by applying those operations to the layer's input. In convergent expansion in a network, two complementary functions, pooling, and nonlinear transformations are beneficial.[41] After this step is finished, the processed feature maps serve as the foundation for the prediction (typically by employing entirely related layers).[41,42]

Deep learning methods have significantly improved performance over most non-deep learning methods.[36] Researchers have recently developed models/algorithms based on deep learning to get features from MRI.[36,43] Deep learning frameworks or techniques primarily focus on binary classification that shows whether an individual has AD.[3,6] However, for an efficient patient diagnosis, various stages of Alzheimer's must be classified, such as MCI, SMC, LMCI, and EMCI.[5,7] Convolutional neural networks have been found to perform well in deep learning when working with large MR image datasets.[44–46] However, the most vital benefit of CNNs over traditional ML methods is that there is no requirement of manually extract features. It can automatically extract efficient features for classifying Alzheimer's stages.[37,44,45] A study[47] presented the use of biomarkers and a feature extraction-based model with DL techniques for Alzheimer's stage detection. It used AD and MCI stages to analyze and detect Alzheimer's in MR Images with 90% accuracy. It has not covered all stages of Alzheimer's, and due to many redundant features and the imbalanced

nature of the dataset taken, having overfitting too. A machine learning-based study[48] proposed an idea with a feature selection model that is based on reciprocal relationships and has acute high-altitude response-like characteristics for brain regions that are predefined. In the machine learning model, feature extraction is manual, which will take too much time for Alzheimer's detection, and these models are also overfitted. An extensive review[49] of the different studies provided a comparative analysis. He showed that related studies only used the detection or prediction of MCI versus AD in their study. The suggested review by many authors showed an exceptionally low accuracy in the proposed deep learning model due to redundant feature sets and overfitting in the model.

Recently, a classification approach[50] for AD patients using a deep learning system was presented. When trying to anticipate AD and normal output classes, this study used autoencoders in conjunction with CNNs to reach a classification accuracy of 98.4%. They did not consider other stages or categories of Alzheimer's that are also essential to detect the correct stage of Alzheimer's. Convolutional neural networks extract discriminative features and categorize Alzheimer's disease and normal control.[51,52] It is still challenging to extract useful information from large unstructured data, even though deep learning-based techniques have achieved significant accomplishments in analyzing MRI big data.[53] This is because the procedures require a significant level of processing capability and extensive model training. Choosing the optimal architecture for the system and the hyperparameters that give it the best performance might be just as difficult.

Helaly HA et al.[54] developed a robust structure for the early diagnosis of Alzheimer's disease with MRI images for four AD stages. They used deep learning-based CNN with four stages of the AD spectrum. Martinez-Murcia et al.[55] used deep convolutional autoencoders to discover the extensive data assessment of AD. The data-driven deconstruction of MRI images enables us to identify features from MRI scans that indicate the underlying neurodegenerative process and an individual's Alzheimer's cognitive symptoms. After doing a regression and classification analysis, the influence of each coordinate of the autoencoder manifold on the brain is calculated. It investigates the allocation of the extracted features across a broad spectrum of possible permutations. Wen et al.[56] managed a transfer learning-based model with the addition of a CNN. They concluded that the methods based on transfer learning performed significantly better than the methods not based on transfer learning. These observations are only influential when applied to the binary AD classification task.

However, many deep learning-based studies have limitations, such as (1) overcoming the challenges associated with data inadequacy impediments in imbalanced datasets; (2) not covering all the stages of Alzheimer's disease, that is, the multiclass classification of Alzheimer's disease; (3) less efficient real-time early detection of Alzheimer's stage, which requires a more precise and tiresome tuning of several arguments that are a source of problems with overfitting and distress the whole efficiency of the working model; and (4) not obtaining generative and efficient attention-aware discriminative features for efficient model training.

Apart from the above research background, many studies worked on feature extraction and model training for Alzheimer's detection and prediction. Some studies worked on CNN and prebuilt models,[44,54,57,58] convolutional network with autoencoder,[55] CBIR with CNN,[37] transfer learning[56] for detecting and predicting Alzheimer's stages, whereas some of them worked on feature extraction and selection using CNN.[55,59,60] Moreover, the existing studies have many problems such as redundant feature extraction and selection, not covering all stages of Alzheimer's, performance degradation due to imbalanced data, and overfitting in the model.

Our study first put forward a novel multistage conjoined twin network-based CNN that uses the hierarchical residual attention-based model to attain high cumulative efficiency for

Alzheimer's disorder diagnosis. After that, we provided hierarchical residual attention training with the residual-based skip association in the residual to enable the attention-aware module to create a more effectual discriminative MR feature set. Furthermore, a multistage correlation filter was used for aggregating discriminant features and data augmentation to scale the insufficient and imbalanced data. The overall proposed study started with 7000 MR brain images and the classification of five Alzheimer's stages. The vital contributions are enumerated as follows:

1. An attention learning based CTNN for multistage classification of Alzheimer's disorder.
2. We used hierarchical residual attention learning-inspired multistage conjoined twin networks to obtain efficient attention-aware and generative features that are discriminative.
3. To use data augmentation approaches with attention learning to scale the insufficient and imbalanced data.
4. To include multistage-initiated correlation filters for aggregating discriminant features in MR images for Alzheimer's stage classification.
5. To mitigate the overfitting of the model by tuning and regularization parameters.

This paper has been further organized: Section 2 explains the material and methods. Section 3 elaborates on the results and validation of the model. Section 4 discusses and compares the same cohort-related work, and Section 5 addresses the conclusion and future scope.

2 | MATERIALS AND METHODS

In our work, the proposed model covered four phases. The first phase covers data augmentation and preprocessing. Feature extraction from taken MRI images is included in the second phase. The third phase covers the hierarchical residual attention learning and multistage correlation filter. The fourth phase covers the classification of Alzheimer's classes. We proposed a conjoined twin-flavored CNN (CTNN) to categorize Alzheimer's stages. We modified the VGG19[38,39] based CNN model[40–42] by adding one additional convolutional layer in the proposed work to obtain maximum discriminative features from an imbalanced dataset. We have used two modified VGG19 layers with 17 convolutional layers, two fully connected layers, five maxpooling layers, four normalizations, and four Gaussian noise layers for feature detection. We presented the parallel architecture of CNN to train the model and extract features from the MR images for improving model performance.

2.1 | Data collection

In our proposed work, we collected the ADNI dataset[61] from the library of neuroimaging (LONI).[5,61] We collected 7000 MRI images from the same dataset, in which 3300 cognitive normal (CN), 996 significance memory concern (SMC), 1430 early mild cognitive impairment (EMCI), 1010 late mild cognitive impairment (LMCI), and 264 Alzheimer's dementia (AD) MRI images were present. The same MRI image data were collected within the age group of 25 to 90 years. We applied image preprocessing steps to identify the efficient features and data augmentation to keep the data balance in each class to improve the learning rate and accuracy. All the MRI images used were 256×256. Nevertheless, we converted them into 224×224 using dimension

TABLE 1 MRI dataset and its demographic information.[5,7,61–64]

Alzheimer stage	No. of males	No. of females	Age (Average)	MMSE (Average)	MoCA (Average)	ADAS13 (Average)	CDR-SB (Average)
CN	**2162**	**1138**	**67.16**	**28.97**	**28.72**	**9.57**	**1.24**
SMC	659	337	66.40	28.70	27.58	9.29	1.12
EMCI	848	582	64.60	25.31	24.12	8.92	0.76
LMCI	596	414	64.13	25.19	23.01	8.67	0.66
AD	133	131	75.71	18.34	16.23	7.94	0.49
Average/total	7000		67.60	25.30	23.93	8.88	0.85

scaling to process them into a CNN. Table 1 gives the information of the MRI dataset, including demographic information for each stage.

2.2 | Image preprocessing and data augmentation

Our proposed model goes through the image preprocessing[65–67] steps for enhancing and transforming MRI images. This improves low variation and insufficient brightness issues in MRI images. We have applied image enhancement methods such as contrast stretching and Linear contrast enhancement[59] (for low variation and poor brightness problems), image acquisition[65,66] (creates the visual features of MRI images, such as the hippocampus, cerebellum cortex, entorhinal cortex, gray vol, white vol, etc), nonlinear filters[60] (to remove certain types of nonadditive noise in the MRI images), and hierarchical clustering[68] (segmentation of intense features).

Data augmentation techniques,[69,70] such as horizontal flipping, cropping, and padding, are often used to train CNNs. Data augmentation is common for deep learning models to obtain better results with more data.[70] Imbalanced small datasets affect model performance by creating problems that are overtuned during model training. To overcome the same problem, the proposed model used the process of data augmentation in this work. Several methods are supported in data augmentation and the pixel resizing method. Nevertheless, we focused on seven basic data augmentation techniques for the selected MRI data. These included shift_width_range, shift_hight_range, horizontal_flip, vertical_flip, rotation_range, shear_range, and zoom_range. The proposed model HRAL-CTNN has solved the problem of imbalanced data using the above data augmentation techniques. The resultant extended part of MR images are generated after seven data augmentation techniques and also taking their precise threshold values. Table 2 presented the proposed model using all data augmentation techniques with their applied parameters. Algorithm 1 presents all the steps involved in image processing and data augmentation.

2.3 | Conjoined CNN with hierarchical residual attention learning

We have proposed a conjoined CNN for model training and testing in this study. We involved a combined platform with a modified CTNN, in which two divisions work similarly. These divisions are (1) a training model division that extracts the proposed features for the training set and weights by using the attention argument engendered with the attention channel and (2) a test model division that extracts the model features for the test set and weights by using the attention

TABLE 2 Data augmentation techniques and their parameter values.

Data augmentation technique	Parameter value
Shift_width_range	0.15°
Shift_hight_range	0.15°
Horizontal_flip	−200 to 200
Vertical_flip	−300 to 300
Rotation_range	5°
Shear_range_factor	0.2°
Zoom_range_factor	1.0, 1.5

Algorithm 1. Image preprocessing with data augmentation

procedure ImagePreprocess **(Input: MRI; Alzheimer's stages: S)**
  Input: MRI Images: $i \in [1, \mathrm{I}]$; Stages: S ← {CN, SMC, EMCI, LMCI, AD}
  Output: P (MRI) ← Preprocessed MR images
 for all MRI Images: $i \in [1, \mathrm{I}]$ **do**
  for all Stages S **do**
  Contrast stretching: $\mathrm{CO_S}\,(i) \leftarrow i \in [1, \mathrm{I}]$.
  $\mathrm{CO_S}\,(i_1, i_2, i_3, \ldots i_n) \leftarrow i \in [1, \mathrm{I}]$
  Image acquisition: $f_s \leftarrow \mathrm{CO_S}\,(i_1, i_2, i_3, \ldots i_n)$
  Nonlinear filter: $N_0 \leftarrow f_s\,(\alpha R + \beta S)$
  Hierarchical Segmentation: $K_C \leftarrow N_0\,(\mathrm{Gx}^2 + \mathrm{Gy}^2)$
  end for
 data augmentation {shift, flip, rotation zoom, shear}
 end for
end procedure

argument engendered with the attention channel.[71,72] The final output of both divisions has been provided for the multistage correlation filter. The primary purpose of using hierarchical initiated residual attention-aware learning was to create more effectual, propagative, and discriminative extracted features. In our model, CNN layers extract local features from Alzheimer's MR images. The proposed CNN model holds individual nodes with ascertainable bias and weights. The weights are in the form of matrix, also known as convolutional kernel. The dimensionality reduction in our proposed model is done by the pooling layer. For the twin CNN as a modified CTNN, we have used the same convolutional layer formula in Equations 1–3. We used the ReLU (rectified linear unit) activation function sequentially layer-by-layer operation. VGG19 stands for the layers of convolutional networks that take some weights. Table 1 gives all the layered and activation function information.

$$H = \frac{\mathrm{IH} - \mathrm{KH} + 2(\mathrm{PD})}{\mathrm{St}} + 1, \tag{1}$$

where H is the layer's height, IH is the image height, KH is the kernel height, PD is padding, and St is strides, which means pixels shift over the input matrix.

$$W = \frac{\text{IW} - \text{KW} + 2(\text{PD})}{\text{St}} + 1, \tag{2}$$

where W is the layer width, IW is the image width, KW is the kernel width, PD is the padding, and St is the stride.

In general,

$$C = \frac{\text{IHW} - \text{F} + 2\ (\text{PD})}{\text{St}} + 1, \tag{3}$$

where C is the convolutional layer, IHW is the image height and width, PD is padding, St is strides, and F is a fitter.

Residual attention learning is a neural network-based attention mechanism.[71,72] It is compatible with the most cutting-edge feed-forward network design and is used for end-to-end model training and discriminating features. We have used hierarchical residual attention learning-inspired multistage conjoined twin network for obtaining efficient attention-aware and generative features that are discriminative. Hierarchical residual attention learning is the subsequent formation of an attention module, written in Equation 4.

$$Y(x) = T(x) * F(x), \tag{4}$$

where $F(x)$ are the features of an MRI image x from the convolutional layer, $T(x)$ is the attention module map generated after the hourglass unit (a feature position estimation),[73] and $Y(x)$ is the attention module-aware developed features after map generation. The asterisk (*) symbol is the matrix product of $T(x)$ and $F(x)$.

The attention module initiated residual network in the form of residual skip connection for upgrading the performance of modified CTNN has been used in this study. We have proposed the same hierarchical initiated residual attention learning method for obtaining effectual, propagative, and discriminant extracted features. The proposed modified hierarchical residual attention learning method has been mentioned in Equation 5. The proposed model cum architecture of a conjoined twin neural network (CTNN) with a hierarchical initiated residual attention module for Alzheimer's stage classification is shown in Figure 1.

$$Y(x) = \sum_{u} T_u(x) * F(x) + F(x), \tag{5}$$

where $T_u\ (x)$ is a hierarchical attention module map with u as the number of attention module maps, $Y(x)$ is the output attention module aware feature set, and $F(x)$ is the features of an MRI image x from the convolutional layer. Here, we combined $F(x)$ with an existing attention mechanism for upgrading the CTNN performance. Algorithm 2 processes all the steps involved in the proposed CTNN model. Algorithm 3 includes all the steps of the attention learning-initiated multistage correlation filter.

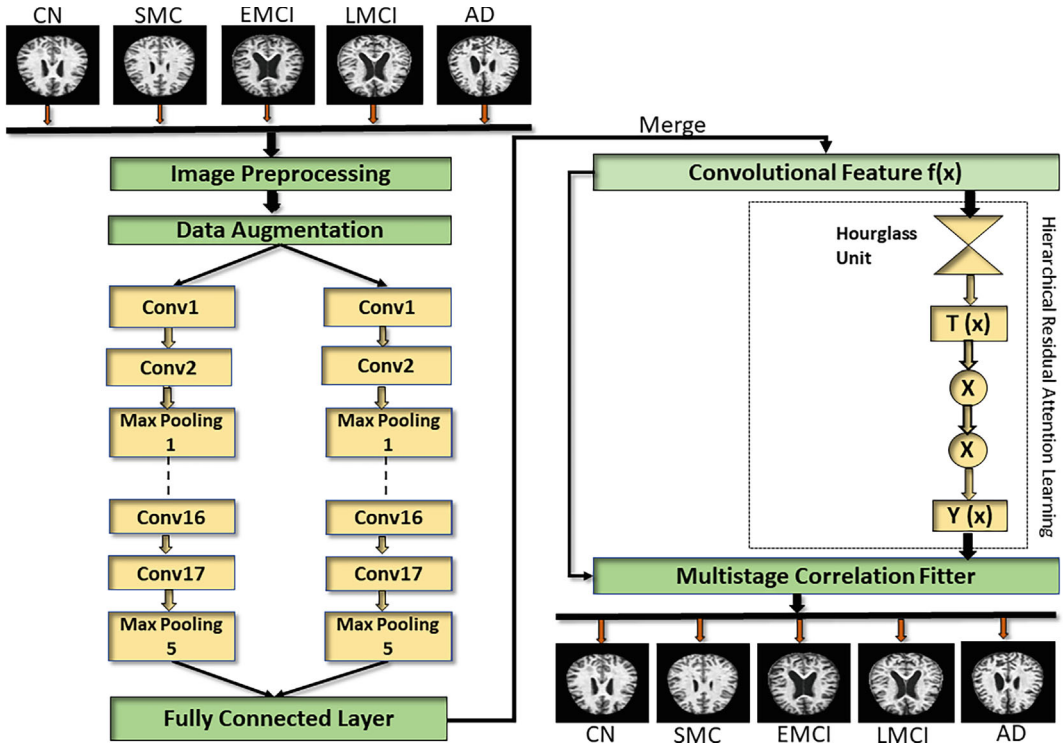

FIGURE 1 Proposed architectural model cum framework.

Algorithm 2. Pseudocode of proposed CTNN model

procedure CTNN **(Input: A_n, Neurons, Epoch, Repeat)**
  Input: $X_F \leftarrow \{A_n$: Attentive features; Neurons; A_f: Activation function$\}$
  Output: $C_d \leftarrow$ Classification {AD, EMCI, LMCI, SMC, CN}
 Train CTNN
  for each $X_{An} \in \{1, 2 \ldots \ldots, A_n\}$ **do**
   for Neurons = 1 to X_{An} **do**
    for Epoch = 1 to 30 **do**
     for Repeat = 1 to 30 **do**
      Train CTNN
       CTNN(TN) $\leftarrow$ (17 Conv, 2FC, 5MP, 4N, 4GN)
       A_f: Activation function $(H) = \frac{IH-KH+2(PD)}{St} + 1$
     end for
    L_w: Layers width$(W) = \frac{IW-KW+2(PD)}{St} + 1$
    Twin convolutional layer$(C) = \frac{IHW-F+2\ (PD)}{St} + 1$
    end for
   end for
  end for
 return C_d
end procedure

2.4 | Multistage correlation filter

Correlation filters[74,75] are the primary classifiers specifically optimized to have sharp peaks in the correlation output to accurately find the MRI images' features. Most correlation filter-based models are not context aware and do not generate minimal context information while tracing. This is generally because the feature area for each image is just outside the target area image. We have included a multistage correlation filter to improve the features tracking discriminant ability. This includes context-aware, regression-enabled, and specialized attribute adaptation MRI image tracking and extraction. We used a correlation filter as a layer in CNN. It has previously been demonstrated that the correlation filter strategy has been formulated in the form of the layer in a neural network. In an MRI image $i \in [1, I]$, we applied optimization with some objective functions about the correlation filter for parameter u and the regression variable v.

$$\arg\min_{u,v} \mathbb{O} = \|M_0\, u - v\|_2^2 + \theta_1\, \|u\|_2^2 + \theta_2\, \|v - v_0\|_2^2 + \theta_3 \sum_{x=1}^{p} \|M_x\, u\|_2^2, \tag{6}$$

where v_0 is the constraint-based relative matrix that helps to maintain the regression variable v, M_0 is the sample of MRI image from I ($M_0 \in I$), M_x is the feature context information concerningM_0, with θ_1, θ_2, and θ_3 used to reduce overfitting or prevent it, known as the regularization parameter, and m_0 and m_x are the base sample of M_0 and M_x circulant matrix. We have modified the objective function given in Equation 6 using a discrete Fourier transform and convex function with M_0 and M_x in the form of a circulant and converting the u filter parameter in a modified form of a multistage filter (Equation 7).

$$\hat{u} = \frac{\theta_2\ \left(\ \hat{m}_0 \odot \hat{v}_0\right)}{\theta_2\ \left(\ \hat{m}_0^* \odot \hat{m}_0 + \theta_3 \sum_{x=1}^{p} \hat{m}_x^* \odot \hat{m}_x\ \right) + \theta_1\ (\ 1 + \theta_2)}, \tag{7}$$

$\hat{m}_0^*$ is the conjugate of $\hat{m}_0$, and $\hat{m}_x^*$ is the conjugate of $\hat{m}_x$ From DFT, the representation of $\odot$ is a Hadamard product or entry-wise product.

Algorithm 3. Pseudocode of novel hierarchical residual attention learning with correlation filter

1: **Input:** $A_n \leftarrow \{F(x)$: MRI features; $T(x)$: Attention map; Hourglass Unit$\}$
2: **Output:** $Y(x) \leftarrow$ attention module aware feature set
3: **While** $F(x) > 0$ **do {**
4: **for** each $X_{An} \in \{1, 2 \ldots\ldots, x\}$ **do**
5: **for** attention map $T(x) = 1$ to n **do**
6: HU ← Hourglass Unit (point estimation)
7: $Y(x) = T(x) * F(x)$
8: **for** hierarchical attention $T_u(x) = 1$ to m **do**
9: Revised HU ← Revised Hourglass Unit (point estimation)
10: $Y(x) = \sum_u T_u(x) * F(x) + F(x)$
11: **end for**
12: **end for**
13: **Generate correlation filter** ($\hat{u}$,v)
14: $\arg\min_{u,v} \mathbb{O} = \|M_0\, u - v\|_2^2 + \theta_1\, \|u\|_2^2 + \theta_2\, \|v - v_0\|_2^2 + \theta_3 \sum_{x=1}^{p} \|M_x\, u\|_2^2$

15: $\hat{u} = \frac{\theta_2 \left(\hat{m}_0 \odot \hat{v}_0 \right)}{\theta_2 \left(\hat{m}_0^* \odot \hat{m}_0 + \theta_3 \sum_{x=1}^{p} \hat{m}_x^* \odot \hat{m}_x \right) + \theta_1 \left(1 + \theta_2 \right)}$

16: **end for**

17: }

18: **return** $Y(x)$

2.5 | Normalization and regularization

Normalization adjusts the data, while regularization adjusts the prediction function.[76–78] When the previously used layer parameter in the CNN is changed, the input of individual layer changes; therefore, the CNN model training is complex. Consequently, the activation function we have used in our model as ReLU quickly drops the gradient. This happens because the deep learning model's learning rate is degraded, and the model slackens progressively. We used hybrid layer normalization to overcome the same problem with the learning rate and deep learning model performance. Hybrid layer normalization decreased the parameter initialization and improved the learning rate. The variance and mean are captured in the input layer during parameter initialization, and the internal covariant is shifted to mitigate it. We formulated the Hybrid Layer Normalization (HLN) in Equations 8–12.

$$l_i = \mathrm{HLN}_{\gamma,\beta,\alpha}\left(r_i\right), \tag{8}$$

$$\mu_s = \frac{1}{k}\sum_{i=1}^{k} r_i, \tag{9}$$

$$\sigma_s^2 = \frac{1}{k}\sum_{i=1}^{k}(r_i - \mu_s)^2, \tag{10}$$

$$\overline{r_i} = \frac{r_i - \mu_s}{\sqrt{\sigma_s^2 + \epsilon}}, \tag{11}$$

$$\left.\begin{aligned} L_i &= \sum l_i \\ l_i &= \gamma.\overline{r_i} + \beta + \alpha \end{aligned}\right\}, \tag{12}$$

where k shows the number of layered batches and μ and σ^2 represent the mean and variance captured in the input layer, respectively. r_i shows each row of covariants with the help of Equations 8–10. We evaluated the μ and r^2 of each activation under layered batches. Equations 11 and 12 are three hyperparameters γ, β, and α, which improve the learning rate. We have incorporated Gaussian distribution noise[79] to improve the regularization and efficiency of the proposed work.

The distribution of the proposed CTNN model is shown in Table 3. The proposed architectural model cum framework is presented in Figure 1. We have presented the overall pseudocode of the proposed model in Algorithms 1–3.

3 | RESULTS

In our proposed work, we evaluate the ability of discriminant features captured by the HRAL. We implement multiclass classification using CN, SMC, LMCI, EMCI, and AD classes. We find the

TABLE 3 The proposed CTNN model layers with a correlation filter.

Layer	Name of layer	Size of kernel	Size of pool	Convolutional filters
1	Conv2D1 + ReLU(f)	3		64
	Normalization			
2	Conv2D2 + ReLU (f)	3		64
	Max-pooling1		2	
3	Conv2D3 + ReLU(f)	3		128
	Gaussian distributed noise			
	Normalization			
4	Conv2D4 + ReLU(f)	3		128
	Max-pooling2		2	
5	Conv2D5 + ReLU(f)	3		256
	Normalization			
6	Conv2D6 + ReLU(f)	3		256
	Gaussian distributed noise			
7	Conv2D7 + ReLU(f)	3		256
8	Conv2D8 + ReLU(f)	3		256
9	Conv2D9 + ReLU(f)	3		256
	Max-pooling3		2	
10	Conv2D10 + ReLU(f)	3		512
11	Conv2D11 + ReLU(f)	3		512
	Gaussian distributed noise			
12	Conv2D12 + ReLU(f)	3		512
13	Conv2D13 + ReLU(f)	3		512
	Max-pooling4		2	
14	Conv2D14 + ReLU(f)	3		512
	Gaussian distributed noise			
15	Conv2D15 + ReLU(f)	3		512
	Normalization			
16	Conv2D16 + ReLU(f)	3		512
17	Conv2D17 + ReLU(f)	3		512
	Max-pooling5		2	
18	Flattening 1			
19	Flattening 2			
20	Merge (2)			
21	Fully connected + ReLU4096			
22	Fully connected + ReLU4096			
23	Softmax (3)			
24	Correlation filter (4)			

training accuracy, training loss, validation accuracy, validation loss, accuracy with normalization, correlation filter, and parameter tuning. Following each iteration of the optimization process, the loss value of a model reveals how well or how poorly the model performs. An accuracy measure is employed to evaluate an algorithm effectively for the proposed model performance. After setting up the parameter tuning of a model, it is common practice to arrive at an estimated accuracy for the model. In our case, this step is performed after the model has already been validated. We have hypertuned the parameter of Hybrid Layer Normalization as γ, β, and α and obtained a stable highest result at $\alpha = 0.33$, $\beta = 0.32$, $\gamma = 0.35$. The procedure of choosing the best possible values for a learning algorithm is recognized as "hyperparameter tuning." A model argument whose value is figured out before learning is called a hyperparameter. Tuning the algorithm's hyperparameters is important to improve deep learning performance. To evaluate the evaluation metrics, we evaluated the accuracy, precision, recall, or sensitivity, f1 score, and specificity for all possible model combinations. Based on the model assessment, we evaluated the confusion matrix. Confusion matrices are measurements often used to show how well a classification model (also called a "classifier") did on a set of test data for which the actual values were already known. It is done by comparing the model's predicted values to the actual ones. We have considered evaluation performance or assessment metrics in Equations 13 and 14. The accuracy is evaluated as follows:

$$\text{Accuracy } (A) = \frac{\text{CO}_p}{\text{All}_p}, \tag{13}$$

where CO_p is the correct prediction and All_p is all predictions.

The precision, recall, F1 score, and specificity are evaluated as follows:

$$\left.\begin{aligned} \text{Precision } (P) &= \frac{\text{TP}}{\text{TP+FP}} \\ \text{Recall } (R) \| \text{ Sensitivity } (S) &= \frac{\text{TP}}{\text{TP+FN}} \\ \text{Specificity(Sp)} &= \frac{\text{TN}}{\text{TN+FP}} \\ \text{F1 Score(F1)} &= \frac{2\ PR}{\text{P+R}} \end{aligned}\right\}, \tag{14}$$

where TP is the true positive, FP is the false positive, FN is the false negative, and TN is the true negative.

This paper proposes HRAL-based CTNN model for training and extracting attention-aware features for classifying Alzheimer's disorder. Figure 2A evaluates the validation and training accuracy by balancing data using data augmentation and Gaussian distribution noise on the Brain MR Images. We have executed the proposed work for 30 epochs and evaluated the maximum validation accuracy of 99.92% and test accuracy of 99.15%. The maximum validation accuracy showed no overfitting in the proposed model.

Figure 2B shows that the model validation loss decreases if we increase epochs (best at 30 epochs), and at the 30th epoch, the validation loss is less than the training loss. Figure 2C presents switch normalization, group normalization, batch normalization, and layered batch normalization. We compared the training loss evaluated from all four normalizations at different epochs. Subsequently, the proposed layered batch normalization produces less loss with higher accuracy than other normalization techniques. In Figure 2D, we evaluated each normalization technique's validation accuracy, which shows layered batch normalization with a validation accuracy of 99.92%. Figure 2E,F present the validation and training accuracy after correlation filtering. The validation accuracy increased by 0.06% after correlation filtering and was evaluated as 99.97%.

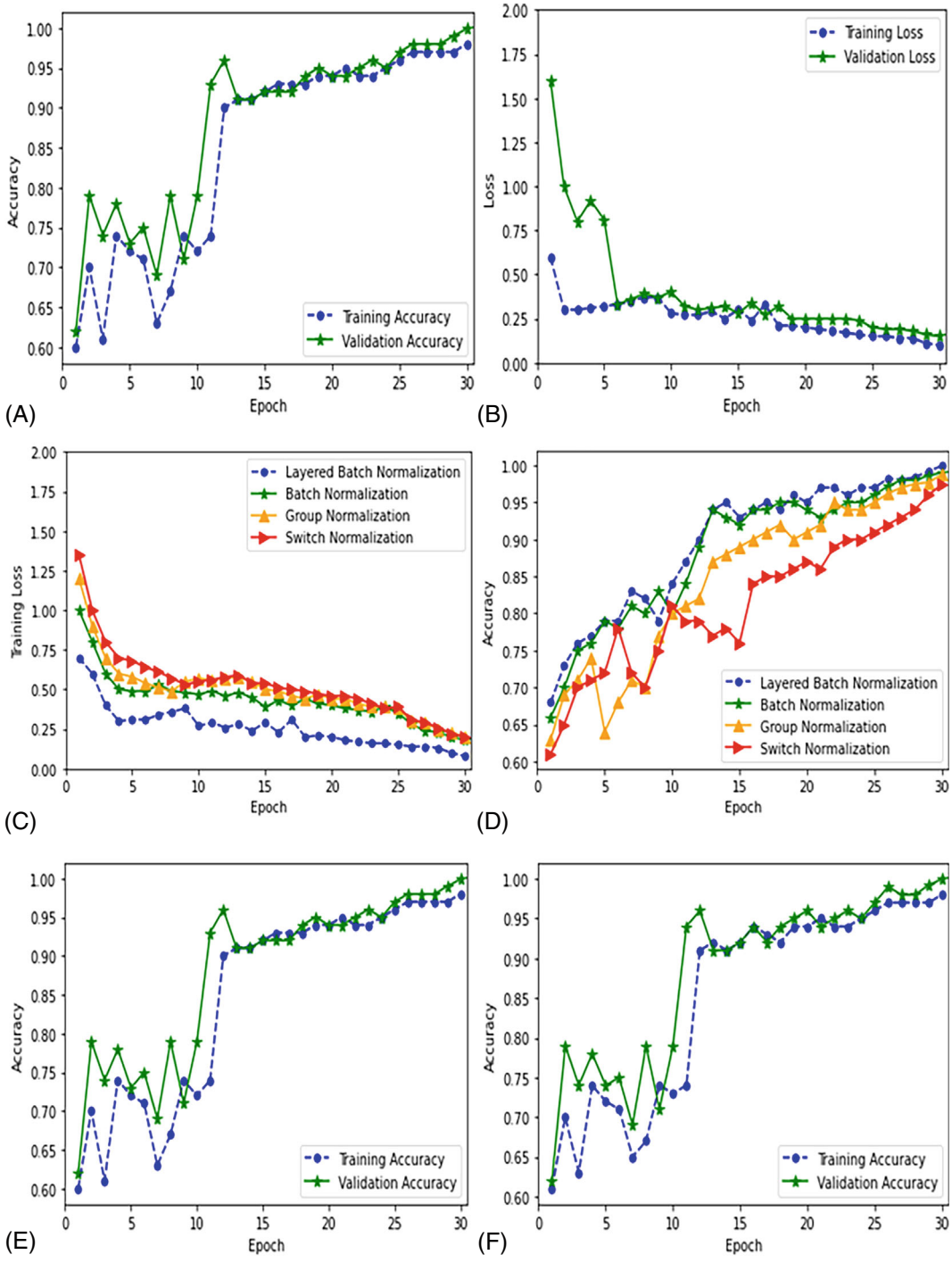

FIGURE 2 Proposed model accuracy and loss: (A) training accuracy and validation accuracy, (B) training loss and validation loss, (C) loss with normalization, (D) accuracy with normalization, (E) accuracy without correlation filter, (F) accuracy with correlation filter.

TABLE 4 Hyperparameter values (γ, β, and α) based on evaluation metrics of the proposed model.

Parameter value: γ, β, and α; Model: (CTNN + HRAL + MCF + NR)	Precision	Recall	Accuracy	F1-score	Specificity
$\alpha = 0.35, \beta = 0.32, \gamma = 0.33$	98.71 ± 0.18	97.91 ± 0.14	98.87 ± 0.07	98.26 ± 0.06	96.01 ± 0.19
$\alpha = 0.35, \beta = 0.33, \gamma = 0.32$	98.56 ± 0.13	97.25 ± 0.11	98.13 ± 0.03	97.99 ± 0.02	95.81 ± 0.15
$\alpha = 0.33, \beta = 0.32, \gamma = 0.35$	98.85 ± 0.19	97.94 ± 0.13	98.91 ± 0.08	98.37 ± 0.06	96.09 ± 0.18
$\alpha = 0.32, \beta = 0.35, \gamma = 0.33$	98.44 ± 0.13	97.21 ± 0.10	98.12 ± 0.03	97.98 ± 0.03	95.71 ± 0.17
$\alpha = 0.33, \beta = 0.32, \gamma = 0.35$	**99.62 ± 0.11**	**98.98 ± 0.10**	**99.97 ± 0.01**	**99.30 ± 0.02**	**97.21 ± 0.12**
$\alpha = 0.32, \beta = 0.33, \gamma = 0.35$	99.22 ± 0.10	98.54 ± 0.12	99.18 ± 0.02	99.05 ± 0.01	97.13 ± 0.10
$\alpha = 0.5, \beta = 0.5, \gamma = 0.0$	96.98 ± 0.14	96.33 ± 0.11	97.22 ± 0.05	97.09 ± 0.01	95.61 ± 0.09
$\alpha = 0.5, \beta = 0.0, \gamma = 0.5$	97.12 ± 0.17	97.71 ± 0.12	97.82 ± 0.06	97.29 ± 0.02	96.67 ± 0.11
$\alpha = 0.0, \beta = 0.5, \gamma = 0.5$	96.99 ± 0.15	96.59 ± 0.15	97.41 ± 0.06	97.13 ± 0.02	95.91 ± 0.10

Note: Bold values are best values of hyperparameter (γ,β,and α).

TABLE 5 Overall evaluation metrics comparison with different models.

Evaluation metrics / Models	Precision	Recall-sensitivity	Accuracy	F1 score	Specificity
Only CNN	94.35 ± 0.15	94.00 ± 0.13	95.00 ± 0.08	94.18 ± 0.04	93.46 ± 0.12
3D-CNN	97.00 ± 0.13	97.15 ± 0.11	98.00 ± 0.05	97.08 ± 0.02	96.19 ± 0.10
CNN + HRAL	97.77 ± 0.14	97.15 ± 0.10	97.80 ± 0.03	97.46 ± 0.03	97.00 ± 0.09
CTNN + HRAL	98.00 ± 0.19	97.30 ± 0.11	98.26 ± 0.04	97.65 ± 0.02	96.77 ± 0.08
CTNN + HRAL + MCF + NR	99.62 ± 0.11	98.98 ± 0.10	99.97 ± 0.01	99.30 ± 0.02	97.21 ± 0.12

For stagewise classification, we showed the combined confusion matrix covering stages such as cognitive normal (CN), significant memory concern (SMC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and Alzheimer's disease (AD). Table 4 presents the hyperparameter values (γ, β, and α) based on evaluation metrics of the best combination with stable validation accuracy. We evaluated a combined F1 score of $99.30 \pm 0.02\%$ from the same confusion matrix, the harmonic mean between the precision of $99.62 \pm 0.11\%$ and recall of $98.98 \pm 0.10\%$. Table 5 presents the evaluation metrics and comparison with the combination of models. In Figure 3, we have covered a different combination of the model's confusion matrices with their comparisons.

One kind of resampling is called cross-validation,[23,50,60] a technique that includes testing and training a model with variable iterations using diverse data subsets. Its principal aim, in which prediction is the goal, as well as an extensive aim, is to evaluate the degree of accuracy a predictive model will accomplish when it is put into action. Cross-validation is a tried-and-true method for preventing the practice of overfitting in statistical and computational models. The entire collection of data has been segmented into several parts. To carry out standard K-fold cross-validation, the data must be partitioned into k folds first. Then, we put the algorithm through its paces by repeatedly folding $k-1$ sets of data while using the remaining holdout fold as a test set. We use cross-validation in our proposed model to prevent overfitting. We use six cross-validation methods

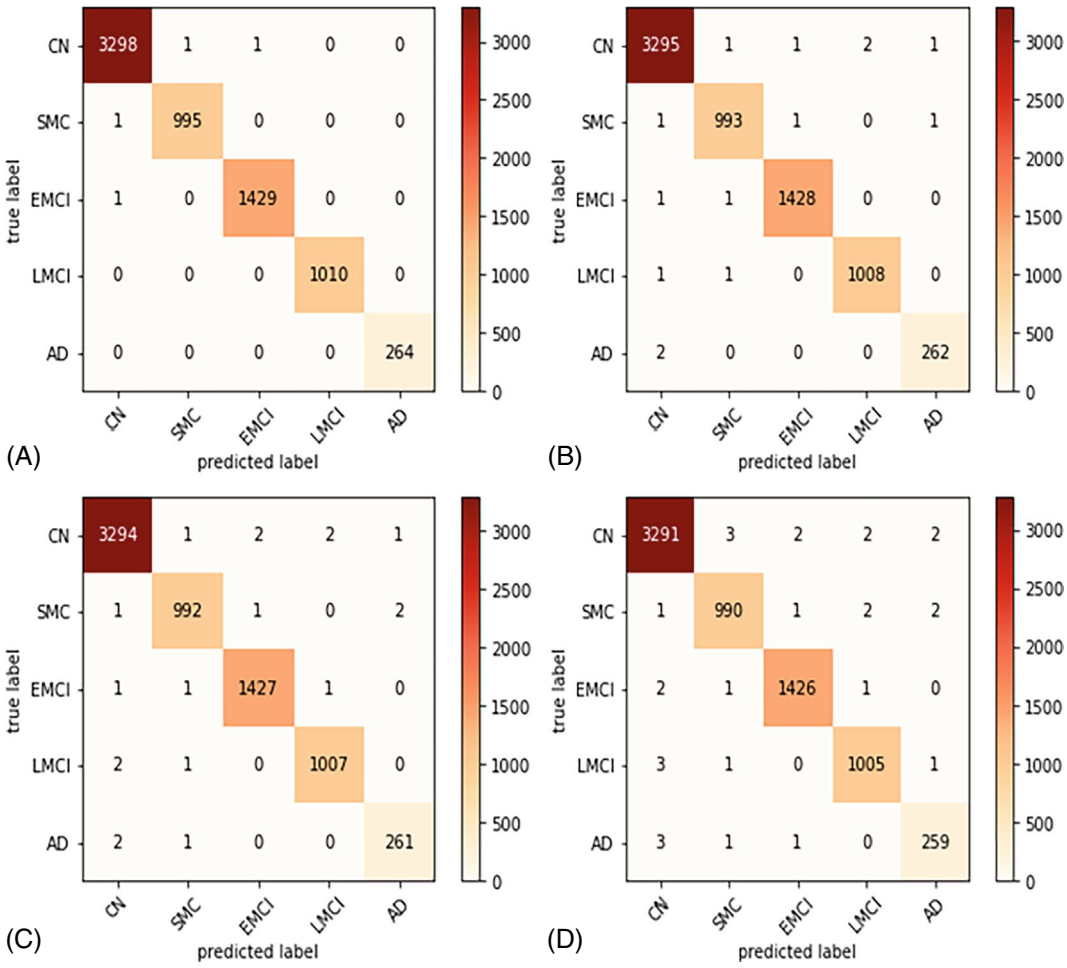

FIGURE 3 Models comparison with confusion matrix: (A) CTNN + HRAL + MCF + NR, (B) CTNN+HRAL, (C) CNN + HRAL, (D) Only CNN.

with different k-values/p-values to find accuracy in Table 6. We have chosen the p-value and k-value as per the static accuracy score. From Table 6, the proposed model has no overfitting and gives the approximate result at $k = 10$ and $p = 200$ compared to Table 5.

4 | DISCUSSION

We have compared the model developed by our group with most of the literature. Kruthika et al.[37] used the concept of a three-dimensional CNN. Computer-assisted diagnostic methods using content-based image retrieval (CBIR) were used to make diagnoses of Alzheimer's disease. The model was evaluated with an imbalanced dataset and correlated features, which caused overfitting and biases in the model. The author evaluated the 98.42% accuracy of their model. Rachna et al.[44] approached the problem by applying transfer learning with a CNN. They trained VGG-16 on the ImageNet dataset, which was used for feature extraction in the classification model. The author reported the same transfer learning-based CNN model with 95.73% accuracy. Basaia et al.[57]

TABLE 6 Accuracy with different cross validation techniques.

***k*-value**	***k* = 3**	***k* = 5**	***k* = 10**	***k* = 15**	***k* = 20**
Cross-validation					
k-folds	97.92 ± 0.28	98.90 ± 0.17	99.58 ± 0.02	99.08 ± 0.03	98.14 ± 0.02
Repeated *k*-folds	98.18 ± 0.15	98.94 ± 0.12	99.94 ± 0.01	99.17 ± 0.07	98.11 ± 0.05
Nested *k*-folds	98.23 ± 0.13	98.93 ± 0.14	99.95 ± 0.02	99.19 ± 0.06	98.14 ± 0.03
Stratified *k*-folds	98.25 ± 0.11	98.93 ± 0.14	99.96 ± 0.03	99.21 ± 0.06	98.15 ± 0.01
***p*-value**	***p* = 1**	***p* = 50**	***p* = 100**	***p* = 200**	***p* = 500**
Leave-one-out	99.95 ± 0.03	—	—	—	—
Leave-*p*-out	—	98.90 ± 0.11	99.18 ± 0.05	99.94 ± 0.04	98.12 ± 0.05

developed a framework based on a CNN, which was focused on mild cognitive impairment (MCI) magnetic resonance imaging. The author reported 75% accuracy of the CNN model. Asl et al.[80] used the 3D-DSA-CNN-based deep learning method with classification accuracy of 97.60% for each stage. They proposed that AD prediction improved with a deep 3D convolutional neural network (3D-CNN) to display the generic features of apprehending AD biomarkers taken from MR brain images. Goceri[58] presented a gradient-based stochastic optimizer-oriented 3D-CNN for AD diagnosis with 98.01% accuracy. The model discussed by Shuangshuang et al.[47] is a patch- and ROI-based feature extraction with deep learning-based model implementation. The limitations of this model are that they have not covered all stages of Alzheimer's, and there is no feature discrimination, due to which there is less accuracy than 90% reported. Shankar et al.[48] presented a novel shared correlation–based feature selection process. They used a discriminative feature selection-initiated supervised machine learning model. They have reported 94%–96% accuracy, but this work's limitation does not include automatic feature extraction, and the data are imbalanced in this model. A review was projected by Grueso et al.,[49] in which they covered review concepts from different studies and gave a comparative analysis of them. A result analysis[54] shows the DL approach, specifically CNN, with an average reporting accuracy between 95%–97%. Only four stages are covered in this study, and the rest of the stages of Alzheimer's are not included in this approach. Due to redundant features extracted from MRI, this model also reports overfitting. A manifold structure[55] of AD by Martinez-Murcia et al. reported prediction accuracy of 80% with two stages of Alzheimer's disease (AD and NC). In the case of the deep learning model, this accuracy is lower, as many discriminative features are not extracted during the model training. A small image dataset is also a significant issue for the lower accuracy of this model. Wen et al.[56] proposed a hybrid transfer learning model with CNN and reported 85%–86% accuracy. They used fivefold cross-validation to validate the data. They have not trained hyperparameters, and the dataset is too minimal to work with a deep learning model. They have performed the classification task with binary AD, and multiclass AD needs to be added to distinguish all stages of Alzheimer's. As discussed by many studies, their proposed method does not cover Alzheimer's stage detection among all stages of Alzheimer's and discriminative selection of features. On average, most studies reported accuracy of 75.4%–78.5%.

The proposed CTNN-based model outperformed existing state-of-the-art models. For stage-wise classification, we covered stages such as cognitive normal (CN), significant memory concern

TABLE 7 Comparative study with the latest literature in the same dataset cohorts.

Author	Methodology	Dataset/Type	Accuracy	Overfitting in model	Data balance	Discriminative feature set	Stages covered
Shuangshuang et al.[47]	Biomarkers-based neural network	ADNI/MRI	90%	✓	×	×	02
Grueso et al.[49]	Deep learning models	ADNI/MRI	75.4%–78.5%.	✓	×	×	02
Helaly et al.[54]	CNN	ADNI/MRI	95%–97%.	✓	×	×	04
Shankar et al.[48]	Supervised classifier	ADNI/MRI	96.00%	✓	×	✓	04
Wen et al.[56]	Transfer learning with CNN	ADNI/MRI	85%–86%	✓	×	×	02
Martinez-Murcia et al.[55]	Deep convolutional autoencoders	ADNI/MRI	80%	✓	×	✓	02
Kruthika et al.[37]	CBIR based CNN	ADNI/MRI	98.42%	✓	×	×	03
Rachna et al.[44]	VGG16 based CNN	ADNI/MRI	95.73%	✓	×	×	02
Basaia et al.[57]	CNN	ADNI/MRI	75%	✓	×	×	02
Asl et al.[80]	3D-CNN based model	ADNI/MRI	97.60%	✓	×	×	03
Goceri[58]	3D-Convolutional Network	ADNI/MRI	98.01%	✓	×	×	02
Proposed model	HRAL based CTNN	ADNI/MRI	99.97 ± 0.01	×	✓	✓	05

TABLE 8 Time spent (elapsed time) comparison of different model's inference.

Models	Individual subjects	Elapsed time (hh: mm: ss)
Only CNN	7000	02: 47: 41
3D-CNN	7000	01: 34: 19
CNN + HRAL	7000	01: 15: 12
CTNN + HRAL	7000	00: 59: 13
CTNN + HRAL + MCF + NR	7000	00: 46: 09

(SMC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and Alzheimer's disease (AD). We have evaluated and presented our model result in Figure 2 with 99.97% accuracy in the same dataset cohorts. In Table 5, we show the evaluation metrics scores of our model with precision of 99.62%, recall of 98.98%, F1-score of 99.30, and specificity of 97.21%. Table 7 shows the comparative study of our work with the latest literature.

The proposed model has also been validated, compared, and evaluated time spent on overall model inference in terms of elapsed time. Table 8 shows the elapsed time comparison of the proposed HRAL-CTNN model with existing models. The proposed model shows less elapsed time compared to existing models.

5 | CONCLUSION

This paper proposes HRAL-based CTNN deep learning architecture for multistage Alzheimer's classification using a modified VGG19 conjoined twin model. We have covered stages of Alzheimer's, such as CN, SMC, EMCI, LMCI, and AD. This study focused on unbalanced data and overfitting the model using data augmentation and hierarchical initiated residual attention learning. We also used discriminant features and analyzed them using a correlation filter. We put forward a novel multistage conjoined neural network-based CNN that utilizes the hierarchical residual attention-based model to reach high-performance efficiency for Alzheimer's disorder diagnosis. To regularize and learn parameters, we used switch-group-batch normalization. The ADNI MRI dataset was used to assess the implementation feasibility of the proposed work. The proposed HRAL-based CTNN attained accuracy of $99.97 \pm 0.01\%$ and F1 score of $99.30 \pm 0.02\%$ for Alzheimer's stage classification. The proposed model outperforms the previously published research about model accuracy and performance scores in this area. This work has a weakness: it cannot make complex decisions beyond earlier training. In the future, we would like to work with more discriminant features and build a large dataset-based transfer learning model.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data sets used in the experiment are publicly available.

ORCID

Venkatesh Gauri Shankar https://orcid.org/0000-0002-0088-5887

Dilip Singh Sisodia https://orcid.org/0000-0001-9845-290X
Preeti Chandrakar https://orcid.org/0000-0002-7387-1582

REFERENCES

1. Beheshti I, Demirel H, Neuroimaging, D. Feature-ranking-based Alzheimer's disease classification from structural MRI. *Magn Reson Imaging*. 2016;34:252-263.
2. Mitelpunkt A, Galili T, Kozlovski T, et al. Novel Alzheimer's disease subtypes identified using a data and knowledge driven strategy. *Sci Rep*. 2020;10:1327. doi:10.1038/s41598-020-57785-2
3. Bondi MW, Edmonds EC, Salmon DP. Alzheimer's disease: past, present, and future. *J Int Neuropsychol Soc*. 2017;23(9–10):818-831. doi:10.1017/S135561771700100X
4. Alzheimer's Association. 2019 Alzheimer's disease facts and figures. *Alzheimers Dement*. 2019;15:321-387.
5. Aisen PS, Petersen RC, Donohue M, Weiner MW, Alzheimer's Disease Neuroimaging Initiative. Alzheimer's disease neuroimaging initiative 2 clinical core: progress and plans. *Alzheimers Dement*. 2015;11(7):734-739. doi:10.1016/j.jalz.2015.05.005
6. Fan Z, Xu F, Qi X, Li C, Yao L. Classification of Alzheimer's disease based on brain MRI and machine learning. *Neural Comput Appl*. 2020;32:1927-1936. doi:10.1007/s00521-019-04495-0
7. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. *Neurology*. 2010;74(3):201-209. doi:10.1212/WNL.0b013e3181cb3e25
8. Brookmeyer R, Johnson E, Ziegler-graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement*. 2007;3:186-191.
9. Alirezaie J, Jernigan ME, Nahmias C. Automatic segmentation of cerebral MR images using artificial neural networks. *IEEE Trans Nucl Sci*. 1998;45:2174-2182.
10. Hosseini-Asl E, Keynton R, El-Baz A. Alzheimer's disease diagnostics by adaptation of 3D convolutional network. Proceedings of the International Conference on Image Processing, Arizona, AZ, USA, 25–28 September 2016; pp. 126–130.
11. Oikarinen H, Karttunen A, Pääkkö E, Tervonen O. Survey of inappropriate use of magnetic resonance imaging. *Insights Imaging*. 2013;4:729-733. doi:10.1007/s13244-013-0276-2
12. Gupta A, Ayhan MS, Maida AS. Natural image bases to represent neuroimaging data. Proceedings of the 30th International Conference on International Conference on Machine Learning, 28 (ICML'13). JMLR.org, III–987–III–994.
13. Zeng K, Yang Y, Xiao G, Chen Z. A very deep densely connected network for compressed sensing MRI. *IEEE Access*. 2019;7:85430-85439. https://ieeexplore.ieee.org/document/8744318
14. Ben Ahmed O, Benois-Pineau J, Allard M, Ben Amar C, Catheline G. Classification of Alzheimer'sdisease subjects from MRI using hippocampal visual features. *Multimed Tools Appl*. 2014;74:1249-1266.
15. Klette R. *Concise Computer Vision*. Springer; 2014.
16. Bishop CM. *Pattern Recognition and Machine Learning*. Springer; 2006.
17. Shankar VG, Sisodia DS, Chandrakar P. An efficient MR images based analysis to predict Alzheimer's dementia stage using random Forest classifier. *Lecture Notes in Networks and Systems*. Vol 521. Springer; 2023. doi:10.1007/978-3-031-13150-9_9
18. Stonnington CM, Chu C, Klöppel S, et al. Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. *Neuroimage*. 2010;51(4):1405-1413.
19. Li S, Shi F, Pu F, et al. Hippocampal shape analysis of Alzheimer disease based on machine learning methods. *Am J Neuroradiol*. 2007;28(7):1339-1345.
20. Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J. Machine learning frame- work for early mri-based alzheimer's conversion prediction in mci subjects. *Neuroimage*. 2015;104(1):398-412.
21. Jha D, Kim JI, Kwon GR. Diagnosis of Alzheimer's disease using dual-tree complex wavelet transform, pca, and feed-forward neural network. *J Healthc Eng*. 2017;2017(5):1-13.
22. Plant C, Teipel SJ, Oswald A, et al. Automated detection of brain atrophy patterns based on MRI for the prediction of alzheimer's disease. *Neuroimage*. 2010;50(1):162-174.
23. Goryawala M, Zhou Q, Barker W, Loewenstein DA, Duara R, Adjouadi M. Inclusion of neuropsychological scores in atrophy models improves diagnostic classification of Alzheimer's disease and mild cognitive impairment. *Comput Intell Neurosci*. 2015;2015(1):1-14.

24. Nozadi SH, Kadoury S. Classification of Alzheimer's and mci patients from semantically parcelled pet images: a comparison between av45 and fdg-pet. *Int J Biomed Imaging*. 2018;2018(1):1-13.
25. Jie B, Liu M, Shen D. Integration of temporal and spatial properties of dynamic connectivity networks for automatic diagnosis of brain disease. *Med Image Anal*. 2018;47(1):81-94.
26. Bi X, Xu Q, Luo X, Sun Q, Wang Z. Analysis of progression toward Alzheimer's disease based on evolutionary weighted random support vector machine cluster. *Front Neurosci*. 2018;12(October):1-11.
27. Ren J, Zhang B, Wei D, Zhang Z. Identification of methylated gene biomarkers in patients with Alzheimer's disease based on machine learning. *Biomed Res Int*. 2020;2020:1-11. doi:10.1155/2020/8348147
28. Sharma D, Soni N, Devi B, Shankar VG. Predem: a computational framework for prediction of early dementia using deep neural networks. *Procedia Comput Sci*. 2022;215:697-705. doi:10.1016/j.procs.2022.12.071
29. Singh AP, Upadhyay N, Shankar VG, Devi B. IHDNA: identical hybrid deep neural networks for Alzheimer's detection using MRI dataset. 2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT), Jaipur, India, 2023, pp. 1–7. doi:10.1109/ICCT56969.2023.10075912
30. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444.
31. Suk H, Lee S, Shen D. Deep ensemble learning of sparse regression models for brain disease diagnosis. *Med Image Anal*. 2017;37:101-113.
32. Hong X, Lin R, Yang C, et al. Predicting Alzheimer's disease using LSTM. *IEEE Access*. 2019;7:80893-80901. doi:10.1109/ACCESS.2019.2919385
33. Taheri Gorji H, Kaabouch N. A deep learning approach for diagnosis of mild cognitive impairment based on MRI images. *Brain Sci*. 2019;9:217.
34. Nawaz H, Maqsood M, Afzal S, Aadil F, Mehmood I, Rho S. A deep feature-based real-time system for Alzheimer disease stage detection. *Multimed Tools Appl*. 2020;80:35789-35807. doi:10.1007/s11042-020-09087-y
35. Zhang B, Lei Z, Sun J, Zhang H. Cross-media retrieval with collective deep semantic learning. *Multimed Tools Appl*. 2018;77(17):22247-22266.
36. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88.
37. Kruthika KR, Rajeswari HD, Maheshappa HD. CBIR system using capsule networks and 3D CNN for Alzheimer's disease diagnosis. *Inform Med Unlocked*. 2019;14:59-68. doi:10.1016/j.imu.2018.12.001
38. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition, International Conference on Learning Representations. 2015.
39. Saikia AR, Bora K, Mahanta LB, Das AK. Comparative assessment of CNN architectures for classification of breast FNAC images. *Tissue Cell*. 2019;57:8-14. doi:10.1016/j.tice.2019.02.001
40. Liu S, Deng W. Very deep convolutional neural network based image classification using small training sample size. Proceedings of the 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, Malaysia, 3–6 November 2015.
41. Yamashita R, Nishio M, Do RKG, et al. Convolutional neural networks: an overview and application in radiology. *Insights Imaging*. 2018;9:611-629. doi:10.1007/s13244-018-0639-9
42. Kim H, Jeong YS. Sentiment classification using convolutional neural networks. *Appl Sci*. 2019;9:2347.
43. Sui X, Zheng Y, Wei B, et al. Choroid segmentation from optical coherence tomography with graph edge weights learned from deep convolutional neural networks. *Neurocomputing*. 2017;237:332-341.
44. Jain R, Jain N, Akshay Aggarwal D, Hemanth J. Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images. *Cogn Syst Res*. 2019;57:147-159. doi:10.1016/j.cogsys.2018.12.015
45. Baron JC, Chételat G, Desgranges B, et al. In vivomapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease. *Neuroimage*. 2001;14:298-309.
46. Al-Shoukry S, Rassem TH, Makbol NM. Alzheimer's diseases detection by using deep learning algorithms: a mini-review. *IEEE Access*. 2020;8:77131-77141. doi:10.1109/ACCESS.2020.2989396
47. Gao S, Lima D. A review of the application of deep learning in the detection of Alzheimer's disease. *Int J Cogn Comput Eng*. 2022;3:1-8. doi:10.1016/j.ijcce.2021.12.002
48. Shankar VG, Sisodia DS, Chandrakar P. A novel discriminant feature selection–based mutual information extraction from MR brain images for Alzheimer's stages detection and prediction. *Int J Imaging Syst Technol*. 2021;32:1-20. doi:10.1002/ima.22685

49. Grueso S, Viejo-Sobera R. Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: a systematic review. *Alzheimers Res Ther*. 2021;13:162. doi:10.1186/s13195-021-00900-w
50. Sarraf S, Anderson J, Tofighi G, et al. DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI. *bioRxiv*. 2016;070441.
51. Payan A, Montana G. Predicting Alzheimer's disease: a Neuroimaging study with 3D convolutional neural networks, arXiv Preprint arXiv:1502.02506.
52. Farooq A, Anwar S, Awais M, Alnowami M. Artificial intelligence based smart diagnosis of Alzheimer's disease and mild cognitive impairment, Smart Cities Conference (ISC2), 2017 International, IEEE, 2017, pp. 1–4.
53. Chen X-W, Lin X. Big data deep learning: challenges and perspectives. *IEEE Access*. 2014;2:514-525.
54. Helaly HA, Badawy M, Haikal AY. Deep learning approach for early detection of Alzheimer's disease. *Cogn Comput*. 2021;21:1-17. doi:10.1007/s12559-021-09946-2
55. Martinez-Murcia FJ, Ortiz A, Gorriz JM, Ramirez J, Castillo-Barnes D. Studying the manifold structure of Alzheimer's disease: a deep learning approach using convolutional autoencoders. *IEEE J Biomed Health Inform*. 2020;24:17-26. doi:10.1109/JBHI.2019.2914970
56. Wen J, Thibeau-Sutre E, Diaz-Melo M, et al. Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation. *Med Image Anal*. 2020;63:101694.
57. Basaia S, Agosta F, Wagner L, et al. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage Clin*. 2019;21:101645. doi:10.1016/j.nicl.2018.101645
58. Goceri E. Diagnosis of Alzheimer's disease with Sobolev gradient-based optimization and 3D convolutional neural network. *Int J Numer Method Biomed Eng*. 2019;35:e3225. doi:10.1002/cnm.3225
59. Justin Joseph R. Periyasamy, a fully customized enhancement scheme for controlling brightness error and contrast in magnetic resonance images. *Biomed Signal Process Control*. 2018;39:271-283. doi:10.1016/j.bspc.2017.08.003
60. Yaroslavsky L. Nonlinear filters in image processing. *Digital Holography and Digital Image Processing*. Springer; 2004. doi:10.1007/978-1-4757-4988-5_12
61. LONI-ADNI. Alzheimer's Disease Neuroimaging Initiative (ADNI). http://adni.loni.usc.edu/data-samples/access-data/, Online Last Accessed: November 18, 2021
62. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment [published correction appears in J Am Geriatr Soc. 2019 Sep;67(9):1991]. *J Am Geriatr Soc*. 2005;53(4):695-699. doi:10.1111/j.1532-5415.2005.53221.x
63. Arevalo-Rodriguez I, Smailagic N, i Figuls MR, et al. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev*. 2015;2015(3):CD010783. doi:10.1002/14651858.CD010783.pub2
64. Kueper JK, Speechley M, Montero-Odasso M. The Alzheimer's disease assessment scale-cognitive subscale (ADAS-Cog): modifications and responsiveness in pre-dementia populations. A narrative review. *J Alzheimer's Dis*. 2018;63(2):423-444. doi:10.3233/JAD-170991
65. Rafel CG, Richard EW. *Digital Image Processing*. 2nd ed. Prince Hall; 2002.
66. King-Sun F, Rosenfeld. Pattern recognition and image processing. *IEEE Trans Comput*. 1976;C-25(12):1336-1346.
67. Khan RU, Tanveer M, Pachori RB. Alzheimer's disease Neuroimaging Initiative (ADNI). A novel method for the classification of Alzheimer's disease from normal controls using magnetic resonance imaging. *Expert Syst*. 2021;38:e12566. doi:10.1111/exsy.12566
68. Sangma JW, Sarkar M, Pal V. Hierarchical clustering for multiple nominal data streams with evolving behaviour. *Complex Intell Syst*. 2022;8:1737-1761. doi:10.1007/s40747-021-00634-0
69. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *J Big Data*. 2019;6:60. doi:10.1186/s40537-019-0197-0
70. Yang J, Zhao Y, Chan JCW, Yi C. The effectiveness of data augmentation in image classification using deep learning. arXiv 2017, arXiv:1712.04621.
71. Zhang X, Yang Q. Transfer hierarchical attention network for generative dialog system. *Int J Autom Comput*. 2019;16:720-736. doi:10.1007/s11633-019-1200-0

72. Xing S, Liu F, Wang Q, Zhao X, Li T. A hierarchical attention model for rating prediction by leveraging user and product reviews. *Neurocomputing*. 2019;332:417-427.
73. Melekhov I, Ylioinas J, Kannala J, Rahtu E. Image-based localization using hourglass networks. 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), 870–877. 2017.
74. Liu J, Luo Z, Xiong X. An improved correlation filter tracking method with occlusion and drift handling. *Vis Comput*. 2020;36:1909-1926. doi:10.1007/s00371-019-01776-6
75. Dai K, Wang D, Lu H. Visual tracking via adaptive spatially regularized correlation filters. Proceedings of the IEEE Computer Vision and Pattern Recognition, pp. 4670–4679. 2019.
76. Scheidegger F, Istrate R, Mariani G, Benini L, Bekas C, Malossi C. Efficient image dataset classification difficulty estimation for predicting deep-learning accuracy. *Vis Comput*. 2020;37:1593-1610. doi:10.1007/s00371-020-01922-5
77. Ramani S, Weller DS, Nielsen J, Fessler JA. Non-cartesian MRI reconstruction with automatic regularization via Monte-Carlo SURE. *IEEE Trans Med Imaging*. 2013;32(8):1411-1422. doi:10.1109/TMI.2013.2257829
78. Pandey G, Ghanekar U. Classification of priors and regularization techniques appurtenant to single image superresolution. *Vis Comput*. 2020;36:1291-1304. doi:10.1007/s00371-019-01729-z
79. Nadarajah S, Kotz S. Statistical distribution of the measure of coherence. *IEEE Trans Biomed Eng*. 2006;53(11):2409-2410. doi:10.1109/TBME.2006.883625
80. Hosseini-Asl E, Ghazal M, Mahmoud A, et al. Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network. *Front Biosci (Landmark Ed)*. 2018;23:584-596. doi:10.2741/4606

How to cite this article: Shankar VG, Sisodia DS, Chandrakar P. An intelligent hierarchical residual attention learning-based conjoined twin neural network for Alzheimer's stage detection and prediction. *Computational Intelligence*. 2023;39(5):783-805. doi: 10.1111/coin.12594