Communications in Mathematics and Applications
Vol. 14, No. 3, pp. 1263–1273, 2023
ISSN 0975-8607 (online); 0976-5905 (print)
Published by RGN Publications
DOI: 10.26713/cma.v14i3.2457

http://www.rgnpublications.com

Special Issue
Recent Trends in Mathematics and Applications
Proceedings of the International Conference of Gwalior Academy of Mathematical Sciences 2022
***Editors*: Vinod P. Saxena and Leena Sharma**

Research Article

Alzheimer's Disease Detection Model Using Rotation Invariant DTCWT Features and Recurrent Neural Network With a Combination of LSTM and GRU

Preeti Topannavar*[1], D. M. Yadav[1] and Varsha Bendre[2]

[1] *Department of Electronics & Telecommunication Engineering, G. H. Raisoni College of Engineering and Management (Savitribai Phule Pune University), Pune 412207, Maharashtra, India*
[2] *Department of Electronics & Telecommunication Engineering, Pimpri Chinchwad College of Engineering (Savitribai Phule Pune University), Pune 411044, Maharashtra, India*
***Corresponding author:** topannavarp@gmail.com

Received: February 12, 2023 **Accepted:** May 26, 2023

Abstract. The characteristic features of brain *magnetic resonance image* (MRI) scans of *Alzheimer's disease* (AD) patients and *normal controls* (NC) are difficult to differentiate because of very similar brain patterns and image intensities at the initial stages. Thus, it is difficult to detect AD at earlier stages. The earlier detection can lead to earlier medication for the patients to avoid further and permanent damage to their brains. Thus, computer-aided systems for early Alzheimer's disease detection are needed. For this purpose, the lossless feature extraction method combined with feature reduction using a selection approach is one of the best possible solutions presented in this paper. The dataset used for experimentation for AD detection is a combination of MRI images with very mild and mild cognitive impairments. The *2D dual-tree complex wavelet transform* (DTCWT) is used for feature extraction. A *recurrent neural network* (RNN) architecture is used for classification purposes. The *long short-term memory* (LSTM) and *gated recurrent unit* (GRU) are used in combination in the proposed architecture. The performance evaluation of combinations of LSTM and GRU and individual layers is performed, in which LSTM sandwiched between two GRU in the proposed model shows better performance.

Keywords. Alzheimer's disease, Dementia, RNN, LSTM, GRU, Performance parameters

Mathematics Subject Classification (2020). 68T07

1. Introduction

Alzheimer's disease (AD) is a very common neurological disorder observed in older people. In brain-related diseases, structure and MRI-based features indicate variations in the gray and white matter tissue. Hippocampal and entorhinal cortical decay are also observed in AD patients. It is difficult for clinicians to detect the smaller changes just by visual assessment. Also, there is a considerable delay between the onset of AD and its diagnosis. Thus, the early detection of AD is very difficult, and there is a need for intelligent means to support clinicians in the early and accurate diagnosis of the disease.

For Alzheimer's disease detection using the MRI imaging technique, a variety of classical methods are available. Some researchers are also focusing on the use of wavelet decomposition methods, in which actual image size reduction is performed by extracting unique characteristic features. On the other hand, *discrete wavelet transform* (DWT) still has a skipping property while extracting features compared to *continuous wavelet transform* (CWT). Also, applying CWT is the right choice for extracting lossless feature coefficients. The combination of discrete and CWT transforms is the best approach for keeping complexity low and extracting lossless features. The 2D DTCWT for feature extraction in medical image processing is explained by Vetova and Ivanov in [16]. The method is modified to get extra angular features in terms of spatial characteristic coefficients. The features obtained are then used for training the classifier in supervised learning and tested against the test dataset, which shows better performance. The *recurrent neural network* (RNN)-based classifier considered in this work is composed of LSTM and GRU. The performance of machine learning algorithms and deep learning algorithms as classifiers is shown in the results and analysis section.

2. Literature Survey

Er and Goularas [5] provided a study of a computer-aided system with a deep learning method for the identification of mild cognitive impairment due to AD. The method identifies different brain regions with the use of a 3D Jacobian-based method. The difference between two consecutive brains is easily maintained using this strategy. The method shows an accuracy of 87.2%.

Lian *et al*. [9] identified limitations in various methods involving neural network-based prediction of dementia from sMRI images. The unique anatomical characteristics of each brain are ignored while identifying common features. Also, patch-based training focuses on limited regions of interest, and hence global structural information about the brain is neglected. The authors provided use of the attention layer in a CNN-based approach for improving performance in the case of multiple combinations of the datasets. Du *et al*. [4] proposed an analysis method using multitask sparse canonical correlation from multimodal *quantitative trait* (QT) brain images. The parameter decomposition and multi-task learning advantages are the main outcomes of using the method. The genetic image analysis with multi view using canonical correlation coefficients and weights is the main advantage of the method. Chan *et al*. [2] have shown the use of the wavelet method for identifying the brain connectivity parts. The identification of Alzheimer's disease is performed using this method. Motion artifact removal is the main target of fNIRS data. The signal processing approach is used for Alzheimer's disease detection. The authors considered topological constraints in filtering brain networks. Even with the sparsity present in the network, it has shown the highest efficiency.

Dadar *et al*. [3] presented a fully automatic segmentation technique. The application considered is a brain MRI image in which aging effects and AD are to be detected. The white matter hyperintensities are segmented using the linear segmentation method. The automatic segmentation shows almost 93% accuracy.

Mahanand *et al*. [10] presented a method for AD detection with an *integer-coded genetic algorithm along with the Extreme Learning Machine classifier* (ICGA-ELM classifier). The morphometric analysis is used with a voxel-based approach, in which the respective best features are selected. The classifier shows 91.8% accuracy with 10 features compared to 86.84% accuracy using a support vector machine classifier for AD detection.

Herrera *et al*. [6] have discussed two problems in AD detection using MRI images. The first one focuses on the identification of AD and normal images from a dataset of 1000 images. The second problem focuses on the identification of mild cognitive impairment patients, AD, and normal individuals from MRI image datasets. The authors' solution to the second issue makes the early diagnosis of dementia possible. In the feature extraction stage, a 2D discrete wavelet transform is used, and then principal component analysis is applied to further reduce the features. The *support vector machine* (SVM)-based linear kernel classifier is used to classify the features. The comparative use of two methods of wavelet decomposition, Daubechies and Haar wavelet, is done. In both methods, performance is evaluated using PCA and without PCA, in which Haar wavelet shows the highest performance of 96.23% accuracy, and applying PCA improves the processing speed of the classifier but reduces accuracy to 94.79%.

Sweety and Jiji [14] used optimisation techniques to select the features from a set of features obtained from an MRI image. The preprocessing with noise removal using the Markove random field method is done. The features are obtained using Eigen vectors, Eigen brain, mean, variance, skewness, kurtosis, standard deviation, area, perimeter, and eccentricity. The *particle swarm optimisation* (PSO) is used to select and reduce the feature vector size, and the *decision tree* (DT) classifier is used to classify the images.

Vetova and Ivanov [16] have shown the use of DTCWT for feature extraction from MRI images. The results are evaluated using a multilevel decomposition of DTCWT. In which vector length, time, and wavelet coefficients are considered. The level 4 decomposition even maintains the uniqueness of the features from MRI images when AD disease is concerned.

Udomhunsakul and Wongsita [15] have shown spatial feature extraction and its application to MRI images. The 5×5-sized filters with 40 sets are used to extract the features, which maintain the scale and edge parameters as characteristic features. The preprocessing of the image is done using Gaussian 3×3 filters, and the impact of noise removal is observed while using Gabor filter banks. Oleiwi [12] used a texture-based segmentation method using k-means. The segmented image is further processed using a *gray-level co-occurrence matrix* (GLCM) to extract the features. The features are then classified in AD and normal sets using the K-nearest neighbour classifier with 86.6% accuracy on the OASIS database.

Alattas and Barkana [1] used simple and basic image processing methods such as thresholding, edge detection, and morphological operations to get brain dimensionality. A comparative study of brain size is performed for AD patients' brain MRI images and normal brain MRI images, in which a dimensional difference study is shown.

3. Proposed Work

The stage of Alzheimer's prediction consists of processing stages as shown in the block diagram in Figure 1.

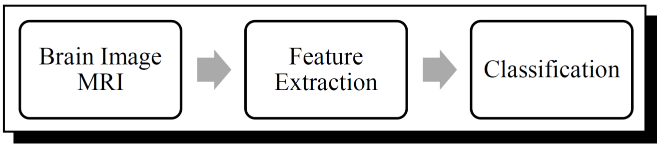

Figure 1. Block diagram of processing for Alzheimer's detection method

The important feature extraction stage should gather all important features that highlight the characteristic features of Alzheimer's disease. In T2 MRI images, the maximum details are seen compared to T1 and PET images. The Kaggle ([1]), the Alzheimer's dataset consists of four stages of Alzheimer's disease with standard images collected from different websites to test the accuracy of the model. The dataset is preferably used to evaluate the performance of the proposed method.

In earlier conventional tools, the Fourier transform was popular, but due to the time and frequency feature extraction capabilities of wavelet transforms, they have shown their impact in machine learning applications. The *2D discrete wavelet transform* (2D-DWT) decomposes the image into four bands, which reduces the complexity of processing by using each band separately. The *complex wavelet transform* (CWT) is a complex-valued extension to the standard *discrete wavelet transform* (DWT). The greater the characteristic features in terms of multi-resolution and sparse representation of CWT, the more unique features can be obtained. The higher degree of shift-invariance in its magnitude is the main reason for obtaining important distinctive features in medical images. A drawback of simple CWT is that it exhibits 2D redundancy in features as compared to separable DWT. On the other hand, for image feature extraction, Kingsbury showed the use of dual tree CWT (DTCWT) in 1998. In DTCWT, two real DWTs are performed while obtaining the subband features, and there is also an imaginary part in the second DWT operation. The features extracted by using the coefficients extraction method should meet the following conditions:

- The decomposition should be perfectly reversible to get the original image back, which shows lossless decomposition and important features even for small regions of interest after decomposition.
- The decomposed parts of the input vector should be a half-sample shift from each other.
- The boundary conditions in $h_0(1)$ and $g_0(1)$ should be satisfied by only one sample shift.

To meet these requirements for 1D signal vectors, Kingsbury proposed orthogonal Q-shift filters. The idea can be extended for 2D by using equations (3.1) and (3.2), which describe complex separable wavelets and scaling, respectively, with the implementation of separable filters on the first columns and then on the rows in 2D DTCWT:

$$\psi_1(x,y) = \varphi(x)(y), \ \psi_2(x,y) = \psi(x)\varphi(y), \ \ \psi_3(x,y) = \psi(x)\psi(y), \tag{3.1}$$

$$\varphi(x,y) = \varphi(x)\varphi(y), \tag{3.2}$$

[1]Figshare, *Brain Tumor Dataset*, URL: https://figshare.com/articles/dataset/brain_tumor_dataset/1512427, First online date: 2017-04-03.

where $\psi(\cdot)$ and $\varphi(\cdot)$ are complex wavelet and complex scaling functions respectively. The proposed method combines four trees on columns and four trees on rows to obtain twelve shift variant wavelets oriented at $\pm 15°$, $\pm 30°$, $\pm 45°$, $\pm 60°$, $\pm 75°$, $\pm 90°$.

Two different sets of filter are used in two wavelet transforms to meet 1st condition mentioned earlier. Consider $h_0(n)$, $h_1(n)$ and $g_0(n)$, $g_1(n)$ represent low pass and high pass pair of filters for upper and lower separable filter banks respectively. Hence complex wavelet can be denoted as,

$$C(t) = h'(t) + jg'(t), \tag{3.3}$$

where $h'(t)$ and $g'(t)$ are real complex wavelet transforms of $h(t)$ and $g(t)$ respectively. Also,

$$g'(t) = \text{Hilbert}\{h'(t)\}. \tag{3.4}$$

Equation (3.4) denotes the relation such that exact Hilbert transform of $h'(t)$ is $g'(t)$ for loss less decomposition method. Another advantage of using DTCWT is that there is no involvement of complex arithmetic during its implementation and also, output data rate is twice to input data rate. In 2D DTCWT the coefficient extraction can be demonstrated as shown in Figure 2. Respective 2D DTCWT is applied for the decomposition of input brain MRI image the resulting four level decomposition is shown in Figure 3.

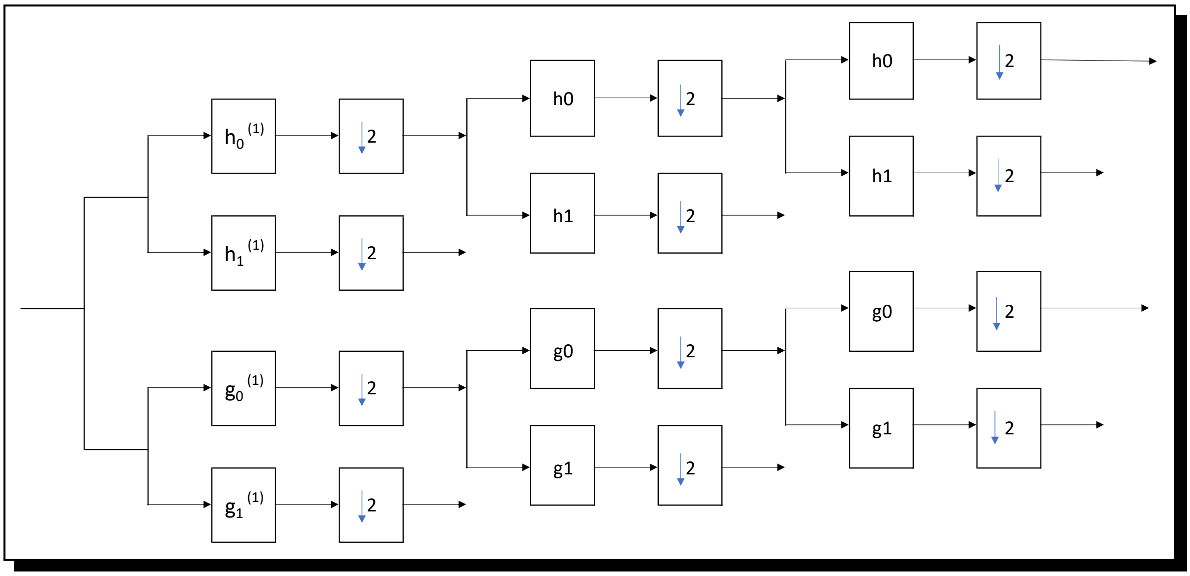

Figure 2. DTCWT decomposition

3.1 Classifiers

Machine Learning Method: The machine learning traditional classifiers are considered for evaluating the performance of the proposed system for Alzheimer detection. The classifiers *support vector machine* (SVM), *Neive Bays* (NB), *Decision Tree* (DT) and KNN are considered for comparative study. These classifiers are used in one versus all mode for multiclass classification.

RNN based Alzheimer's disease detection architecture:

In the suggested architecture, the RNN based deep learning method with *long short term memory* (LSTM) and *graded rated unit* (GRU) layers are used. For RNN based experimentation

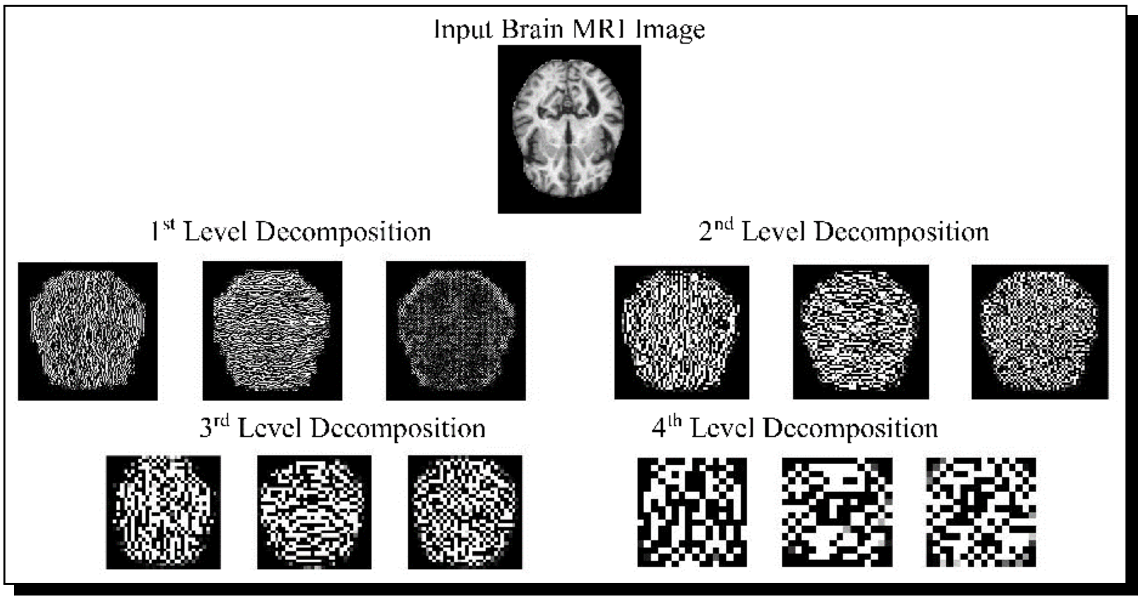

Figure 3. 2D DTCWT four level decomposition

for first model, three LSTM layers for total five layers are considered, as shown in MODEL1 (Figure 4).

For MODEL2 as shown in figure, the GRU layers is chosen for experimentation. Sandwiching of GRU with two LSTM is considered in MODEL 3 and two GRU with one LSTM layer is considered in MODEL 4. The dense layer, which is the last layer of the model is responsible for giving the binary classification output for each type of model shown in Figure 4. MODEL4 is Alzheimer's Detection Model that has been proposed.

Model architectures 1, 2, 3, and 4 are tested for algorithms 1 and 2. Once with data clustering and once without data clustering. The total number of features obtained is dependent on the dimensions of the input vector. The size of the input vector is chosen at 45, and it can be mapped into 300 learnable features during the design of the network. Hence, respective settings in basic layers in sequential models are considered for accurate settings of these values. This gives error-free input to the output interface of each layer during the design of the model. The further fitting status of the model depends on the dataset size and number of iterations in which care has been taken to avoid under- and over-fitting.

The significant difference between GRU and LSTM is that GRU's bag has two gates that are reset and update, while LSTM has three gates that are input, output, and forget. GRU is less complex than LSTM because it has fewer gates. If the dataset is small, then GRU is preferred; otherwise, LSTM is preferred for the larger dataset. GRU exposes the complete memory and hidden layers, but LSTM doesn't. The individual advantages result in improved performance in the combined model, as provided. The work presented in this paper shows four different architectures in which LSTM alone and GRU alone are also considered, and combinations are also considered in which GRU acts as a full memory network at the input and output of the LSTM network. In a 3-layer combination architecture, the performance of model 4 is better due to less forgetting even for large datasets, and at the same time, it is less complex compared to all three LSTM layers. The results obtained with 1 layer combined with 2 layers are very poor. But increasing the number of layers shows more tuning, resulting in improved performance with increased complexity. This overall strategy is considered for experimentation for finalising the architecture with a combination of GRU and LSTM in a three-layered architecture. Also, combinations of layers may lead to the problem of overfitting the model, which can be assessed with experimentation by considering a divided training and validation dataset.

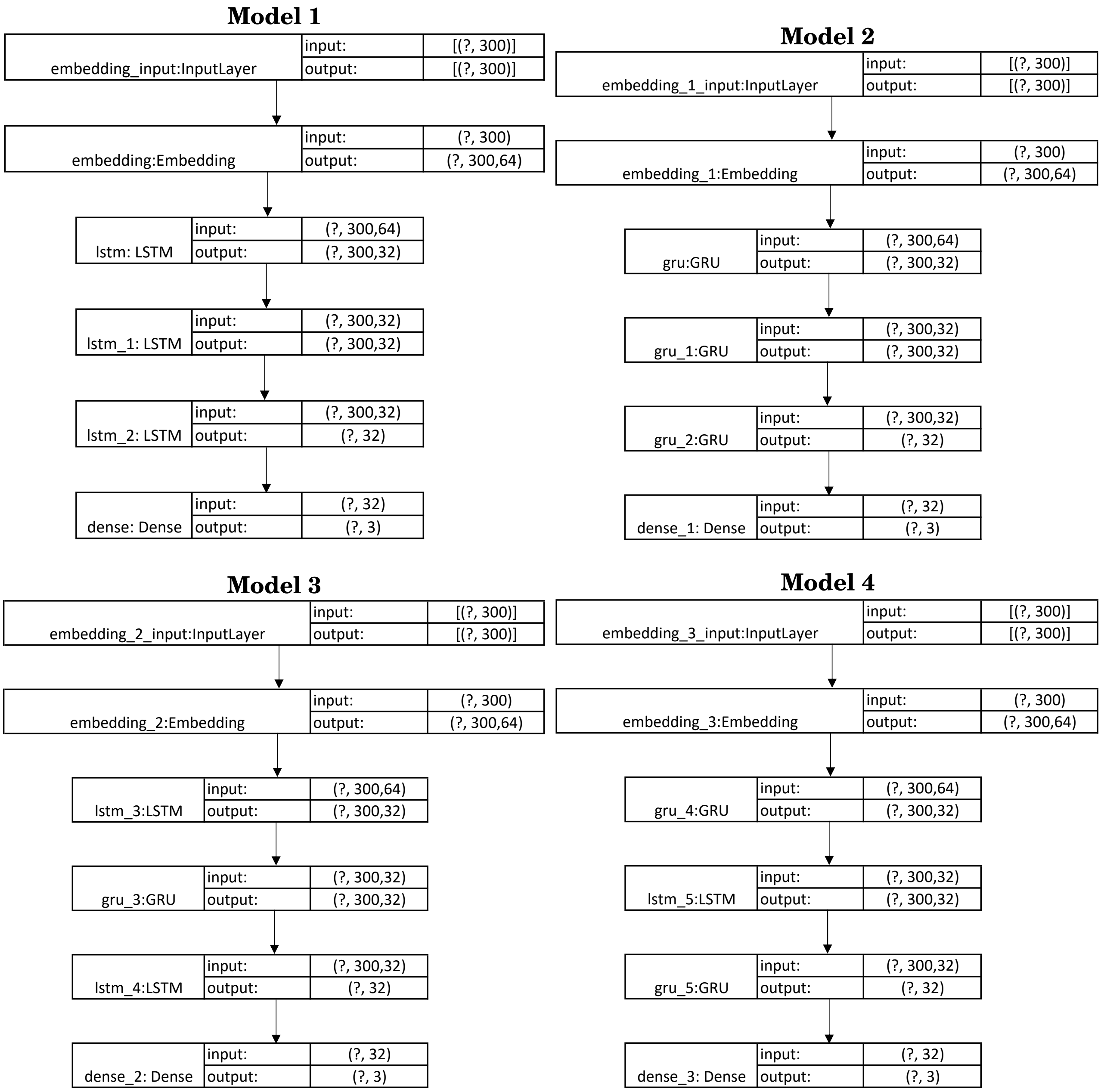

Figure 4. RNN based Deep learning models (Model 1, 2, 3, 4)

4. Results and Analysis

In the dataset, there are three subsets of Alzheimer disease images with variation in stages from very mild to moderate, and hence each case has a different approach for considering confusion matrix parameters. The one against all for each Alzheimer class is considered separately, and average performance is calculated to obtain the parameters accuracy, sensitivity, and specificity.

The loss rate analysis during training of RNN models is done as follows: The training of the model is carried out using a training dataset over 100 epochs with 10 steps per epoch. Thus, the total output of 1000 epochs is observed in multiple experiments, and graphical analysis of results is done using average results. During the training steps, the loss rate is monitored, and the loss analysis graph is plotted as shown in Figure 5.

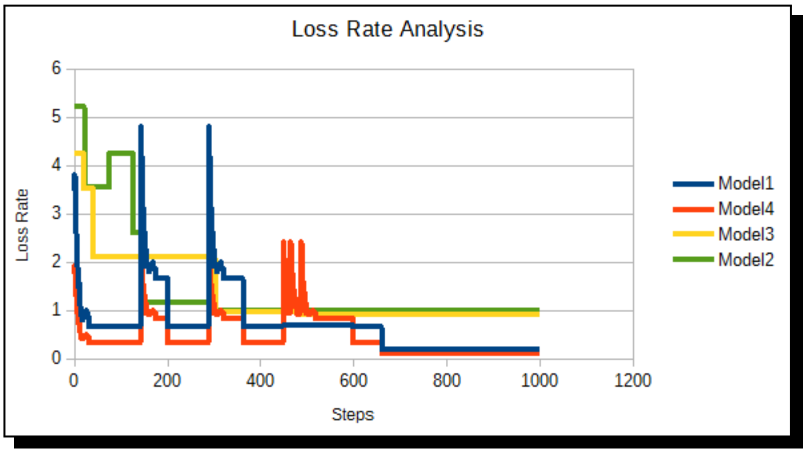

Figure 5. Loss rate analysis during training of DL models

After around 700 iterations out of a total of 1000, it is observed that models 1 and 4 have the lowest loss rates, and thus total convergence is achieved when compared to models 2 and 3, which have a loss rate of less than 0.9 in all iterations. Also, all four models achieve a minimum loss after iteration 700. The earlier stability of loss at the minimum value contributes to the training complexity and fit of the model. It is also observed that, even after 700 iterations, the model is not getting overfitted, and hence the total trainable parameters are at the best choice level while designing the model. Model 4 is more stable compared to models 1, 2, and 3 during training loss analysis. This is considered an Alzheimer's disease detection model, and further analysis is performed on the testing dataset.

Table 1. Evaluation parameters for performance analysis

Parameter	Formula
Sensitivity	$TPR = \frac{TP}{TP+FN}$
Specificity	$TNR = \frac{TN}{TN+FP}$
Accuracy	$ACC = \frac{TP+TN}{TP+TN+FP+FN}$

In Table 2, TP means true positive, in which the input given is Alzheimer and the output is Alzheimer. TN means true negative, in which the input is Alzheimer's and the output is normal. FP stands for false positive, in which the output is normal and the input is Alzheimer's, and FN means false negative, in which the input and output are both normal.

Table 2. Performance comparison of deep neural network models

	Accuracy	Specificity	Sensitivity
Model 1 (3 LSTM layers)	0.842	0.831	0.810
Model 2 (3 GRU layers)	0.850	0.857	0.855
Model 3 (Sandwiching of LSTM and GRU)	0.912	0.900	0.820
Model 4 (Sandwiching of LSTM with two GRU layers)	0.985	0.910	0.905

Figure 6 shows the plot of the comparative analysis of performance parameters: accuracy, specificity, and sensitivity of learning classifiers. The various algorithms are considered for classification and comparison, such as SVM, K-nearest neighbour, *Naive Bays* (NB), and *decision tree* (DT). The above-mentioned evaluation parameters were obtained after experimentation. It can be observed that, the results of the SVM algorithm are superior in terms of accuracy. The reason for obtaining better results with SVM is the improved convergence of the SVM model for different stages of Alzheimer's disease, i.e., from mild to moderate.

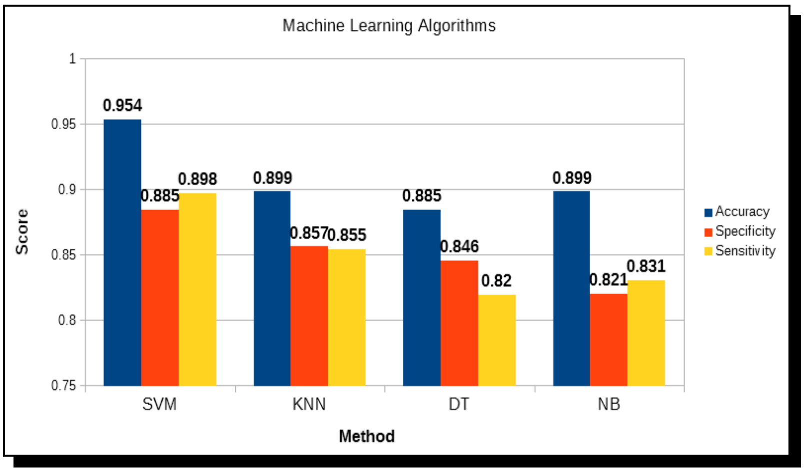

Figure 6. Performance of machine learning methods

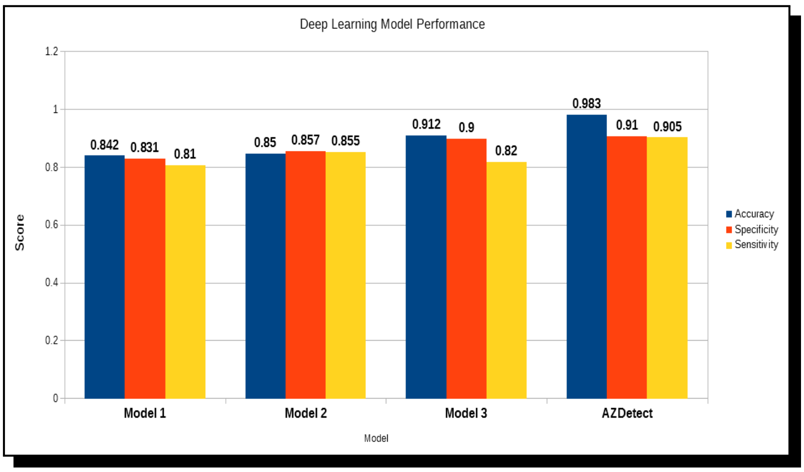

Figure 7. Performance of deep neural network models

Figure 7 shows the plot of performance parameters for various deep learning models. Figures 6a and 6b show the comparison of performance parameters for machine learning methods and deep learning models. The performance of deep learning models is admirably better. Also, enhanced results are observed after applying clustering.

5. Conclusion

This paper contributes in terms of RNN-based Alzheimer's disease detection along with features extracted using the 2D DTCWT method. The results of classification using machine learning algorithms and RNN are compared. The Alzheimer's disease detection model composed of one LSTM layer sandwitched between two GRU layers shows better performance for the dataset, which contains very mild, mild, and moderate Alzheimer's. The performance of 98.5% accuracy is seen as satisfactory over the performance of SVM at 96%.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

All the authors contributed significantly in writing this article. The authors read and approved the final manuscript.

References

[1] R. Alattas and B. D. Barkana, A comparative study of brain volume changes in Alzheimer's disease using MRI scans, in: *Proceedings of the 2015 Long Island Systems, Applications and Technology*, Farmingdale, NY, USA, 2015, pp. 1 – 6, (2015), DOI: 10.1109/LISAT.2015.7160197.

[2] Y. L. Chan, W. C. Ung, L. G. Lim, C.-K. Lu, M. Kiguchi and T. B. Tang, *Automated thresholding method for fNIRS-based functional connectivity analysis: Validation with a case study on Alzheimer's disease*, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **28**(8) (2020), 1691 – 1701, DOI: 10.1109/TNSRE.2020.3007589.

[3] M. Dadar, T. A. Pascoal, S. Manitsirikul, K. Misquitta, V. S. Fonov, M. C. Tartaglia, J. Breitner, P. Rosa-Neto, O. T. Carmichael and C. Decarli, Validation of a regression technique for segmentation of white matter hyperintensities in Alzheimer's disease, *IEEE Transactions on Medical Imaging* **36**(8) (2017), 1758 – 1768, DOI: 10.1109/TMI.2017.2693978.

[4] L. Du, F. Liu, K. Liu, X. Yao, S. L. Risacher, J. Han, A. J. Saykin and L. Shen, Associating multi-modal brain imaging phenotypes and genetic risk factors via a dirty multi-task learning method, *IEEE Transactions on Medical Imaging* **39**(1) (2020), 3416 – 3428, DOI: 10.1109/TMI.2020.2995510.

[5] F. Er and D. Goularas, Predicting the prognosis of MCI patients using longitudinal MRI data, *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **18**(3) (2021), 1164 – 1173, DOI: 10.1109/TCBB.2020.3017872.

[6] L. J. Herrera, I. Rojas, H. Pomares, A. Guillén, O. Valenzuela and O. Baños, Classification of MRI images for Alzheimer's disease detection, in: *Proceedings of the 2013 International Conference on Social Computing*, Alexandria, VA, USA, 2013, pp. 846 – 851, (2013), DOI: 10.1109/SocialCom.2013.127.

[7] B. S. Khehra and A. P. S. Pharwaha, Comparison of genetic algorithm, particle swarm optimization and biogeography-based optimization for feature selection to classify clusters of microcalcifications, *Journal of The Institution of Engineers (India): Series B* **98**(2) (2017), 189 – 202, DOI: 10.1007/s40031-016-0226-8.

[8] A. Kulkarni and N. Kulkarni, Fuzzy neural network for pattern classification, *Procedia Computer Science* **167** (2020), 2606 – 2616, DOI: 10.1016/J.PROCS.2020.03.321.

[9] C. Lian, M. Liu, Y. Pan and D. Shen, Attention-guided hybrid network for dementia diagnosis with structural MR images, *IEEE Transactions on Cybernetics* **52**(4) (2022), 1992 – 2003, DOI: 10.1109/TCYB.2020.3005859.

[10] B. S. Mahanand, S. Suresh, N. Sundararajan and M. A. Kumar, ICGA-ELM classifier for Alzheimer's disease detection, in: *Proceedings of 2013 Indian Conference on Medical Informatics and Telemedicine* (ICMIT13), Kharagpur, India, pp. 48 – 52, (2013), DOI: 10.1109/IndianCMIT.2013.6529407.

[11] A. V. Nandedkar and P. K. Biswas, A fuzzy min-max neural network classifier with compensatory neuron architecture, *IEEE Transactions on Neural Networks* **18**(1) (2007), 42 – 54, DOI: 10.1109/TNN.2006.882811.

[12] W. K. Oleiwi, Alzheimer disease diagnosis using the K-means, GLCM and K_NN, *Journal of University of Babylon for Pure and Applied Sciences* **26**(2) (2017), 57 – 65, DOI: 10.29196/jub.v26i2.474.

[13] P. K. Simpson, Fuzzy min-max neural networks. I. Classification, *IEEE Transactions on Neural Networks* **3**(5) (1992), 776 – 786, DOI: 10.1109/72.159066.

[14] M. E. Sweety and G. W. Jiji, Detection of Alzheimer disease in brain images using PSO and decision tree approach, in: *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, Ramanathapuram, India, 2014, pp. 1305 – 1309, (2014), DOI: 10.1109/ICACCCT.2014.7019310.

[15] S. Udomhunsakul and P. Wongsita, Feature extraction in medical ultrasonic image, in: *Proceedings of the 3rd Kuala Lumpur International Conference on Biomedical Engineering 2006* (IFMBE'06), F. Ibrahim, N. A. A. Osman, J. Usman and N. A. Kadri (editors), Vol. **15**, Springer, Berlin — Heidelberg (2007), DOI: 10.1007/978-3-540-68017-8_69.

[16] S. Vetova and I. Ivanov, Image features extraction using the dual-tree complex wavelet transform, in: *Advances in Applied and Pure Mathematics*, J. Balicki (editor), pp. 277 – 282, WSEAS Press, (2014), URL: http://www.wseas.us/e-library/conferences/2014/Gdansk/MATH/MATH-37.pdf.

[17] H. Zhang, J. Liu, D. Ma and Z. Wang, Data-core-based fuzzy min-max neural network for pattern classification, *IEEE Transactions on Neural Networks* **22**(2) (2011), 2339 – 2352, DOI: 10.1109/TNN.2011.2175748.