Alzheimer's Classification Using Transfer Learning on MRI Images

1st Kirtanaa Anandakumaran
Computer Science Engineering
Vellore Institute of Technology
Chennai, India
kirtanaa.anandakumaran2021@vitstudent.ac.in

2nd Asam Ignatious Abhinaya Reddy
Computer Science Engineering
Vellore Institute of Technology
Chennai, India
asamignatious.abhinaya2021@vitstudent.ac.in

***Abstract*—Alzheimer's disease is a progressive neurodegenerative disorder that leads to cognitive and behavioral decline, ultimately progressing to fatality in its advanced stages. Accurate staging of Alzheimer's is essential for timely and appropriate intervention. In this study, we perform and analyse the results of multiclass classification of Alzheimer's stages using MRI scans of the brain. Our methodology employs a transfer learning approach with VGG16 as the base model to distinguish among various stages of the disease, leveraging MRI imaging data. This approach aims to enhance clinical decision-making by offering insights into the progression and stage-specific characteristics of Alzheimer's disease.**

***Index Terms*—image processing, transfer learning, Alzheimer's disease**

I. Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder and the most common cause of dementia globally. It is characterized by cognitive decline, memory impairment, and behavioral disturbances that worsen over time, ultimately leading to fatality in its advanced stages. Timely diagnosis and accurate staging of Alzheimer's are critical for effective clinical intervention, enabling personalized patient care and improving the quality of life for both patients and caregivers. Despite significant advancements in medical imaging and diagnostic tools, the early detection and precise classification of Alzheimer's stages remain challenging due to the complex progression of the disease and subtle differences between its early stages.

In recent years, advancements in artificial intelligence (AI), particularly deep learning, have shown immense potential in the medical imaging domain. Deep learning models, specifically convolutional neural networks (CNNs), have demonstrated remarkable capabilities in learning complex patterns from medical images such as Magnetic Resonance Imaging (MRI) scans. These models enable the detection of subtle changes in brain structure and functionality that are indicative of Alzheimer's progression. However, while deep learning models achieve high accuracy, their lack of interpretability often limits their adoption in clinical settings, where explainability and trust are paramount for decision-making. Transfer learning emerges as a noteworthy alternative in performing image classification. Transfer learning is the process of applying pre-trained deep learning models on specific image classification tasks. One such model is the VGG16 model (Visual Geometry Group) which is renowned for its exceptional ability to classify images in various different categories with high accuracy. This model is a type of convolutional neural network that contains 16 layers - 13 convolutional layers and 3 fully connected layers. It is pre-trained on the ImageNet dataset and we utilize this pre-trained model for our analysis.

This study addresses the critical need for accurate and interpretable classification of Alzheimer's disease stages using MRI scans. Leveraging the Augmented Alzheimer MRI Dataset, which includes four distinct stages of the disease—non-demented, very mild demented, mild demented, and moderate demented—we employ deep learning-based multiclass classification to distinguish between these stages.

The objectives of this research are to process and analyze MRI images for Alzheimer's detection and to evaluate the performance of the transfer learning in classifying Alzheimer's stages, thereby bridging the gap between computational methods and clinical practice. By advancing the understanding of Alzheimer's progression and providing interpretable predictions, this work aims to support clinicians in making informed decisions and contribute to the broader field of this disease research.

II. Related Work

In recent years, techniques employing deep learning to diagnose AD have gained prominence [8,9,10,11,12,13,14,15,16]. Deep learning is a type of machine learning that is particularly well suited for the analysis of complex medical images, as it has the ability to automatically learn and extract features from large datasets. In addition to medical applications, deep learning is used in other applications [20].

For example, Darehnaei et al. [1] presented an approach for multiple vehicle detection in UAV images using swarm intelligence ensemble deep transfer learning (SI-EDTL). The presented method has the potential to enhance the effectiveness of various applications, such as surveillance and disaster response. A number of studies have explored the use of deep learning for AD detection using various imaging modalities, including structural MRI, functional MRI, PET, and amyloid imaging. However, in this paper, we focused on the methods that used the same dataset of MRI images. These studies

have demonstrated the potential of deep learning to accurately classify the AD disease. Several different approaches have been used to develop deep learning models for AD diagnosis using MRI images.

[2] presented an approach for the classification of AD using deep neural networks and MRI. The approach involves preprocessing the MRI scans, extracting features from segmented brain images using a combination of 2D and 3D CNNs, and classifying the scans into AD and non-AD using a fully connected neural network. The authors achieved promising results with an accuracy of 95.34%, a sensitivity of 96% and a specificity of 94.67%. The presented method has the potential to significantly improve the early detection and treatment of AD. However, further validation on larger and more diverse datasets is necessary to assess its generalizability and robustness. Another hybrid model based on deep learning and traditional classifiers was presented by Mohammed et al.

Menagadevi et al. [8] developed a computer-aided diagnosis system for detecting AD based on a combination of a deep learning model with traditional classification methods. They first start with preprocessing stages on the input MRI images to enhance the images. After that, they perform segmentation on the preprocessed images to obtain the region of interest. Then, they extract the features using the presented multiscale pooling residual autoencoder model. Finally, they used separate classifiers such as K-Nearest Neighbor (KNN) and Extreme Learning Machine (ELM) for final classification. They obtained an overall accuracy of 96.88% using the KNN classifier and 98.97% using the ELM classifier for the binary classification task. However, the study focused on only one imaging modality, MRI, and used relatively small datasets. Murugan et al.

[9] introduced a deep learning modality called "DEMNET" for diagnosing AD from MRI images. They used several image processing techniques, such as preprocessing, oversampling, and splitting the input data. After that, they fed the split data to the presented deep model for feature extraction and classification. They obtained an overall accuracy of 95.23% for the multi-classification task. However, similar to Menagadevi et al. [8], the study only focused on MRI imaging and used small datasets. Loddo et al.

[10] presented a fully automatic model based on ensemble deep learning approaches for diagnosing AD from MRI images. They employed three pretrained deep models: AlexNet, ResNet 101, and InceptionResNetV2. After that, they used an average strategy to generate the ensemble output. They obtained the best accuracy of 96.57% for the binary classification task and an accuracy of 97.7% for the multi-classification task. However, the study did not use any image preprocessing techniques, and the ensemble approach may not always improve the performance of deep learning models. Sharma et al.

[11] presented a hybrid modality called "HTLML" based on AI approaches for the detection of AD from MRI images. They perform the first preprocessing stage on the input MRI images. After that, they fed these preprocessed images in parallel into two pretrained models, such as DenseNet201 and DenseNet121. Then, they perform classification using separate classifiers for each pretrained model. Finally, they combine the output for each classifier using the voting strategy to obtain the final decision. They obtained an overall accuracy of 91.75% for the multi-classification task. However, the study did not employ any data augmentation techniques and used relatively small datasets. Hazarika et al.

[12] for the early diagnosis of AD from MRI images. The authors first enhanced the input MRI images using several preprocessing techniques. After that, they fed the preprocessed images to the presented deep model, which is a convolutional neural network (CNN) model for extracting the features. Finally, these features are fed to a separate classifier, such as a support vector machine (SVM), for final classification. They worked on a multi-classification task and obtained an overall accuracy of 94.80%. However, the use of a traditional classifier may limit the performance of the model.

Balasundaram et al. [13] used the ResNet50 pretrained model for the diagnosis of AD from MRI images. They used preprocessing techniques such as resizing and thresholding on the input images. After that, they fed these images to the presented pretrained model for final classification. They obtained an overall accuracy of 94.1% on the multi-classification task. However, the study did not employ any data augmentation techniques, and the use of a single pretrained model may limit the performance of the model. Bangyal et al.

[14] applied deep learning techniques to MRI images to detect AD. A comparative analysis between them proves that deep learning approaches can detect AD better than traditional machine learning approaches. They finally obtained an overall accuracy of 94.63% using deep learning approaches on a multi-classification task from MRI images. However, the study did not employ any data augmentation techniques and used relatively small datasets. Ahmed et al.

[15] presented a classification method called "DAD-Net" using an optimized neural network for the early diagnosis of AD. They split the data and performed preprocessing techniques on the input MRI images. After that, they fed these images to the presented deep classification model for extracting features and final classification. They obtained an overall accuracy of 90% for the multi-classification task.

Tuvshinjargal and Hwang [16] presented a combination model between the VGG-C transform and CNN for the prediction of AD from MRI images. They use Z-score scaling to preprocess the input images and quantize pixel intensity. After that, they fed these images to the VGG pretrained model for final prediction. They obtained an overall testing accuracy of 77.46 when working on a multi-classification task. However, the study used a relatively simple deep learning model and achieved lower performance compared to the other studies. Balaji et al.

[17] presented a hybridized deep learning approach for detecting AD using MRI images. The authors combine a CNN and a long short-term memory (LSTM) network to learn spatial and temporal features from MRI scans. The authors report an accuracy of 98.50% in classifying MRI scans into

AD or normal cases using the presented hybridized deep learning approach. However, the study requires a large amount of data to learn complex features and patterns accurately. In addition, the use of this combination can be computationally expensive, which may limit the scalability of the model. Hu et al.

[18] introduced a deep learning model for a short-term longitudinal study of MCI using brain structural MRI (sMRI) as the main biomarker. The VGG-TSwinformer model combines a VGG-16-based CNN and Transformer to extract and encode features from longitudinal sMRI images, and it uses sliding-window and temporal attention mechanisms to integrate local and distant spatial features for MCI progression prediction. They obtained an accuracy of 77.20% for the binary classification task. The study still has some limitations, such as not mining 2D local features inside slices, not adopting an effective feature fusion method for axial, coronal and sagittal plane slices, and not taking full advantage of available crosssectional biomarkers.

The study [21] employs a deep learning-based approach leveraging Convolutional Neural Networks (CNNs) for the classification of Alzheimer's disease stages using MRI scans. The methodology integrates a transfer learning strategy by utilizing a pre-trained VGG16 model, with the final layers fine-tuned to adapt to the specific classification task. The paper makes several significant contributions. First, it proposes a robust pipeline for leveraging transfer learning in medical image classification tasks, demonstrating high accuracy on Alzheimer's disease detection using MRI scans. It also systematically explores the application of CNNs in multiclass classification of Alzheimer's stages, providing insights into class imbalances and performance metrics. However, the study has some limitations. The dataset used, though augmented, may not fully capture the diversity of real-world clinical data, which could limit generalizability. Additionally, the model's reliance on transfer learning from a general-purpose image classification model (VGG16) might overlook domain-specific nuances of MRI data. Lastly, the interpretability of the model's predictions is not explored, which is critical for clinical adoption.

III. Proposed Methodology

A. Data Acquisition

The data used for this research comes from the "Augmented Alzheimer MRI Dataset', which was obtained from Kaggle. It consists of MRI images of the brain taken during various stages of Alzheimer's: non-demented, very mildly demented, mildly demented and moderately demented. The dataset comprises augmented data which is used as the training data, and original data which is used for testing and validation.

B. Data Preprocessing

To ensure that the images are compatible with the VGG16 model, they are resized to 224 x 224 pixels. The image pixel values are then normalized to the range [0, 1] using

TABLE I
AUGMENTED DATA

Class	Number of Images
Non demented	9600
Very Mild Demented	8690
Mild Demented	8690
Moderate Demented	6464

TABLE II
ORIGINAL DATA

Class	Numberof Images
Non demented	3200
Very Mild Demented	2240
Mild Demented	869
Moderate Demented	64

TensorFlow's 'Rescaling' layer. Labels are converted to one-hot encoded format, facilitating categorical classification for the four disease stages. These preprocessing techniques ensure that the data is suitable for input into the CNN model. The dataset is batched with a size of 32 for efficient processing during training and evaluation. Each batch contains normalized image tensors and one-hot encoded label tensors. The dataset pipeline is optimized using 'tf.data' operations for real-time augmentation and loading during training.

C. Model Architecture

Our study employs a transfer learning approach using the VGG16 model which is pre-trained on the ImageNet dataset. The VGG16 model is a convolutional neural network architecture consisting of 16 layers. The convolutional base is frozen, except for the last four layers. The earlier layers of the model usually learn generic features while the later layers take into account task-specific features. This approach enables the model to adapt to the specific classification task while still keeping in mind the general features learned during pre-training. This model acts as the base and its architecture is extended with:

- A Global Average Pooling layer to reduce spatial dimensions, which reduces overfitting and retains the most important features.
- A Batch Normalization layer which normalizes the inputs to each layer. This reduces sensitivity to weight initialization and stabilizes learning.
- A dense layer with ReLU activation.
- A dropout layer of 30% to prevent overfitting.
- Another dense layer with ReLU activation.
- A final dense layer with a softmax activation function for multiclass classification of the four disease stages.

The model is compiled with the Adam optimizer (learning rate of 0.0001), a categorical cross-entropy loss function, and accuracy as the evaluation metric. A learning rate scheduler ('ReduceLROnPlateau') adjusts the learning rate dynamically based on the validation loss, halving it if no improvement is observed for three consecutive epochs. This helps maintain stable training and prevent stagnation. The model is trained

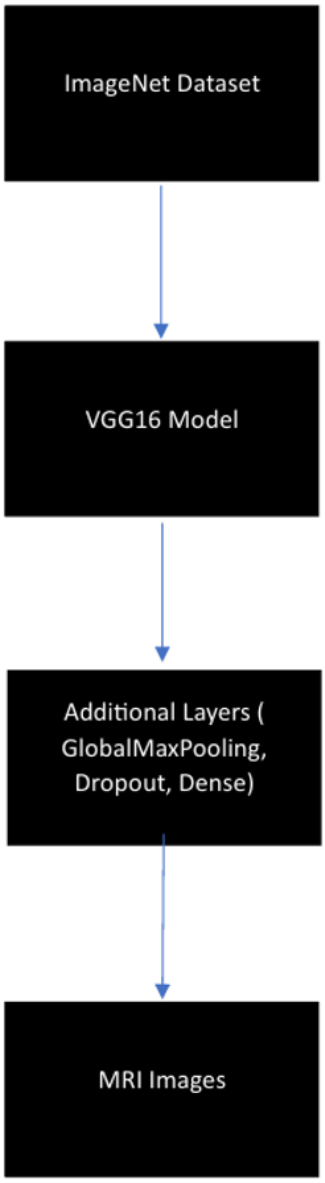

Fig. 1. Proposed Methodology

for 20 epochs, with performance monitored on the validation set. Training metrics, including accuracy and loss, are logged for analysis.

IV. Results and Discussion

A. Evaluation Metrics

To evaluate the proposed methodology, accuracy and F1 metrics have been employed. These metrics are calculated as shown in Equation 1 and 2

- F1-score: F1-score is the evaluation matrix that combines precision and recall, into a single metric by taking their harmonic mean.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (1)$$

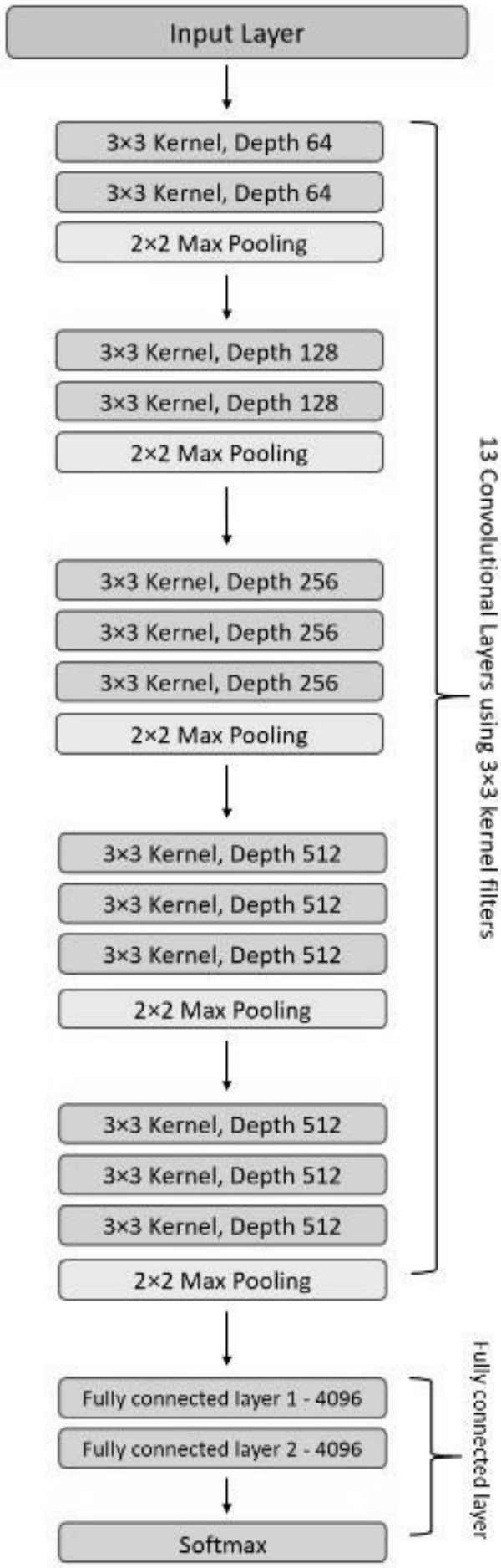

Fig. 2. Model Architecture

- Accuracy: Accuracy is a metric that measures how often a machine learning model correctly predicts the outcome.

$$Accuracy = \frac{TP + TN}{FP + FN + TP + TN} \tag{2}$$

Where:

$$TP = \text{True Positives}$$
$$TN = \text{True Negatives}$$
$$FP = \text{False Positives}$$
$$FN = \text{False Negatives}$$

B. Results

The proposed model achieves an accuracy of 99.33% and a weighted F1 score of 0.99.

TABLE III
RESULTS

Class	Precision	Recall	F1 score
Non demented	1.00	0.99	1.00
Very Mild Demented	0.96	1.00	0.98

Based on the above table, we can infer that all the predictions of the non-demented class were correct as the precision value is 1. From the recall score we can derive that almost all of the actual instances of the class were correct. For the very mild demented class, there were a few incorrect predictions, but all of true instances were correctly identified.

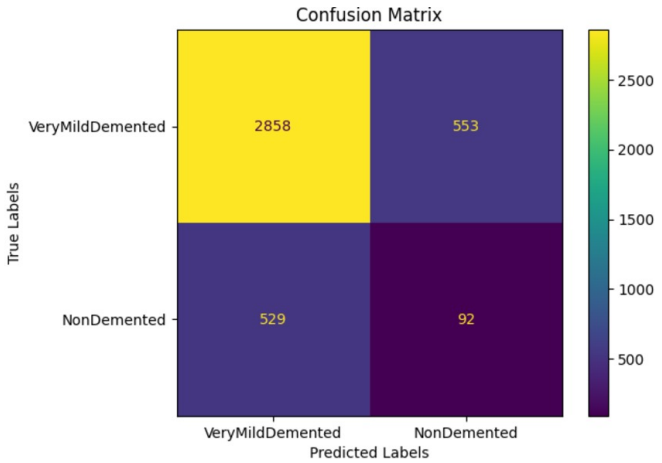

Fig. 3. Confusion Matrix

The image represents the confusion matrix of the VGG16 model in performing multiclass classification. Each diagonal element indicates the number of correctly classified instances for its respective class, while off-diagonal elements signify misclassifications. 2858 samples of the very mild demented class and 92 samples of the non-demented class were were classified correctly.

C. Discussion

Overall, the model performs exceptionally well in performing multiclass classification of the various stages of Alzheimer's, backed by the 99.3% accuracy rate. The misclassifications are owed to the imbalanced dataset leading the model to perform better for the majority class. This issue of class imbalance can be rectified using resampling or SMOTE techniques. This would make the model more robust and allow to identify and distinguish between the features of each class.

D. Future Work

Although the proposed method excels in performing multiclass classification of the various stages of Alzheimer's, there is scope for improvement in a few areas. The dataset used in this study is considerably imbalanced, with only 6000 images for the moderate demented class. This affects the ability of the model to distinguish between the features of the non demented and mild demented stages compared to other stages. Future work can focus on using SMOTE and other techniques to generate relevant images for the minority classes and help the model learn better. To enhance clinical applicability, we can also explore integrating explainable AI (XAI) techniques, specifically SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations), to provide insights into feature importance at the individual scan level. These techniques not only improve the transparency of the classification process but also aid in identifying critical regions of interest in the brain that contribute to stage-specific predictions.

V. CONCLUSION

In conclusion, our study aims to add more insight into the ongoing research about Alzheimer's disease. Perfecting classifications would revolutionize our understanding and the diagnostics of this illness. This study demonstrates the effectiveness of the VGG16 model in classifying Alzheimer's disease stages using MRI scans. By leveraging its deep hierarchical feature extraction capabilities, the model achieved high accuracy in distinguishing between various stages of the disease, highlighting its potential as a reliable tool for early diagnosis and disease progression monitoring. Furthermore, the integration of explainable AI techniques, such as SHAP and LIME, provided valuable insights into the model's decision-making process, enhancing its clinical interpretability and trustworthiness.

Despite its promising performance, the study faced certain limitations, such as computational demands and dependency on high-quality datasets. Future work should focus on optimizing the model for resource-constrained environments, expanding its generalizability across diverse populations, and incorporating multi-modal data to further improve classification accuracy. Overall, this research underscores the potential of deep learning models like VGG16 in advancing the early detection and management of Alzheimer's disease, ultimately contributing to better patient outcomes.

REFERENCES

[1] DeTure M.A., Dickson D.W. The Neuropathological Diagnosis of Alzheimer's Disease. Mol. Neurodegener. 2019;14:32. doi: 10.1186/s13024-019-0333-5.

[2] Kocaelli H., Yaltirik M., Yargic L.I., Özbas H. Alzheimer's Disease and Dental Management. Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod. 2002;93:521–524. doi: 10.1067/moe.2002.123538.
[3] Chassain C., Cladiere A., Tsoutsos C., Pereira B., Boumezbeur F., Debilly B., Marques A.-R., Thobois S., Durif F. Evaluation of Common and Rare Variants of Alzheimer's Disease-Causal Genes in Parkinson's Disease. Park. Relat. Disord. 2022;97:8–14. doi: 10.1016/j.parkreldis.2022.05.007.
[4] Kalaria R. Similarities between Alzheimer's Disease and Vascular Dementia. J. Neurol. Sci. 2002;203:29–34. doi: 10.1016/S0022-510X(02)00256-3.
[5] Sarasso E., Gardoni A., Piramide N., Volontè M.A., Canu E., Tettamanti A., Filippi M., Agosta F. A Multiparametric MRI Study of Structural Brain Damage in Dementia with Lewy Bodies: A Comparison with Alzheimer's Disease. Park. Relat. Disord. 2021;91:154–161. doi: 10.1016/j.parkreldis.2021.09.003.
[6] Simic G., Stanic G., Mladinov M., Jovanov-Milosevic N., Kostovic I., Hof P. Does Alzheimer's Disease Begin in the Brainstem? Neuropathol. Appl. Neurobiol. 2009;35:532–554. doi: 10.1111/j.1365-2990.2009.01038.x.
[7] Desai A.K., Grossberg G.T. Diagnosis and Treatment of Alzheimer's Disease. Neurology. 2005;64((Suppl. 3)):S34–S39. doi: 10.1212/WNL.64.12_suppl_3.S34.
[8] Menagadevi M., Mangai S., Madian N., Thiyagarajan D. Automated Prediction System for Alzheimer Detection Based on Deep Residual Autoencoder and Support Vector Machine. Optik. 2023;272:170212. doi: 10.1016/j.ijleo.2022.170212.
[9] Murugan S., Venkatesan C., Sumithra M.G., Gao X.Z., Elakkiya B., Akila M., Manoharan S. DEMNET: A Deep Learning Model for Early Diagnosis of Alzheimer Diseases and Dementia from MR Images. IEEE Access. 2021;9:90319–90329. doi: 10.1109/ACCESS.2021.3090474. [
[10] Loddo A., Buttau S., Di Ruberto C. Deep Learning Based Pipelines for Alzheimer's Disease Diagnosis: A Comparative Study and a Novel Deep-Ensemble Method. Comput. Biol. Med. 2022;141:105032. doi: 10.1016/j.compbiomed.2021.105032.
[11] Sharma S., Gupta S., Gupta D., Altameem A., Saudagar A.K.J., Poonia R.C., Nayak S.R. HTLML: Hybrid AI Based Model for Detection of Alzheimer's Disease. Diagnostics. 2022;12:1833. doi: 10.3390/diagnostics12081833.
[12] Mohammed B.A., Senan E.M., Rassem T.H., Makbol N.M., Alanazi A.A., Al-Mekhlafi Z.G., Ghaleb F.A. Multi-Method Analysis of Medical Records and MRI Images for Early Diagnosis of Dementia and Alzheimer's Disease Based on Deep Learning and Hybrid Methods. Electronics. 2021;10:2860. doi: 10.3390/electronics10222860.
[13] Balasundaram A., Srinivasan S., Prasad A., Malik J., Kumar A. Hippocampus Segmentation-Based Alzheimer's Disease Diagnosis and Classification of MRI Images. Arab. J. Sci. Eng. 2023:1–17. doi: 10.1007/s13369-022-07538-2. online ahead of print .
[14] Bangyal W.H., Rehman N.U., Nawaz A., Nisar K., Ibrahim A.A.A., Shakir R., Rawat D.B. Constructing Domain Ontology for Alzheimer Disease Using Deep Learning Based Approach. Electronics. 2022;11:1890. doi: 10.3390/electronics11121890.
[15] Ahmed G., Er M.J., Fareed M.M.S., Zikria S., Mahmood S., He J., Aslam M. DAD-Net: Classification of Alzheimer's Disease Using ADASYN Oversampling Technique and Optimized Neural Network. Molecules. 2022;27:7085. doi: 10.3390/molecules27207085.
[16] Tuvshinjargal B., Hwang H. VGG-C Transform Model with Batch Normalization to Predict Alzheimer's Disease through MRI Dataset. Electronics. 2022;11:2601. doi: 10.3390/electronics11162601.
[17] Varalakshmi P., Priya B.T., Rithiga B.A., Bhuvaneaswari R., Sundar R.S.J. Diagnosis of Parkinson's Disease from Hand Drawing Utilizing Hybrid Models. Park. Relat. Disord. 2022;105:24–31. doi: 10.1016/j.parkreldis.2022.10.020.
[18] Inguanzo A., Sala-Llonch R., Segura B., Erostarbe H., Abos A., Campabadal A., Uribe C., Baggio H., Compta Y., Marti M., et al. Hierarchical Cluster Analysis of Multimodal Imaging Data Identifies Brain Atrophy and Cognitive Patterns in Parkinson's Disease. Park. Relat. Disord. 2021;82:16–23. doi: 10.1016/j.parkreldis.2020.11.010.
[19] Liu C., Huang F., Qiu A., Alzheimer's Disease Neuroimaging Initiative Monte Carlo Ensemble Neural Network for the Diagnosis of Alzheimer's Disease. Neural Netw. 2023;159:14–24. doi: 10.1016/j.neunet.2022.10.032.
[20] Lahmiri S. Integrating Convolutional Neural Networks, kNN, and Bayesian Optimization for Efficient Diagnosis of Alzheimer's Disease in Magnetic Resonance Images. Biomed. Signal Process. Control. 2023;80:104375. doi: 10.1016/j.bspc.2022.104375.
[21] Raza N, Naseer A, Tamoor M, Zafar K. Alzheimer Disease Classification through Transfer Learning Approach. Diagnostics. 2023; 13(4):801. https://doi.org/10.3390/diagnostics13040801