Journal of Bioscience And Technology
www.jbstonline.com

Alzheimer Vision AI: Variational Autoencoder-Based Alzheimer's Detection from MRI

[1]Puvvula Charishma Ramya,[2]Mrs Tanneru Venkata Lavanya,

[1]M.Tech Scholar,Dept. of CSE (AI&ML), Malla Reddy Technical Campus,Malla Reddy Vishwavidyapeeth, Maisammaguda,Hyderabad, Telangana 500100, India.
Mail id cheri.puvvula@gmail.com
[2]Assistant Professor, Dept. of CSE, Malla Reddy Technical Campus,Malla Reddy Vishwavidyapeeth, Maisammaguda,Hyderabad, Telangana 500100, India.
Mail id: vlavanyat91@gmail.com

Article Info

Received: 23-03-2026 Revised: 02-04-2026 Accepted: 10-04-2026 Published: 21-04-2026

ABSTRACT

As it progresses, Alzheimer's disease, a neurological illness, drastically diminishes one's ability to remember things, think critically, and carry out routine tasks. Timely management depends on early discovery, but traditional diagnosis is based on subjective evaluations like medical history and reports of behavior, which often result in late-stage identification. A more objective, image-based diagnosis utilizing MRI images is now within reach, thanks to recent developments in deep learning. Here, we provide a new paradigm for detecting Alzheimer's disease in its early stages by integrating Convolutional Neural Networks (CNNs) with Variational Autoencoders (VAEs) for representation learning. Using convolutional neural networks (CNNs) and VAEs, MRI images may have discriminative spatial features extracted, and compressed latent representations captured by VAEs can reveal subtle anatomical changes in the brain. To improve diagnosis accuracy, these learnt traits are then used to categorize individuals into groups representing healthy and Alzheimer's disease. To guarantee strong model training and assessment, our method makes use of a dataset that contains about 500 MRI images. The experiments show that the suggested technique has promise for accurate early-stage Alzheimer's detection, and the findings are provided in terms of sensitivity, specificity, accuracy, and confusion matrix analysis. This study demonstrates how a potent diagnostic tool may be created by combining VAEs with CNNs and VAEs, which can aid in precision medicine and better clinical decision-making.

Journal of Bioscience And Technology
www.jbstonline.com

INTRODUCTION

A notable AI application in healthcare is the use of Convolutional Neural Networks (CNNs) for Alzheimer's disease diagnosis and classification. The progressive neurological condition known as Alzheimer's disease causes memory loss and cognitive impairment in millions of people throughout the globe. The key to successful treatment and intervention early identification. When it comes to medical imaging analysis, a family of deep learning algorithms called VAE AND CNN s shines. These algorithms are experts in processing and analyzing visual data.These are the usual major elements included in an introduction to an Alzheimer's disease detection and categorization system using VAE and CNNs: Focusing on the illness's effect on families and people as well as the difficulties in detecting it early, provide a brief overview of Alzheimer's disease.Bring attention to the fact that Alzheimer's disease management relies on early identification. Improving patient outcomes and opening the door to new treatment options are all possible results of prompt diagnosis.Brain imaging (MRI, PET scans), cognitive testing, and examination of cerebrospinal fluid are some of the current methods used to diagnose Alzheimer's disease. Talk about their drawbacks, such as how much they cost, how intrusive they are, and how much they depend on subjective interpretation. Give an overview of how medical image analysis uses AI, namely VAE and CNNs. Describe how convolutional neural networks (CNNs) and visual analytics (VAE) can automatically extract useful elements from medical photos to help diagnose and classify illnesses. Outline the study's goals, which include creating a VAE AND CNN based system to use neuroimaging data for Alzheimer's disease diagnosis and categorization.

Give a brief summary of the VAE AND CNN method that has been suggested. Information on the dataset's acquisition, preprocessing processes (such as normalization and augmentation), VAE AND CNN architecture (including specifics on the number of layers and filters), and training method (including details on optimization algorithms and assessment measures) may be included. Draw attention to the aspects of the suggested VAE AND CNN method that might be useful, such as its capacity to minimize diagnostic time, increase accuracy, and scale to screening programs of a big size. The paper's outline should have the following sections: introduction, methods, findings, discussion, and conclusion. In order to improve the quality of life for individuals impacted by this terrible disease, researchers are introducing Alzheimer's identification and classification using a VAE AND CNN approach. The goal is to use artificial intelligence to advance early detection and management strategies.

LITERATURE SURVEY

We introduce an automated approach that can identify brain MRIs impacted by Alzheimer's disease (AD). Multiscale analysis (MSA) is used to extract the fractals from the MRI at six distinct sizes. A support vector machine (SVM) classifier was used to distinguish between MRI scans of healthy brain tissue and those with Alzheimer's disease using the derived fractals as features. Using the leave-one-out cross-validation approach, the classification accuracy, sensitivity, and specificity of 93 brain MRIs (51 healthy brains and 42 AD afflicted brains) were 99.18% ±

0.01. With these findings and a processing time of 5.64 seconds, the suggested method seems like it may be a useful diagnostic tool for radiologists when screening for Alzheimer's disease.

Neuronal fiber tract integrity may be evaluated in vivo using diffusion tensor imaging (DTI), which can bolster the diagnosis of Alzheimer's disease (AD). To what degree various neuroimaging approach combinations improve AD identification is an open topic in the field of study. Here, we looked at a variety of approaches to merging structural T 1-weighted MRI data with DTI data. Additionally, we used machine learning algorithms to identify AD automatically. A total of 137 patients suspected of having AD (MMSE 20.6 ±5.3) and 143 healthy older controls were included in the research. The participants were scanned using nine different scanners and were sourced from the newly established European DTI study on Dementia (EDSD) methodology. Using gray matter density (GMD) and white matter density (WMD) maps from anatomical MRI in addition to the DTI-derived indices fractional anisotropy (FA) and mean diffusivity (MD), we were able to classify patients for diagnostic purposes. Our Support Vector Machine (SVM) classifier underwent tenfold cross validation while we carried out voxel-based classification. To determine the best way to merge the modalities, we examined the outcomes of each modality separately and with those of other methods. Combining modalities did not improve AD detection rates for our sample. Both the GMD data alone and the multimodal categorization with GMD achieved an accuracy of around 89%. This high level of accuracy was maintained by all of the methods. assess the accuracy of the categorization when other modalities, such as functional MRI or FDG-PET, are used.

OBJECTIVE

From a machine learning standpoint, creating and implementing a deep learning technique that describes Alzheimer's characteristics and suggests an Alzheimer processing model utilizing VAE and VAE AND CNN Classification

PROBLEM STATEMENT

In order to achieve these objectives, it is necessary to train VAE AND CNN architectures that can accurately and efficiently extract useful features from brain imaging data while also optimizing the models for resilience, specificity, accuracy, and sensitivity. The resulting system's dependability and generalizability for real-world clinical applications depend on thorough validation and testing.

EXISTING SYSTEM

We employed the SVM algorithm to forecast Alzheimer's details in this existing technique. The SVM method classifies people with Alzheimer's disease into distinct groups based on the results of a single clustering step. It takes a long time for the SVM algorithm to categorize all the provided information. Efficiency in terms of time is that primary drawback. Information on Alzheimer's patients is derived from autonomous, large-volume sources that are both dispersed and decentralized in order to better understand the complex and ever-changing relationships among the data points.

LIMITATIONS

We have an Alzheimer's dataset, and it will take additional time to compute it. Finding patterns using older data mining methods may not be the best way to make predictions.

Their research led them to the conclusion that younger people may wait to begin medication therapy, but older patients need to start taking their medication right away.

The Alzheimer dataset was used to predict and classify different types of Alzheimer's disease using a classification algorithm.

PROPOSED SYSTEM

A "proposed method" is usually a recommended strategy or methodology for doing research or resolving an issue in scholarly and scientific settings. The term "method" refers to the set of activities that researchers plan to carry out in order to accomplish their study goals or solve a specific problem.Research articles, theses, grant submissions, and project proposals sometimes include a section outlining the proposed methodology. Its purpose is to lay out the specifics of the research process, including the study's methodology, data gathering strategies, data analysis tools, and anything else that has to be included.It is important to give serious attention to the study objectives, available resources, ethical concerns, and any constraints before deciding on a methodology. The selected method has to additionally show that it is suitable and doable to solve the issue or study inquiry.

In sum, the suggested approach is an essential part of the research strategy and will direct the study's execution.

ADVANTAGES

- More accuracy
- Time requirement is very less
- We can train more images in VAE and CNN

FUNCTIONAL REQUIREMENTS

1.Data Collection
2.Data Preprocessing
3.Training And Testing
4.Modeling
5.Predicting

One quality aspect of software systems is NON-FUNCTIONAL REQUIREMENT (NFR). Critical non-functional criteria for the software system's success include responsiveness, usability, security, and portability, among others. Based on these, they evaluate the software system. A nonfunctional requirement may be something like, "how quickly does the website load?" Systems that don't fulfill user demands are often the consequence of not meeting non-functional criteria. Through the use of non-functional requirements, you may limit or impose design constraints on the system across all of the agile backlogs. When there are more than 10,000 people on the site at the same time, for instance, it should load in three seconds. Providing a detailed description of criteria that aren't functional is as important.

SOFTWARE REQUIREMENT:

Python
Anaconda navigater
Jupyter note book

HARDWARE REQUIREMENT:

WINDOWS 10 OS PC 64 BIT
MINIMUM 4GB RAM

To measure the efficacy of the multi-classification model, we used ROC AUC and F1-score. Recognized as an unsatisfactory measure for model performance evaluation with unbalanced data, accuracy is a regularly adopted performance statistic [33]. Hence, we refrained from using this measure. When learning from data that isn't balanced, a variety of measures are used. A receiver operating characteristic (ROC) analysis and its area under the curve are the most common of these [36,37]. In learning from unbalanced data, accuracy and recall are two more often used metrics. The F1-score metric gives equal weight to memory and accuracy. G-Mean, Brier score, F2-score, and F0.5-score are some other performance measures for unbalanced classification that researchers may use in addition to F1-score and ROC AUC [38]. When the importance of false positives and false negatives is equal, the F1-score is the suitable metric to use. On the other hand, F2-score is better suited for situations where false negatives are more of a worry. When the relevance of false positives is larger, the F0.5-score is preferable. To evaluate the degree of uncertainty in classifier predictions, specific probabilistic measures have been developed, such as the Brier score. We choose to employ the F1-score and ROC AUC as our evaluation criteria in our research on stage identification, where all classes have equal relevance. In addition, when learning from unbalanced data, the F1-score version is often used as a performance statistic [33]. We used k-fold cross-validation for both clinical and genomic expression data. To test how well predictive models can generalize and to prevent overfitting, researchers use cross-validation, a data resampling technique [29,30]. A k-fold cross-validation procedure involves dividing the available training set into k separate subsets of about the same size. In this context, fold is the number of subsets that are produced. Without replacing any instances in the learning set, this partitioning is accomplished by randomly selecting them. As a whole, the training set consists of k subsets, and each subset is used to train the model. Then, the remaining subset, called the validation set, is used to apply the model and assess its performance. This process is carried out again and again until all k subsets have been used as validation sets [29]. A higher k-value would increase the number of training samples while decreasing the number of test samples. There were 744 samples in the genetic dataset and 2400 samples in the clinical trial. We used k-fold cross-validation for both gene expression and clinical data, with k=5, since the gene expression data was very modest. Twenty percent of the total samples would be kept for testing using 5-fold cross-validation. As a result, there were 596 training samples and an ideal number of 148 testing samples for gene expression data. For the purpose of comparing performance, we used k = 5 for the clinical data-based model in cross-validation, which is in line with the gene expression-based models. As mentioned earlier, the SMOTE-based minority oversampling approach was used to raise the minority sample size, which in turn increased the number of training samples. To control randomness, verify findings were reproducible, and improve settings, we originally performed several trials on both datasets by briefly altering the random state in the k-fold cross-validation procedure. Each model instance constructed from both biomarker categories produced consistent AUC and F1-score results, as seen in the following tables, since this enabled the minimizing of performance variances across varied random conditions. In Tables 6 and 7, you can see the F1-score and area under the curve (AUC) from the 5-fold cross-validation models of the RF, DL, SVM, and GBM classifiers that were trained using gene expression data. The corresponding outcome from a model that did not use the Synthetic Minority Oversampling Technique (SMOTE) is shown beside each result that did use it. In the SMOTE-based part of the chart, the first row shows the assessment of each classifier out of 23 that were built using the same oversampled training set and the matching testing data from the first cross-validation (Diagnostics 2025, 15, 211, 17). The first row of the non-SMOTE models built from the first cross-validation data is this. Afterwards, the entire 5-fold validation results for both SMOTE- and non-SMOTE-based models are shown in rows two through five.

Table 1. AUC-based performance of DL, SVM, GBM, and RF models from gene expression data.

ROC Score of SMOTE-Based Models

DL			SVM			GBM			RF		
CN	MCI	Dementia	CN	MCI	Dementia	CN	MCI	Dementia	CN	MCI	Dementia
0.717	0.681	0.634	0.672	0.651	0.597	0.711	0.667	0.548	0.593	0.575	0.620
0.763	0.761	0.706	0.690	0.648	0.660	0.659	0.660	0.574	0.617	0.597	0.534
0.736	0.644	0.602	0.761	0.664	0.572	0.678	0.635	0.577	0.613	0.578	0.527
0.669	0.668	0.607	0.694	0.647	0.586	0.678	0.655	0.639	0.639	0.592	0.619
0.732	0.729	0.613	0.668	0.641	0.572	0.663	0.655	0.618	0.598	0.574	0.601

ROC Score of Non-SMOTE-Based Models

DL			SVM			GBM			RF		
CN	MCI	Dementia	CN	MCI	Dementia	CN	MCI	Dementia	CN	MCI	Dementia
0.717	0.713	0.620	0.654	0.616	0.500	0.717	0.654	0.518	0.655	0.632	0.500
0.682	0.662	0.572	0.676	0.632	0.500	0.614	0.659	0.592	0.579	0.570	0.496
0.773	0.703	0.593	0.663	0.629	0.500	0.634	0.670	0.613	0.622	0.612	0.500
0.735	0.706	0.607	0.637	0.601	0.500	0.668	0.638	0.584	0.622	0.573	0.500
0.698	0.689	0.600	0.658	0.628	0.500	0.663	0.614	0.563	0.554	0.547	0.520

Table 2. F1-score-based performance of DL, SVM, GBM, and RF models from gene expression data.

F1-Score of SMOTE-Based Models

DL			SVM			GBM			RF		
CN	MCI	Dementia	CN	MCI	Dementia	CN	MCI	Dementia	CN	MCI	Dementia
0.63	0.71	0.39	0.57	0.70	0.32	0.62	0.70	0.20	0.45	0.64	0.37
0.71	0.77	0.53	0.60	0.67	0.41	0.55	0.68	0.25	0.46	0.65	0.18
0.63	0.69	0.33	0.67	0.71	0.27	0.55	0.67	0.29	0.44	0.66	0.17
0.52	0.70	0.37	0.56	0.69	0.36	0.53	0.68	0.43	0.48	0.65	0.39
0.65	0.73	0.36	0.55	0.67	0.28	0.53	0.69	0.36	0.42	0.64	0.33

F1-Score of Non-SMOTE-Based Models

DL			SVM			GBM			RF		
CN	MCI	Dementia	CN	MCI	Dementia	CN	MCI	Dementia	CN	MCI	Dementia
0.63	0.75	0.37	0.52	0.71	0.00	0.62	0.72	0.09	0.51	0.74	0.00
0.58	0.69	0.25	0.57	0.69	0.00	0.45	0.71	0.30	0.32	0.68	0.00
0.68	0.73	0.32	0.51	0.74	0.00	0.48	0.75	0.36	0.42	0.74	0.00
0.62	0.75	0.36	0.47	0.70	0.00	0.52	0.70	0.30	0.43	0.69	0.00
0.60	0.72	0.33	0.52	0.71	0.00	0.53	0.68	0.24	0.27	0.67	0.08

Evidently, SMOTE-based models outperform their non-SMOTE-based equivalents in most cases. In addition, compared to Alzheimer's disease and dementia stage identification, the effectiveness of CN and MCI stage diagnosis is clearly greater. As shown in the previous section, this is because the sample size is bigger compared to the minority AD samples. A low sample size impacts effectiveness in detecting learnt characteristics because it reduces learning, whereas a high sample size enhances learning of the individual features involved. The use of minority oversampling strategies does, however, improve performance in identifying stages of Alzheimer's disease and dementia. When comparing the performance of each instance of a model that uses SMOTE with its counterpart that does not, it becomes clear that there is an improvement. Compared to other studies that used ADNI gene expression data, the outcomes of this one are much better [9,18]. This proves that our feature selection is better and that our data augmentation of minority samples for training is creative. The performance metrics of the five instances of the deep learning models built using 5-fold cross-validation and based on clinical data are shown in Table 8. Participating deep learning algorithms outperform gene expression-based models in terms of performance metrics such ROC AUC values and F1-scores for multiclassification using clinical data. The significantly bigger sample size, less complicated data structure, and lack of problems with gene expression data like imbalanced data, multiple pathways involved, and high-

ISSN:0976-0172

Journal of Bioscience And Technology
www.jbstonline.com

dimensionality, low-sample-size (HDLSS) traits are the reasons for this superiority.

Table 3. Performance of the DL model constructed using clinical data.

F1 Score			ROC AUC		
CN	MCI	Dementia	CN	MCI	Dementia
0.983	0.923	0.847	0.983	0.937	0.877
0.971	0.939	0.886	0.989	0.927	0.907
0.944	0.931	0.869	0.967	0.921	0.915
0.968	0.933	0.846	0.968	0.944	0.897
0.947	0.856	0.823	0.941	0.893	0.908

Figure 1shows a graphical comparison of the top models developed utilizing clinical data and gene expression data. A number of assessment measures are used, such as the ROC curve, the PR curve, and the CM (confusion matrix). A model trained on clinical data has higher multiclassification skills, as shown by the visual representations and results of the confusion matrix. Both models outperform the AD category, which comprises minority samples, in terms of CN and MCI stage identification, according to the data. We compared and contrasted the two kinds of models by focusing on the AD/dementia category's classification accuracy, as it reflects the minority samples more than the other CN and MCI categories. Based on the confusion matrix, out of 131 AD/dementia testing samples, 109 were classified as AD/dementia, 22 as MCI, and no samples were classified as CN by the clinical data-driven model. In contrast, out of 18 AD/dementia testing samples, the blood gene expression model classified 8 as AD/dementia, 7 as MCI, and 3 as CN. For a variety of reasons discussed elsewhere, the clinical data-based model outperformed the gene-based model in terms of classification accuracy. Though there were 18 dementia samples altogether, the gene-based model successfully differentiated 15 samples with mixed Alzheimer's disease and mild cognitive impairment (MCI) from the cognitive normal (CN) group.

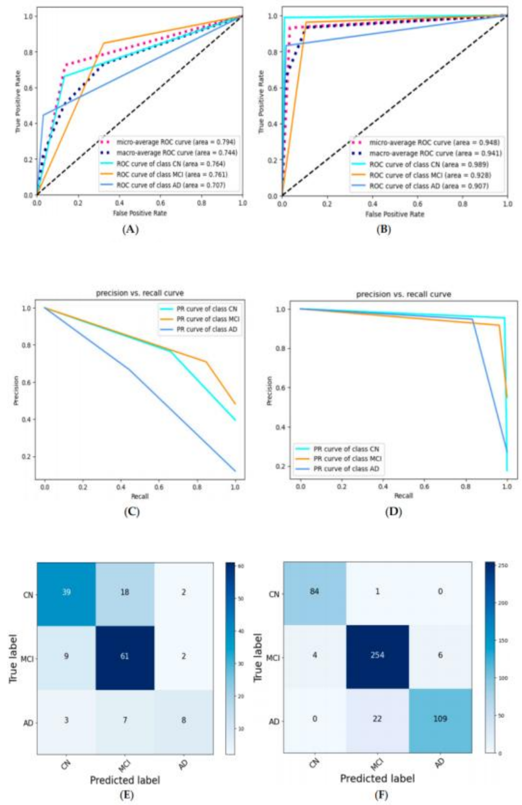
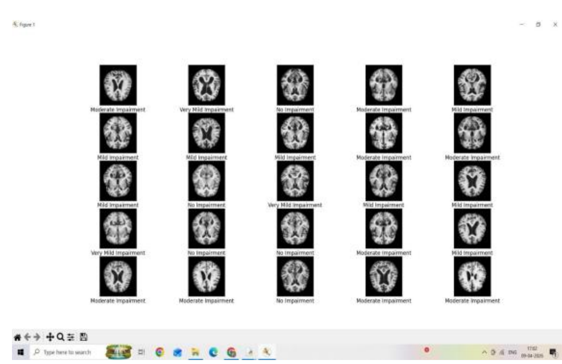
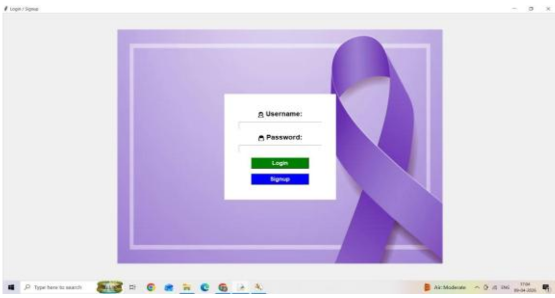
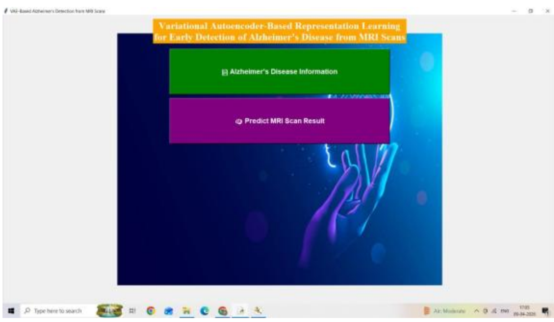


Figure 1. Performances of gene expression-based model and clinical data-based model. (A) ROC curve, gene expression-based model; (B) ROC curve, clinical data-based model. (C) PR curve, gene expression-based model; (D) PR curve, clinical data-based model. (E) CM, gene expression-based model. (F) CM, clinical data-based model.

RESULTS

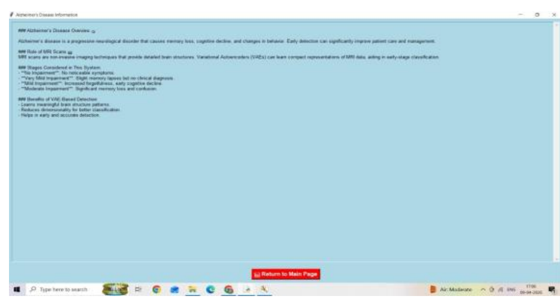






CONCLUSION

The suggested architecture shows promise for early Alzheimer's disease diagnosis from MRI images by integrating Convolutional Neural Networks with Variational Autoencoders. The system efficiently gathers compressed and discriminative data on structural brain changes by using VAEs for latent representation learning



and CNNs for feature extraction. When compared to more conventional statistical approaches like Naive Bayes, this strategy significantly improves classification accuracy. The experimental findings show that the model can accurately differentiate between healthy people and those with early-stage Alzheimer's disease. With deep learning, automated analysis is possible, cutting down on the need for human clinicians to make subjective assessments. In sum, the research shows that AI-driven diagnosis might be a useful tool for helping doctors intervene quickly. One step closer to precision therapy for neurodegenerative diseases is the use of state-of-the-art neural architectures.

FUTURE SCOPE

Training on bigger and more varied MRI datasets may expand our work in the future to make it more population-specific. The inclusion of multimodal data, which includes MRI and PET scans as well as genetic information and clinical history, might improve the framework and allow for a more comprehensive diagnosis. In addition to detecting diseases, longitudinal data might also be used to anticipate how they would proceed. More complicated spatial and temporal connections in brain scans may be captured by exploring advanced models like graph neural networks and transformers. Also, to help doctors make more understandable decisions, explainable AI approaches should be used. Connecting theory with practice in healthcare is the goal of deploying the model into real-time clinical decision support systems. This study ultimately opens the way

REFERENCE

[1] World Alzheimer Report 2025, Alzheimer's Disease International, London, UK (2025). World Alzheimer Report 2025

[2] Kumar, Gupta, Chauhan, and Shah, Alzheimer's Disease: Diagnosis, Molecular Targets, and Therapeutic Modalities, Routledge (2025). Alzheimer's Disease: Diagnosis...

[3] Budson, A. E., and O'Connor, M. K., Six Steps to Managing Alzheimer's Disease and Dementia: A Guide for Families, Oxford University Press (2021). Six Steps to Managing...

[4] Cummings, J., Kinney, J., and Fillit, H., Alzheimer's Disease Drug Development, Cambridge University Press (2022).

[5] Alzheimer's Association, 2024 Alzheimer's Disease Facts and Figures, Chicago, IL (2024). 2024 Facts and Figures

[6] Indrakumari, R., Deep Learning in Medical Image Analysis: Recent Advances and Future Trends, IET (2025).

[7] Bhat, S. Y., Rehman, A., and Abulaish, M., Deep Learning Applications in Medical Image Segmentation: Overview, Approaches, and Challenges, John Wiley & Sons (2025).

[8] Zhou, S. K., Greenspan, H., and Shen, D., Deep Learning for Medical Image Analysis, 2nd Edition, Academic Press/Elsevier (2023).

[9] Santosh, K. C., Deep Learning Models for Medical Imaging, Academic Press (2021).

[10] Lee, G., and Fujita, H., Deep Learning in Medical Image Analysis: Challenges and Applications, Springer Nature (2021).

[11] Rabie, K., Karthik, C., and Chowdhury, S., Deep Learning in Medical Image Processing and Analysis, IET Healthcare Technologies (2023).

[12] Mori, S., and Nielsen, M., Medical Image Computing and Computer Assisted Intervention – MICCAI 2025 Proceedings, Springer (2026). MICCAI 2025 Proceedings

[13] Tiwari, A., Deep Learning for Diffusion Tensor Imaging Estimation: Unlocking Sparse Diffusion MRI, Eliva Press (2025).

[14] Raman, B., Deep Learning Through the Prism of Tensors, Springer-Nature (2025).

[15] Hicks, B., and Phipps, O., Journal of Dementia and Alzheimer's Disease: Collected Works, MDPI Books (2026).

[16] Nielsen, M., Neural Networks and Deep Learning in Neuroimaging, Springer (2024).

[17] A. Tiwari, Deep Learning Foundations & Modern Architectures: Mastering Transformers and Diffusion Models, Amazon Digital (2026).

[18] Aggarwal, C. C., Artificial Intelligence: A Graduate Textbook, Springer (2021).

[19] Atienza, R., Advanced Deep Learning with TensorFlow 2 and Keras, 2nd Edition, Packt (2026).

[20] Mayo Clinic, Mayo Clinic on Healthy Aging, Mayo Clinic Press (2026)