International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

Review paper of Alzheimer's Disease Detection for Imbalance Dataset using machine Learning

[1]Md Ashif Karim, [2]Ms. Ruchi Dronawat, [3]Rupali Chaure
Department of Computer Science and Engineering (Data Science), Sagar Institute of Research and Technology, Bhopal[1]
Department of Computer Science and Engineering, Sagar Institute of Research and Technology, Bhopal[2,3]

ABSTRACT
Alzheimer's Disease is a progressive neurodegenerative disorder that affects memory, thinking ability, and cognitive functions, primarily occurring in elderly individuals. Early detection of Alzheimer's disease is essential for effective treatment planning, patient monitoring, and slowing disease progression. Traditional diagnostic methods rely on clinical examinations, cognitive assessments, and brain imaging techniques, which are often time-consuming, costly, and dependent on expert analysis. In recent years, machine learning approaches have shown significant potential in the automated detection and classification of Alzheimer's disease using medical datasets and neuroimaging data. However, one of the major challenges in Alzheimer's disease prediction is the imbalance present in medical datasets, where the number of healthy samples is significantly higher than diseased cases, leading to biased model performance and poor minority-class prediction accuracy.
This review paper presents a comprehensive analysis of machine learning techniques used for Alzheimer's disease detection on imbalance datasets. The study also discusses different imbalance handling techniques including Synthetic Minority Oversampling Technique (SMOTE), undersampling, oversampling, cost-sensitive learning, and hybrid sampling approaches. In addition, preprocessing methods, feature extraction techniques, neuroimaging datasets, and evaluation metrics such as accuracy, precision, recall, sensitivity, specificity, and F1-score are analyzed. Comparative analysis indicates that machine learning models combined with imbalance handling techniques significantly improve Alzheimer's disease prediction performance and reduce classification bias. The review concludes that intelligent machine learning systems provide reliable and efficient solutions for early Alzheimer's disease diagnosis and can support healthcare professionals in clinical decision-making and patient care management.
Keywords: Deep Learning; Alzheimer's Disease; MRI; Early Diagnose.

1. INTRODUCTION
Alzheimer's Disease (AD) creates memory loss and cognitive decline; it is a neurological disorder, and neuro degenerative type of dementia, Alzheimer's occurs because of brain cell death [1]. The size of brain gets shrinks with Alzheimer's - the tissue has increasing minimum nerve cells connections. The following fig. 1 shows the existence of AD on neuron cells.
While they can't be seen or tried in the living brain influenced by Alzheimer's disease, postmortem/autopsy will dependably indicate minor incorporations in the nerve tissue, called

plaques and tangles:

Plaques – it is identified among the dying cells in the brain - from the increase of a protein called beta-amyloid.

Tangles – it is identified within the brain neurons- due to the disintegration of another protein, called tau.

Fig. 1: Nerve cells (neurons) in the brain, in Alzheimer's

The unusual protein bunches, in the cerebrum tissue is constantly present with the disease, however there could be another fundamental procedure that is really causing the AD, and researchers are not yet beyond any doubt [2, 3]. This kind of progression in brain nerves is seen in different scatters, and analysts need to discover something other than that there are protein variations from the norm - they additionally need to know how these grow with the goal that a cure or aversion may be found. Specialists don't completely comprehend why the progressions that prompt Alzheimer's illness. A few distinct components are accepted to be included [4].

2. ALZHEIMER'S DISEASE

Symptoms:- Indications can be analyzed at any phase of Alzheimer's dementia and the movement through the phases of the infection is observed after an underlying analysis when creating the manifestations directs how mind is overseen. Obviously, the side effects can mystify both a patient and the general population around them, with various levels of seriousness. Consequently, and on the grounds that indications could flag any of various analyses, it is constantly beneficial consulting with a neuro specialist. For neuro specialists to make an underlying conclusion of Alzheimer's sickness, they should first be fulfilled that there is an existence of dementia. It includes cognitive or behavioral symptoms that demonstrate a decrease from past levels of "working and performing" and meddle with capacity to operate at work or at common exercises.

Risk Factors:- A few things are all the more ordinarily connected with Alzheimer's ailment - not seen so frequently in individuals without the turmoil. These components may along these lines have some immediate association. Some are preventable or modifiable variables for instance, reducing the danger of diabetes or coronary illness may thus cut the danger of dementia. On the off chance that analyst's acquire comprehension of the hazard factors, or logically demonstrate any “cause” association for Alzheimer's, this could discover approaches to anticipate it or create medications. Hazard factors related with Alzheimer's sickness incorporate.

Unavoidable risk factors:- Age - the turmoil is more probable in older individuals and a more prominent extent of more than 85-year-olds , than of more than 65's.

Family history - having Alzheimer's in the family is related with higher hazard. This is the second greatest hazard factor after age. Having a specific quality puts a man, contingent upon their particular hereditary qualities, at three to eight times more hazard than a man without the quality. Various different qualities have been observed to be related with Alzheimer's disease. Factors which maximizes the blood vessel risk - involves diabetes, high cholesterol and high blood pressure. While a traumatic brain injury does not essentially direct to Alzheimer's, some research links have been drawn, with increasing risk tied to the harshness of trauma history.

3. LITERATURE REVIEW

Y. P. Reddy et al. [1], proposed an Alzheimer's disease detection framework using deep feature extraction and explainable machine learning techniques. The study focused on improving the accuracy and interpretability of Alzheimer's disease diagnosis using neuroimaging data and intelligent classification models. The researchers utilized deep learning-based feature extraction methods to automatically identify significant patterns from brain MRI images associated with cognitive decline and neurodegenerative changes. After extracting deep features, machine learning classifiers such as Support Vector Machine (SVM), Random Forest, and ensemble learning techniques were employed for disease classification. The study also integrated explainable artificial intelligence methods to provide transparent prediction results and help clinicians understand the contribution of different brain features in diagnosis. Various preprocessing techniques including normalization, image enhancement, and feature selection were implemented to improve model performance and reduce computational complexity. Experimental evaluation demonstrated that the proposed explainable deep feature extraction framework achieved higher accuracy, precision, recall, and F1-score compared to conventional diagnostic approaches. The authors concluded that combining deep learning with explainable machine learning can significantly improve reliable and interpretable Alzheimer's disease detection systems for healthcare applications.

A. Basat, et al. [2], presented an Alzheimer's disease analysis system using machine learning approaches for early-stage disease prediction and classification. The study aimed to develop an intelligent healthcare model capable of identifying Alzheimer's disease symptoms from clinical and neuroimaging datasets. Several supervised machine learning algorithms including Decision Tree, Naïve Bayes, Random Forest, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) were implemented and compared to determine the most efficient classification model. Data preprocessing techniques such as feature extraction, normalization, missing value handling, and dimensionality reduction were applied to improve classification performance and reduce data inconsistencies. The researchers analyzed the effectiveness of machine learning models using performance evaluation metrics such as accuracy, precision, sensitivity, specificity, and F1-score. Experimental findings revealed that ensemble and tree-based learning models achieved superior diagnostic accuracy compared to traditional statistical methods. The study concluded that machine learning approaches can provide fast, efficient, and accurate Alzheimer's disease diagnosis systems for smart healthcare environments.

D. Alatrany et al. [3], proposed a SHAP-based interpretability framework for Alzheimer's disease classification using ensemble learning techniques. The study primarily focused on improving the transparency and reliability of machine learning-based medical diagnosis systems through explainable artificial intelligence methods. The researchers implemented ensemble learning classifiers for Alzheimer's disease prediction using patient and neuroimaging datasets and integrated SHAP (SHapley Additive exPlanations) techniques to interpret the prediction outcomes. Various preprocessing methods such as normalization, feature engineering, and dimensionality reduction were applied to enhance the learning process and improve classification accuracy. The proposed ensemble framework demonstrated high predictive performance and provided detailed explanations regarding the importance of different clinical and imaging features in disease prediction. Experimental results showed improved accuracy, recall, and precision compared to standalone machine learning approaches. The study highlighted that explainable machine learning techniques are highly beneficial in healthcare applications because they increase physician trust and improve the interpretability of automated diagnostic systems.

O. A. Dara et al. [4], presented a comprehensive survey on Alzheimer's disease diagnosis using machine learning techniques. The study reviewed various machine learning and deep learning approaches utilized for automated Alzheimer's disease detection from neuroimaging data, cognitive assessments, and clinical datasets. The survey analyzed different classification models including Support Vector Machine (SVM), Random Forest, Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and hybrid learning approaches. The researchers also discussed preprocessing methods, feature extraction strategies, dimensionality reduction techniques, and publicly available Alzheimer's disease datasets used in previous studies. Comparative analysis was conducted based on performance metrics such as accuracy, sensitivity, specificity, and computational efficiency. The survey highlighted that deep learning and ensemble machine learning methods achieved superior diagnostic performance due to their capability to learn complex brain patterns automatically. The study concluded that intelligent machine learning systems have strong potential for early Alzheimer's disease diagnosis, clinical support, and cognitive disorder monitoring in modern healthcare systems.

K. N. Rao et al. [5], proposed a machine learning-based framework for prediction and classification of Alzheimer's disease using 3D MR brain images. The research focused on improving disease detection accuracy by utilizing three-dimensional MRI scans that provide detailed structural information about brain tissues affected by Alzheimer's disease. Various preprocessing techniques including image enhancement, normalization, segmentation, and feature extraction were implemented to improve MRI image quality and extract significant brain features related to neurodegeneration. The researchers applied multiple machine learning algorithms for classification and compared their performance using evaluation metrics such as accuracy, sensitivity, specificity, and precision. Experimental analysis demonstrated that machine learning techniques effectively classified Alzheimer's disease stages from 3D MR images with improved prediction reliability. The study concluded that intelligent MRI-based machine learning systems can support healthcare professionals in early Alzheimer's disease diagnosis, disease progression monitoring, and effective patient treatment planning.

Jin Liu et al. [6], proposed a Median Filter approach for the removal of salt and pepper noise and Poisson noise away from the images. The function of a median filter is such that, the output intensity value of the pixel that is being processed is obtained by the sliding of a window all along the image and then the median intensity value of the pixels inside the window results to be the output intensity. Also, the median filter maintains the edges in an image with random noise being reduced. The value of every pixel is fixed to median value of the pixels close to the respective input pixels. Then this filter is used for removing these noises and afterwards the bounding box approach is used for detecting the tumor position.

Jun Zhang et al. [7], have researched the evolution of order statistics filters that lets a simple and efficient technique to reduce the noise from medical images. This approach uses an integrated median filtering and mean filtering to determine about the pixel value present in the no-noise image. It is also used for eliminating the Rician noise that affects the images.

A. Chincarini et al. [8], have developed an anisotropic filter for the elimination of background noise and thereby protecting the edge points present in the image. This technique is involved with concurrent filtering and contrast stitching. The selection of a Diffusion constant is related to the noise gradient and it smoothens by eliminating noises present in the background with the filtering done with a right value of threshold.

Li et al. [9], introduced a new technique for the enhancement of MRI image that, in turn, is dependent on the Modified Tracking Algorithm, Histogram Equalization and Center Weighted Median (CWM) filter. This technique comprises of two schemes. The first scheme is the application of the modified tracking algorithm to eliminate the film artifacts, labels and skull region and therefore using the Histogram Equalization and CWM filter methods individually for the images enhancement.

Jin Liu et al. [10], provided a count on the voxel value histogram in significant anatomical areas that could be acquired either by image segmentation or the registering of a brain atlas over the image. But, the anatomical parcellation of brain is not an easy task and might not adapt to the pathology. Permits the examination of focal differences in brain anatomy employing the statistical parametric mapping (SPM), and therefore hugely helps in extracting voxel-based features.

Heung-Il Suk et al. [11], first applied voxel morphometry analysis for the extraction of few of the most potential AD-relevant features of the brain images from the actual MRI volumes and also the Gray Matter (GM) segmentation volumes. The features should acquire the most discriminative characteristics, which differ between the brain of a healthy and Alzheimer affected individual. Then, a Principal Component Analysis (PCA)-based dimension reduction over the acquired features is conducted for an analysis that is quicker and still adequately accurate. In order to take the best advantage of the extracted features, a hybrid manifold learning framework is presented that includes the feature vectors in a subspace.

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

4. METHODOLOGY

Machine Learning: Machine Learning is a branch of artificial intelligence (AI) that enables computers and systems to learn from data and make predictions or decisions without being explicitly programmed. Machine learning algorithms identify hidden patterns, relationships, and trends in datasets and improve their performance automatically through experience. It is widely used in various applications such as healthcare, image processing, cybersecurity, finance, agriculture, speech recognition, recommendation systems, and autonomous vehicles.

Machine learning works by training models using historical or labeled data. During training, the algorithm learns the relationship between input data and output results. After training, the model can predict outcomes for new unseen data. The overall machine learning process generally includes data collection, preprocessing, feature extraction, model training, testing, and performance evaluation.

Deep Learning: Deep learning is one of AI techniques. It comprises of progressive organized layers that can interpret input information to significant yields in a discovery model. Deep learning strategies include wide applications inside various research investigation, for example, graphical displaying of information, neural systems, parameters streamlining, picture investigation, design acknowledgment furthermore, signal preparing. Numerous profound learning models in different applications depend on the model, proposed by Yann LeCun for transcribed acknowledgments by utilizing profound directed backpropagation convolutional organize [6].

Deep neural system is a straightforward ordinary neural system with progressively concealed layers with the goal that it gets further. The profound neural engineering can be considered as the speculation of a direct or on the other hand calculated relapse neural system designs. Each neuron is enacted in a straight mix of information and a few learning parameters, which are trailed by a component savvy nonlinear development.

A neural system design comprises of various layers L of a weighted neuron through which enactment is performed. Multi-Layer Perceptron (MLP) is a class of feedforward neural system with at least two layers between information and yield layers. The feedforward implies that information owes one way from contribution to yield layers. The backpropagation learning calculation is utilized to prepare the MLP. MLP is utilized in numerous applications, for example, design classification, acknowledgment, guess, and expectation. MLP for the most part takes care of the issues which are not directly distinct as appeared in Fig. 2.

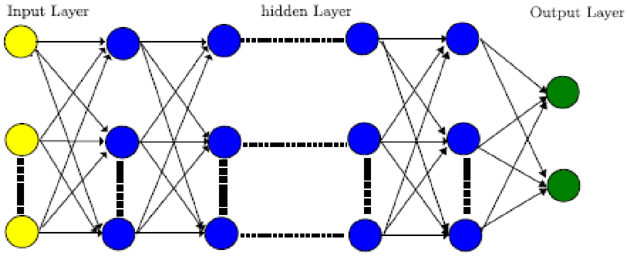

Fig. 2: Deep neural network

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

In any case, there are elective strategies to prepare an entire profound neural arrange start to finish in a directed manner. These other option strategies can be better executed by specified sort of neural arrange, i.e., the convolutional neural system (CNN).

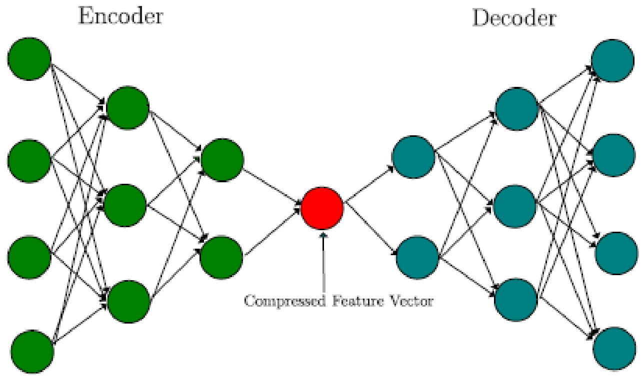

Fig 3: Deep auto-encoder (AE)

These days, CNN gets increasingly well known in clinical picture preparing what's more, turns into a decent decision for the analysts in clinical picture investigation. The accompanying subsections clarify the fundamental sorts of neural systems and bit by bit usage of the technique with their impediments and points of interest.

Image Segmentation:- Image segmentation is the procedure that subdivides an image into its constituent parts or objects. Segmentation is a standout amongst the most broadly utilized strides in reducing images to data (John, 1999). Segmentation techniques were utilized to confine the coveted object from the scene with the goal that estimations have been made on it consequently. The level to which this subdivision is done relies upon the issue being tackled, i.e., the segmentation should end when the objects of enthusiasm for an application have been separated from the image and afterward to portion the substance of the road down to potential vehicles. The algorithms for segmentation of images depend on discontinuity or similarity in pixel esteems. In the irregularity approach an image is partitioned on the premise of sudden changes in intensity, edges, points, blobs, and lines in the image. In similar approach, an arrangement of pre-characterized criteria is utilized for partitioning the image into areas that are comparable as per the model set. Thresholding, region growing, and region splitting and merging are cases of similarity-based segmentation techniques.

Feature Extraction:- Feature extraction portrays the significant shape of data contained in an example, so the work of arranging the example is made simple by a formal system. In pattern recognition and in image processing, feature extraction is an exceptional type of dimensionality decrease.

The primary objective of feature extraction is to acquire the most important data from the first information and speak to that data in a lower dimensionality space. At the point when the input information to an algorithm is huge to ever be handled and it is suspected to be repetitive then the input information will be changed into a diminished portrayal set of features. Transforming the input information into the arrangement of features is known as feature extraction. On the off chance that the features separated are painstakingly picked it is normal that the features set will extricate the applicable data from the information keeping in mind at the end goal to play out the coveted assignment utilizing this decreased portrayal rather than the full size input

5. CONCLUSION

In conclusion, machine learning techniques have become highly effective tools for the detection and classification of Alzheimer's Disease due to their ability to analyze complex medical and neuroimaging datasets with improved accuracy and efficiency. Various supervised learning approaches such as Support Vector Machine (SVM), Random Forest, Decision Tree, K-Nearest Neighbor (KNN), ensemble learning, and deep learning models have demonstrated significant performance in identifying Alzheimer’s disease at early stages. The reviewed studies indicate that preprocessing methods, feature extraction techniques, explainable artificial intelligence, and hybrid learning frameworks further enhance prediction reliability and diagnostic interpretability. In addition, imbalance handling techniques such as SMOTE, oversampling, undersampling, and cost-sensitive learning help improve minority-class prediction performance and reduce classification bias in imbalanced medical datasets. Machine learning-based diagnostic systems provide fast, cost-effective, and reliable solutions that can assist healthcare professionals in clinical decision-making and patient monitoring. Therefore, intelligent healthcare systems integrated with machine learning offer promising opportunities for early Alzheimer’s disease diagnosis and improved patient care management. Future research can focus on multimodal data integration, lightweight deep learning models, explainable AI frameworks, and real-time healthcare applications to further improve diagnostic performance and clinical reliability.

REFERENECS

[1] Y. P. Reddy, S. V. M and J. E. S, "Alzheimer's Disease Detection using Deep Feature Extraction and Explainable Machine Learning," 2026 5th International Conference on Sentiment Analysis and Deep Learning (ICSADL), Birendranagar, Nepal, 2026, pp. 135-140.

[2] A. Basat, A. M. Ramly, P. S. Chakolthi and M. S. Sharif, "Alzheimer's Disease Analysis Using Machine Learning Approach," 2025 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Sakhir, Bahrain, 2025, pp. 1-6.

[3] D. Alatrany, M. Abdulzahra, K. Qader and M. Qader, “SHAP based interpretability method for Alzheimer's disease classification using ensemble learning,” International Journal of Online and Biomedical Engineering, vol. 20, no. 3, pp. 55–71, 2024.

[4] O.A. Dara, J.M. Lopez-Guede, H.I. Raheem, J. Rahebi, E. Zulueta and U. Fernandez-Gamiz, “Alzheimer's disease diagnosis using machine learning: a survey,” Applied Sciences, vol. 13, no. 14, p. 8298, 2023.

[5] K.N. Rao, B.R. Gandhi, M.V. Rao, S. Javvadi, S.S. Vellela and S.K. Basha, “Prediction and classification of Alzheimer's disease using machine learning techniques in 3D MR

images," in Proc. 2023 Int. Conf. Sustainable Computing and Smart Systems (ICSCSS), 2023, pp. 85–90.

[6] Jin Liu, Jianxin Wang, Bin Hu, Fang-Xiang Wu and Yi Pan, "Alzheimer's Disease Classification Based on Individual Hierarchical Networks Constructed With 3-D Texture Features," IEEE Trans on Nano Bioscience, vol. 16, pp. 428-435, September 2017.

[7] Jun Zhang, Yue Gao, Yaozong Gao, Brent C. Munsell and Dinggang Shen, "Detecting Anatomical Landmarks for Fast Alzheimer's Disease Diagnosis," IEEE transections on medical imaging, vol. 35, NO. 12, December 2016.

[8] Li, F., Tran, L., Han Thung, K. et al, "A Robust Deep Model for Improved Classification of AD/MCI Patients," IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 5, pp. 1610-1619, 2015.

[9] Jin Liu, Min Li, Wei Lan, Fang-Xiang Wu, Yi Pan, and Jianxin Wang, "Classification of Alzheimer's Disease Using Whole Brain Hierarchical Network," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 14, NO. 8, August 2015.

[10] Alzheimer's disease from healthy controls. A European Alzheimer's Disease Consortium (EADC) study," Neuro Image: Clinical, vol. 7, pp.34-42, 2015.

[11] Manhua Liu, Daoqiang Zhang, Dinggang Shen and The Alzheimer's Disease Neuroimaging Initiative, "Ensemble sparse classification of Alzheimer's disease," Neuro Image, vol. 60, 1106–1116, January 2012.

[12] E. Gerardin et al., "Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging," Neuro Image, vol. 47, no. 4, pp. 1476– 1486, 2009.

[13] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," IEEE Trans. Med. Imag., vol. 17, no. 1, pp. 87–97, Feb. 1998.

[14] Friston, K.J., Holmes, a. P., Worsley, K.J., Poline, J.-P., Frith, C. D., Frackowiak, R.S.J.: Statistical parametric maps in functional imaging: Ageneral linear approach. Hum. Brain Mapp. 2, 189–210 (1995).