

## Explainable AI Models for Predicting Neurological Disorders

**Muhammad Arham**

Department of Computer Science, The University of Faisalabad.

[drmuhammadarham4@gmail.com](mailto:drmuhammadarham4@gmail.com)

**Aqsa Maqsood**

Institute of Management Sciences Peshawar. [Aqsamaqsood66@gmail.com](mailto:Aqsamaqsood66@gmail.com)

### ABSTRACT

The core of neurological diseases is the top in the world, as these diseases comprise the Alzheimer disease, Parkinson disease, epilepsy and stroke. And there is an urgent necessity to forecast such conditions in the earliest possible time and with accuracy to treat and administer the disease. It was discovered that the deep neural network (DNN) and machine learning systems could analyze relatively complicated neurological information, neuroimaging features and electrophysiological changes and medical history. However, the complexity of AI models, which is the black-box of most of them, makes them not applicable to using them in clinical practice, since clinicians require a predictive system that is transparent and interpretable and can be expected to rely on it. The field of explainable artificial intelligence (XAI) is inspiring as a promising approach to these issues as it offers a way to interpret model predictions in a human-understandable manner. The current review discusses the existing XAI methods and their use in predicting neurological disorders with a particular focus on their role in improving clinical judgment, patient trust, and regulatory adherence. Besides, we present the issues concerning data quality, generalizability, and ethical issues and the future perspectives of applying XAI to the routine in neurological practice.

**Keywords:** Explainable AI (XAI), Neurological disorders, Machine learning, Deep learning; Interpretability, Clinical decision support, Neuroimaging; Alzheimer's disease, Parkinson's disease, Epilepsy prediction.

### 1. INTRODUCTION

Some of the most urgent health issues in the world health are neurological disorders that are associated with high rates of morbidity, mortality and economic costs. Diseases like Alzheimer, Parkinson, epilepsy, stroke among others impact millions of

persons with various diseases increasing in frequency as people grow older and environmental stressors mount up. These conditions are typified by gradual cognitive loss, motor dysfunction, sporadic seizures or acute brain stroke and they may lead to long term disability that burdens healthcare systems and caregivers with an enormous load [1, 2]. There is yet no perfect diagnosis and prognosis of neurological diseases despite the great advances that are being made in the area of neuroimaging, molecular genetics and clinical diagnosis. In the vast majority of cases, the clinical manifestations become apparent after the neuronal harm is already inflicted and the disease-modifying therapies, or preventive action, cannot work [3, 4]. This disparity produces a requirement to implement, urgently, computational techniques capable of possessing fine disease signatures capable of producing a clinical manifestation more subtly in a subtle manner. One of the most radical technologies that occur in the contemporary field of medicine is artificial intelligence (AI). In particular, machine learning (ML) and deep learning (DL) algorithms have discovered dazzling capabilities when it comes to handling high, multi-modal data e.g., MRI and CT, electroencephalographic (EEG) and records of electronic health (EHR) [5, 6]. Whatever patterns the human expertise will disregard, AI can change the quality of the diagnostic process, the prognostic analysis, and the personalized treatment plans. Other works in the medical field demonstrated that convolutional neural networks (CNNs) can detect structural brain defects associated with Alzheimer disease or multiple sclerosis with a relatively high accuracy comparable to a trained radiologist [7-9]. In this regard, recurrent neural networks (RNNs) and transformer-based models have been used to predict EEG sequences of seizures, which have been shown to surpass human accuracy. However, AI has not been adopted as easily as expected into the everyday life of a neurologist. One of the most significant causes of this reluctance is the lumpiness of most AI systems. Although they are accurate, deep learning models have been criticized as being a black box as clinicians cannot access their internal decision-making processes [10]. Such deficiency in interpretability is not merely an inconvenience in technical terms, but a basic obstacle in clinical medicine, where the decisions made by a physician have to be clarified, rationalized and reported to patients and regulatory authorities [11].

## 1.1. Explainable AI (XAI) is needed in Clinical Prediction

Explainable artificial intelligence (XAI) has been driven by the need to provide transparency in AI-based medical devices. The reason is obvious: no one is going to trust the predictions which cannot be explained, no matter how accurate they are statistically. In neurology, where treatment choices may involve invasive interventions, e.g. deep brain stimulation or surgery to treat epilepsy, clinicians need to be confident that the outputs of their models depend on the physiological plausibility features, and not on spurious correlations. XAI aims to fill this gap by generating intelligible explanations to human beings, consequently, enhancing trust, accountability, and ethics compliance [12, 13]. The fact that explainability is required is not limited to clinicians. Clients and their families would frequently need assurance of the diagnostic outcomes as the patients are already mentally strained by the symptoms of the neurological conditions. In addition, medical AI tools transparency is turning into a legal or regulatory necessity, such as the European Union AI Act or the U.S. Food and Drug Administration (FDA) regulations. Even the good predictive models are subject to be rejected in clinical setting unless they have their explainability [14]. Therefore, XAI is not a desirable factor but an obligatory instrumentation in ensuring that AI can be adequately applied in the clinical use of neurology in a safe and ethical manner. The traditional machine learning or the deep learning Architectures can process neurological data in a variety of modalities, and explainability layers can make the results readable to the final user clinician [15, 16]. This workflow, as depicted in Figure 1.1, does not only assist in ensuring the predictive models are accurate, but also interpretable, which consequently leads to the trust of the physician and patient-centered care.

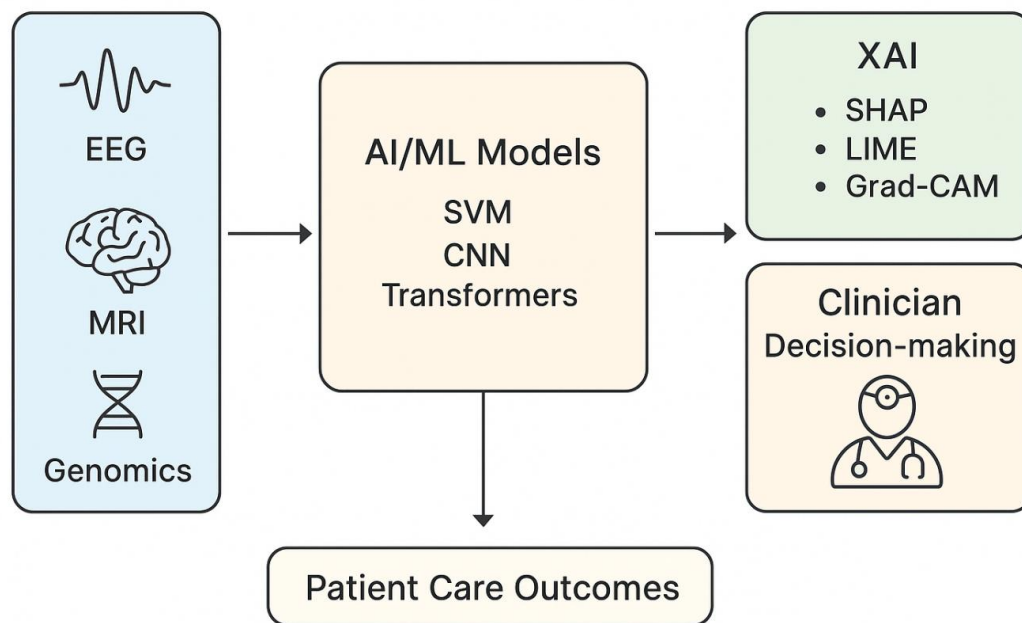


**FIGURE 1.1:** Conceptual framework showing the integration of AI/ML models with explainable AI techniques to enhance clinical decision-making in neurological disorders.

## 1.2. Neurological Disorders and Predictive Challenges

However, predictive modeling is facing its own difficulties, even though the potential of AI in the field of neurology is extremely large. Neurological disorders are disparate in exhibiting and the impetus of the illness, and commonly have identical symptoms, and this casts doubt on their differentiation [17-19]. In other words, in normal aging, and in mild stages of Alzheimer disease, memory impairment may be discovered, and tremors may be precipitated by Parkinson, essential tremor, or side effects developed due to the administration of some medications. Such clinical overlaps need predictive models which in addition to detection error-free are sensitive.

Another issue that makes prediction difficult is consideration of challenges in data. The neurological data is generally nonhomogeneous and hence carries imaging, electrophysiology, genomics and clinical notes among all these kinds of information requires specific preprocessing and integration [20]. Missing data is a perennial

problem with patients lost in the long-term studies and failure to take all the diagnosing tests [21]. Imbalance in classes is another huge challenge: when dealing with a large population slide, individuals with the disease is usually under-represented to the groups that are not quality to the disease, biased towards the most prevalent of the models. In addition, skeptics of AI models challenge that the sample sizes of rare neurological disorders are not large enough to generalise [22, 23].

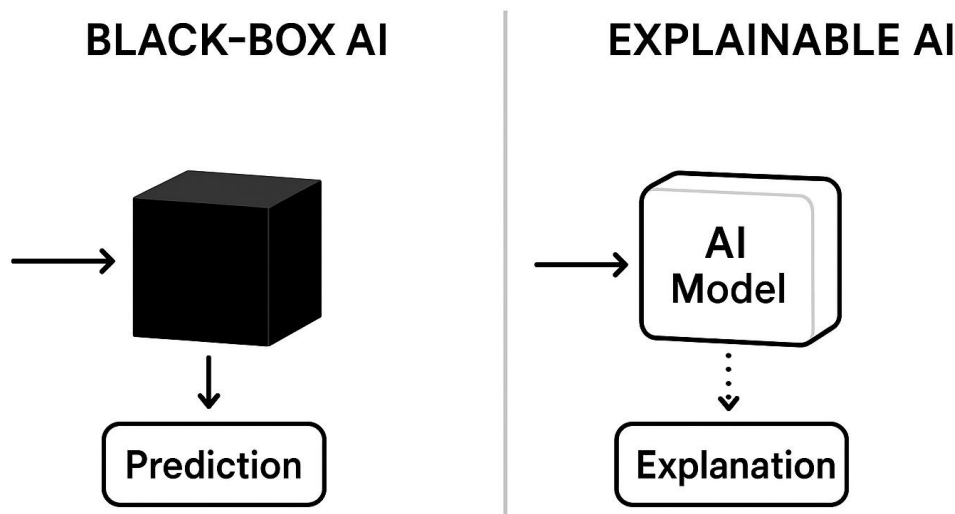
These challenges pose the issue of interpretability. An opaque model can even be observed to be very precise, yet with unrelated or misleading features arriving at clinical unsafe conclusions. By contrast, an explainable system may allow clinicians to validate the existence of predictions with known markers of disease such as hippocampal atrophy in Alzheimer or beta-band follies in Parkinson [24, 25]. This interpretability, as well as the fact that it enhances clinical reliability, contribute to the identification of new patterns that could extend the current knowledge on the mechanisms of the diseases.

### 1.3. Clinical Decision-Making Importance of Interpretability

In health care, interpretability equates with accountable so to speak. The physicians have a moral duty to explain to the patients, as well as the families and the oversight committees, their decision. The black-box predictions, which cannot be explained, cannot be employed in clinical practice, particularly at the high-stakes level [26, 27]. An example of this dilemma is in the field of neurology since in acute stroke; the choice to implant a deep brain stimulator, provide anti-epileptic drugs, or initiate thrombolysis must be anchored in sound evidence. The interpretability is as well synergistic to the thoughts of humans, where AI prioritizes comprehensive human conceptualization of variables dominating a given prediction [28, 29]. Also, explainable AI facilitates participative decision making. And patients who know more about what the disease is about of what they are going to see on their MRI scan is ready to take the advice of treatment. The interpretability can also be helpful as a teaching tool with medical trainees, who can be trained to relate the computational predictions and pathophysiological processes [30, 31]. Instead, the interpretability is the most important aspect as to the safe, detectable, and moral application of AI within the industry of neurology.

## 1.4. Explainable Artificial Intelligence (XAI)

Explainable AI is a shift in paradigm in opaque black-box systems to transparent and accountable systems. XAI is a collection of approaches aimed at explaining the way AI systems reach their outputs. XAI is fundamentally guided by the principles of transparency, interpretability, trustworthiness and accountability [32, 33]. Following these principles, XAI can not only enhance the clinical adoption but also comply with the ethical and regulatory requirements in the medical field. Another important difference that is crucial to make between traditional AI which is a black-box and explainable AI is their interpretability [34, 35]. Black-box models, as illustrated in Figure 1.4, can generate only predictions, unlike the XAI techniques that can provide clinicians with a better understanding of model predictions as it can present its reasoning.



Comparison of black-box and explainable AI

**FIGURE 1.4:** Comparison between black-box AI models that provide only predictions and XAI models that generate interpretable and transparent outputs.

The field is represented by two broad approaches. Post-hoc explainability is a method of using interpretability methods to explain a trained complex model. Notable ones are SHAP (Shapley Additive Explanations), which measures the value of each feature to a prediction, and LIME (Local Interpretable Model-Agnostic

Explanations), which tries to explain local model behavior with simpler interpretable ones. Grad-CAM (Gradient-weighted Class Activation Mapping) has been applied to imaging in imaging to indicate the localizations of interest with the strongest predictive power to allow clinicians to compare output with anatomical knowledge [36-38]. Intrinsic explainability is the second approach, which puts priority on interpretability by coming up with models that are transparently designed. They are decision trees, rule-based classifiers, and attention-based neural networks, where users can follow the way predictions were made by input features directly. In certain cases, these models might not be as predictive accurate as deep neural networks but have major clinical trust and usability benefits [30, 39]. These benefits of XAI may be demonstrated in comparison to black-box AI systems. The black-box models can be superior although may be less understandable and so cannot be used in clinical practice where explanations of the decisions are of utmost importance [40]. XAI, on the other hand, offers a mechanism used in trade-off, XAI maintains significant predictive powers of complex models, and simultaneously, render the results clinically significant and comprehensible. An example to demonstrate this is that an XAI-enhanced model to predict Alzheimer disease using MRI does not simply give a probability rating, but further distinguishes an execution of both gradient narrowing or vanishing hipmozomes, are the factors that would decide the decision-making [41]. This level of both accuracy and interpretability is precisely what can take XAI to the height of becoming the future of the neurological care. The fields of machine learning (ML) and artificial intelligences (AI) have particularly quickly become a new regulatory resource in the field of neurology and offer new prospects of prediction, diagnosis and surveillance of multifaceted brain problems [42]. The initial efforts to introduce computational models in clinical neurology were hugely dependent on classical machine learning methods including support vector machines (SVM), decision trees and random forests. These models were good at the management of structured datasets such as clinical scores, laboratory measurements and image-based features. As an example, structural MRI biomarkers were useful in identifying the difference between healthy and patients with early-stage Alzheimer's disease using SVM and random forests were effective in classifying Parkinsonian syndromes based on the combination of gait parameters and voice analysis [43].

Despite the fact that such methods provided sensible predictive accuracy, they were poor in predicting due to manual feature engineering, and limited ability to generalize between heterogeneous groups of patients.

The emergence of deep learning transformed the field of neurology as it enabled models to automatically extract high level representations of raw data of high dimensions. The use of convolutional neural networks (CNNs) was also found especially useful in neuroimaging, where it was used to identify cortical thinning, hippocampal atrophy, and subtle changes in white matter that are caused by dementia or epilepsy. Recurrent neural networks (RNNs) and their more modern counterparts, including long short-term memory (LSTM) networks, found common use in the analysis of sequential data such as electroencephalography (EEG) and magnetoencephalography (MEG) signals to predict attacks of epileptic seizures minutes before they occur [44, 45]. More recently, transformer-based architectures have started to have an impact on neurological prediction, using attention mechanisms to capture long-range multimodal dependencies between multimodal data, including combined imaging, genomics, and clinical records. These developments have enhanced precision greatly, yet have raised the issue of interpretability, which has become a greater issue as these developments have become more and more complex, thus increasing the issue of the black-box nature of this type of system. In order to counter the problems of single-model methods, scientists have resorted to hybrid and ensemble strategies to integrate advantages of an array of algorithms [46, 47]. Bagging, boosting and stacking have been used to combine predictions of different learners so as to enhance robustness and generalization in noisy clinical data. In the neurology field, there has been evidence of hybrid systems based on CNNs to extract features and other conventional classifiers such as SVMs as a way of stroke lesion segmentation and dementia staging. In a similar manner, multimodal groups which combine neuroimaging, electrophysiology, and clinical text data have provided superior results in complicated diagnostic work [48]. As much as these methods counter certain weaknesses of the individual models, they tend to increase the lack of transparency of the decision-making, making it even harder to identify the reasoning behind the predictions among the clinicians. The ever-increasing need to be transparent has

driven the use of explainable artificial intelligence (XAI) in neurological prediction [49]. XAI approaches are designed to explain the most significant aspects affecting the outputs of models, thus closing the gap between algorithmic and clinical reasoning. Some of the most popular methods include feature attribution methods that include Shapley Additive explanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME) and Integrated Gradients [50]. SHAP, which is founded on the cooperative game theory, uses importance scores on all features in all the possible model coalitions, and can be used to find biomarkers that are most likely to aid in the classification of Alzheimer disease based on imaging or genetic data. LIME offers local explanations through approximating complex models using the easiest to interpret surrogates to allow a clinician to know why a model predicted that a particular patient would experience imminent seizure activity. In deep neural networks, Integrated Gradients can identify areas of input data including particular voxels in an MRI scan that may have helped the classification decision [41, 51]. Such approaches have increased interpretability, but still have the problems of stability, computational cost and at times misleading explanations. Along with feature attribution, there exist other avenues of transparency with rule-based and symbolic approaches. These models attempt to produce human-read rules, lines of decision or symbolic representations, which may be cross-validated with current medical knowledge. As an example, symbolic regression methods have been employed to discover mathematical correlations between EEG signal patterns and epileptic activity and also decision rules obtained by interpretable neural-symbolic systems have been used to warn of early motor symptoms of Parkinsonism. These methods are particularly attractive because they are consistent with clinical reasoning, but in many cases they tend to make trade-offs with predictive accuracy against deep black-box models, which increases the trade-off between performance and interpretability [52, 53].

Neurological prediction case studies also indicate the opportunities and difficulties of XAI. Deep learning models of MRI and PET scans had a high accuracy in identifying early Alzheimer disease in dementia research, but saliency maps and SHAP values made the researchers be able to justify that the center of attention of the model was in the clinically relevant area of the brain like the hippocampus and

temporal lobe [54]. With epilepsy, interpretable models that used intracranial EEG data were able to give predictions of seizure onset that could be explained by patterns of spatiotemporally that could be recognized by the neurologists and hence, the predictions became more credible in their clinical usage [55]. Heatmap visualization of predictive stroke models with hybrid AI systems has been enhanced to be more clinically applicable to physicians by enabling the overlaying of heatmaps on imaging data, which enable doctors to identify the regions that make the greatest contribution to ischemic classification [56, 57]. These visualization methods are especially effective in the process of closing the divide between computational output and the cognitive systems of clinicians. The usefulness of XAI can be further seen in the disorder-specific applications. Models that emphasized explainability in the case of Alzheimer disease and other dementias not only enabled the earlier diagnosis of the disease but also aided in the identification of new imaging biomarkers and genetic risk factors, which inform possible targets of treatment [58, 59]. In the case of Parkinson, interpretable systems of gait dynamics, voice records, and wearable sensors data have offered an understanding of what characteristics are most predictive of early motor deterioration and can aid in disease staging. In epilepsy, explainable deep learning seizure prediction models have allowed clinicians to predict important events, as well as understand the neurophysiological origin of alerts, which have made them more acceptable to practice in the real world [24, 60]. XAI has also been useful in stroke prediction where explainable imaging models have enhanced triage decision-making by explaining the nature of lesions and vascularity. In some rare cases of neurological diseases, where machine learning may not work well due to a lack of data, interpretable frameworks have enabled the incorporation of small, heterogeneous data sets without losing clinician trust to transparency [34, 61].

Explainable AI has much more than technical interpretability to offer to the practice of clinical care. Medical adoption is built on trust; with opaque systems, there is no likelihood of clinicians trusting them despite their statistical effectiveness [62]. The ability to explain the logic of predictions leads to the development of trust in AI results and, as a result, can enable the assessment of the use of AI in day-to-day care. Besides, explainability is capable of improving physician decisions by

serving as a second opinion and not a substitute, enabling clinicians to confirm AI results with their experience [63]. Transparent models are also beneficial to patient-centered care because it is possible to explain the information in comprehensible terms to patients and families and use it in shared decision-making and enhance compliance with the treatment plans. Resultatively, explainability can be associated with increasing ethical and legal concerns of transparency, responsibility, and fairness in medical AI [14]. The benefits of integrating AI section in healthcare industry have been emphasized by the FDA, the European Union, and the other agencies similar in healthcare sector that the tools implemented must be capable of providing interpretable results; this is why the XAI is beyond benefit assisting in requirement compliance [64].

It would be better put, though, that the need of trying to seek explainability is not uninhibited. The other question is to determine whether the current XAI practices adequately capture the decision processes, against which the models are meant to make approximations or whether they convey any sensible or even erroneous stories. Moreover, there is the question of the problem of interpretability and predictive performance trade off that is not conclusively decided up till now; the more transparent the model is the more accurate it might be and vice versa [65, 66]. These obstacles must be carried out through a very careful and gentle conduction in logic on medicine, scientific influences and morals. However, the introduction of explainable AI to the neurology field represents a radical solution here, which would yield more open, more trustful, and patient-oriented computational medicine.

## 2. AIMS AND OBJECTIVES

- To critically discuss the role of AI and explainable AI methods in the prediction of neurological disorders.
- To determine the use of XAI in most significant neurological diseases and its clinical decision-making effects.
- To outline major issues, constraints, and future prospects of using XAI in the practice of neurology.

## 3. CHALLENGES AND LIMITATIONS

Nevertheless, explaining AI in the field of neurology continues to be a challenge

even though substantial progress has been made. Among the initial issues is the data privacy and security as the neurological data is sensitive, including imaging, EEG, and genomic data. By sharing and centralizing such data to model-train, confidentiality breaches are prone to occur especially in a multi-institutional partnership. The other problem is generalizability of models, algorithms that are trained on a given sample of a population do not work when applied to a wide range of cohorts, demonstrating biases in training samples. Another weakness is that it is hard to strike a balance between accuracy and interpretability. Traditional interpretable systems are sometimes defeated by deep learning models, but due to their opaque nature, they are not accepted in clinical practice. On the other hand, very transparent models can negatively affect the accuracy of diagnosis. Lastly, clinical integration barriers have not been reduced yet, such as workflow incompatibility, regulatory ambiguity, and the absence of clinical training on AI literacy, all of which are impediments to clinical integration in the real-world neurological setting.

#### 4. FUTURE DIRECTIONS

The research omissions Medsafe has been pointing to must be worked out by their other works in the future through developing some creative alternatives. Privacy preserving and federated learning models have good potentials that can defend sensitive neural data and jointly develop a model. Another boundary branch is multimodal data fusion, in which EEG, MRI and genomic data are joined together to model all of these complexities of neurological disorders. The cross-modal interactions can also be explained with such models and they express new information about the disease processes. Human-in-the-loop strategies will be instrumental in ensuring clarity of the clinical relevance as the physicians will be tasked with fine-tuning the model and testing in the real life the purely theoretical explanations. The other is the fact that clinically validated XAI frameworks appear, which need to be actively tested in prospective trials to substitute proof-of-concept studies. Only after this progress have the explainable AI passing through the experimental novelty phase to a trusted partner in the neuinguistic deal-making department.

## 5. CONCLUSION

Put simply, explainable AI holds an immense potential regarding the response to the development and treatment of neurological issues. By providing a path to realize improved patient-centered care, predictive accuracy, and interpretability, they can result in improved trust, physician decision-making, and predictor of patient outcomes by XAI. However, it must address the issue of data privacy, generalizability and clinical integration in order to enable their widespread application. In the future, the way to go is privacy-preservation, multimodal fusion and responses that are more intimate to clinicians and their AI systems. As these technologies pass through the phase of maturity, the explicable AI can be consequently transformed into a practical tool, however, not before being translated into a transparent and accountable algorithmism, and hence becoming an inseparable element of the procedure that constitutes being a precise neurologist.

## 6. REFERENCES

1. Beghi E. 30-year trends of neurological disorders and impact on health-care systems. *Journal of the Neurological Sciences*. 2021;429.
2. Ding C, Wu Y, Chen X, Chen Y, Wu Z, Lin Z, et al. Global, regional, and national burden and attributable risk factors of neurological disorders: The Global Burden of Disease study 1990–2019. *Frontiers in public health*. 2022;10:952161.
3. Liu Z, Song S-Y. Genomic and transcriptomic approaches advance the diagnosis and prognosis of neurodegenerative diseases. *Genes*. 2025;16(2):135.
4. Domínguez-Fernández C, Egiguren-Ortiz J, Razquin J, Gómez-Galán M, De las Heras-García L, Paredes-Rodríguez E, et al. Review of technological challenges in personalised medicine and early diagnosis of neurodegenerative disorders. *International Journal of Molecular Sciences*. 2023;24(4):3321.
5. Mohsen F, Ali H, El Hajj N, Shah Z. Artificial intelligence-based methods for fusion of electronic health records and imaging data. *Scientific Reports*. 2022;12(1):17981.
6. Amal S, Safarnejad L, Omiye JA, Ghanzouri I, Cabot JH, Ross EG. Use of multi-modal data and machine learning to improve cardiovascular disease care. *Frontiers in cardiovascular medicine*. 2022;9:840262.

7. Sarvamangala D, Kulkarni RV. Convolutional neural networks in medical image understanding: a survey. *Evolutionary intelligence*. 2022;15(1):1-22.
8. Shourie P, Anand V, Upadhyay D, Singh V, Singh S, editors. AI-Powered Alzheimer's Diagnosis from Brain MRI Images Using CNNs. 2024 3rd International Conference for Innovation in Technology (INOCON); 2024: IEEE.
9. Afzal HR, Luo S, Ramadan S, Lechner-Scott J. The emerging role of artificial intelligence in multiple sclerosis imaging. *Multiple Sclerosis Journal*. 2022;28(6):849-58.
10. Poon AI, Sung JJ. Opening the black box of AI-Medicine. *Journal of gastroenterology and hepatology*. 2021;36(3):581-4.
11. Hakkoum H, Abnane I, Idri A, editors. A Systematic Map of Interpretability in Medicine. *HEALTHINF*; 2022.
12. Inukonda J, Tetala VRR, Hallur J. Explainable artificial intelligence (XAI) in healthcare: enhancing transparency and trust. *Int J Multidiscip Res*. 2024;6:30010.
13. Lötsch J, Kringel D, Ultsch A. Explainable artificial intelligence (XAI) in biomedicine: Making AI decisions trustworthy for physicians and patients. *BioMedInformatics*. 2021;2(1):1-17.
14. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Consortium PQ. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC medical informatics and decision making*. 2020;20(1):310.
15. Ratnakar A, Sawant S, Karajagikar J. Explainable AI-driven deep learning for neurological disease diagnosis using MRI: a systematic review and future directions. *Int J Sci Res Arch*. 2025;14:1799-832.
16. Yang CC. Explainable artificial intelligence for predictive modeling in healthcare. *Journal of healthcare informatics research*. 2022;6(2):228-39.
17. Voigtlaender S, Pawelczyk J, Geiger M, Vaio EJ, Karschnia P, Cudkowicz M, et al. Artificial intelligence in neurology: opportunities, challenges, and policy implications. *Journal of neurology*. 2024;271(5):2258-73.
18. Kalani M, Anjankar A. Revolutionizing neurology: the role of artificial intelligence in advancing diagnosis and treatment. *Cureus*. 2024;16(6).

19. Ienca M, Ignatiadis K. Artificial intelligence in clinical neuroscience: methodological and ethical challenges. *AJOB neuroscience*. 2020;11(2):77-87.
20. Rita FY, Arefeen SS, Zakaria RM, Shimanto AH. Advancing the Prediction of Neurological Disorders Through Innovative Machine Learning Methodologies and Clinical Data Analysis. *Journal of Computer Science and Technology Studies*. 2025;7(2):668-80.
21. Pham TM, Pandis N, White IR. Missing data, part 1. Why missing data are a problem. *American Journal of Orthodontics and Dentofacial Orthopedics*. 2022;161(6):888-9.
22. Tasci E, Zhuge Y, Camphausen K, Krauze AV. Bias and class imbalance in oncologic data—towards inclusive and transferrable AI in large scale oncology data sets. *Cancers*. 2022;14(12):2897.
23. Ganeshnath R, Pillai VM, Kunju N, editors. A Class Imbalance Learning Approach to Build an Efficient Machine Learning Model for the Diagnosis of Parkinson's Disease. 2024 1st International Conference on Trends in Engineering Systems and Technologies (ICTEST); 2024: IEEE.
24. Ndlovu B, Maguraushe K, Mabikwa O. Machine Learning and Explainable AI for Parkinson's Disease Prediction: A Systematic Review. *The Indonesian Journal of Computer Science*. 2025;14(2).
25. Magesh PR, Myloth RD, Tom RJ. An explainable machine learning model for early detection of Parkinson's disease using LIME on DaTSCAN imagery. *Computers in biology and medicine*. 2020;126:104041.
26. Ozdemir O, Fatunmbi TO. Explainable AI (XAI) in Healthcare: Bridging the Gap between Accuracy and Interpretability. *Journal of Science, Technology and Engineering Research*. 2024;2(1):32-44.
27. Oetl FC, Oeding JF, Samuelsson K. Explainable artificial intelligence in orthopedic surgery. *Journal of Experimental Orthopaedics*. 2024;11(3):e12103.
28. Shah SP, Heiss JD. Artificial Intelligence as A Complementary Tool for Clinical Decision-Making in Stroke and Epilepsy. *Brain Sciences*. 2024;14(3):228.
29. Ghosh A, Kandasamy D. Interpretable artificial intelligence: why and when. *American Journal of Roentgenology*. 2020;214(5):1137-8.
30. Arjunan G. Implementing Explainable AI in Healthcare: Techniques for

- Interpretable Machine Learning Models in Clinical Decision-Making. International Journal of Scientific Research and Management. 2021;9(5):597-603.
31. Prentzas N, Kakas A, Pattichis CS. Explainable AI applications in the medical domain: A systematic review. arXiv preprint arXiv:230805411. 2023.
32. Thalpage N. Unlocking the black box: Explainable artificial intelligence (XAI) for trust and transparency in ai systems. J Digit Art Humanit. 2023;4(1):31-6.
33. Mishra S. The Age of Explainable AI: Improving trust and transparency in AI models. International Journal of Artificial Intelligence, Data Science, and Machine Learning. 2020;1(4):41-51.
34. Rane N, Choudhary S, Rane J. Explainable artificial intelligence (XAI) in healthcare: interpretable models for clinical decision support. Available at SSRN 4637897. 2023.
35. Sheu R-K, Pardeshi MS. A survey on medical explainable AI (XAI): recent progress, explainability approach, human interaction and scoring system. Sensors. 2022;22(20):8068.
36. Narkhede J, editor Comparative evaluation of post-hoc explainability methods in ai: Lime, shap, and grad-cam. 2024 4th International Conference on Sustainable Expert Systems (ICSES); 2024: IEEE.
37. Sathyan A, Weinberg AI, Cohen K. Interpretable AI for bio-medical applications. Complex engineering systems (Alhambra, Calif). 2022;2(4):18.
38. Salih AM, Raisi-Estabragh Z, Galazzo IB, Radeva P, Petersen SE, Lekadir K, et al. A perspective on explainable artificial intelligence methods: SHAP and LIME. Advanced Intelligent Systems. 2025;7(1):2400304.
39. Kumar A, Veeraiah V, Gongada TN, Ahamad S, Khan H, Gupta A, editors. Explainable Machine Learning Models for Clinical Decision Support Systems. 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT); 2024: IEEE.
40. Muhammad D, Bendeche M. Unveiling the black box: A systematic review of Explainable Artificial Intelligence in medical image analysis. Computational and structural biotechnology journal. 2024;24:542-60.
41. Taiyeb Khosroshahi M, Morsali S, Gharakhanlou S, Motamedi A,

- Hassanbaghlou S, Vahedi H, et al. Explainable artificial intelligence in neuroimaging of Alzheimer's disease. *Diagnostics*. 2025;15(5):612.
42. Bibi N, Courtney J, McGuinness K. Enhancing brain disease diagnosis with XAI: a review of recent studies. *ACM Transactions on Computing for Healthcare*. 2025;6(2):1-35.
43. Lampe L, Niehaus S, Huppertz H-J, Merola A, Reinelt J, Mueller K, et al. Comparative analysis of machine learning algorithms for multi-syndrome classification of neurodegenerative syndromes. *Alzheimer's research & therapy*. 2022;14(1):62.
44. Jiwani N, Gupta K, Sharif MHU, Adhikari N, Afreen N, editors. A LSTM-CNN model for epileptic seizures detection using EEG signal. 2022 2nd International Conference on Emerging Smart Technologies and Applications (eSmarTA); 2022: IEEE.
45. Xu G, Ren T, Chen Y, Che W. A one-dimensional CNN-LSTM model for epileptic seizure recognition using EEG signal analysis. *Frontiers in neuroscience*. 2020;14:578126.
46. Moss L, Corsar D, Shaw M, Piper I, Hawthorne C. Demystifying the black box: the importance of interpretability of predictive models in neurocritical care. *Neurocritical care*. 2022;37(Suppl 2):185-91.
47. Chen P, Dong W, Wang J, Lu X, Kaymak U, Huang Z. Interpretable clinical prediction via attention-based neural network. *BMC Medical Informatics and Decision Making*. 2020;20(Suppl 3):131.
48. Venugopalan J, Tong L, Hassanzadeh HR, Wang MD. Multimodal deep learning models for early detection of Alzheimer's disease stage. *Scientific reports*. 2021;11(1):3254.
49. Woodbright MD, Morshed A, Browne M, Ray B, Moore S. Toward transparent ai for neurological disorders: A feature extraction and relevance analysis framework. *IEEE Access*. 2024;12:37731-43.
50. Mane D, Magar A, Khode O, Koli S, Bhat K, Korade P. Unlocking machine learning model decisions: A comparative analysis of LIME and SHAP for enhanced interpretability. *J Electr Syst*. 2024;20(2):1252-67.
51. Vimbi V, Shaffi N, Mahmud M. Interpreting artificial intelligence models: a

- systematic review on the application of LIME and SHAP in Alzheimer's disease detection. *Brain Informatics*. 2024;11(1):10.
52. Parziale A, Senatore R, Della Cioppa A, Marcelli A. Cartesian genetic programming for diagnosis of Parkinson disease through handwriting analysis: Performance vs. interpretability issues. *Artificial intelligence in medicine*. 2021;111:101984.
53. Sahota A, Roguski A, Jones MW, Rolinski M, Whone A, Santos-Rodriguez R, et al. Interpretable classification of early stage parkinson's disease from eeg. *AI for Health Equity and Fairness: Leveraging AI to Address Social Determinants of Health*: Springer; 2024. p. 219-31.
54. Shad HA, Rahman QA, Asad NB, Bakshi AZ, Mursalin SF, Reza MT, et al., editors. Exploring Alzheimer's disease prediction with XAI in various neural network models. *TENCON 2021-2021 IEEE Region 10 Conference (TENCON)*; 2021: IEEE.
55. Jemal I, Mezghani N, Abou-Abbas L, Mitiche A. An interpretable deep learning classifier for epileptic seizure prediction using EEG data. *IEEE Access*. 2022;10:60141-50.
56. Shafaat O, Bernstock JD, Shafaat A, Yedavalli VS, Elsayed G, Gupta S, et al. Leveraging artificial intelligence in ischemic stroke imaging. *Journal of Neuroradiology*. 2022;49(4):343-51.
57. Vasantha S, Sujiya S, Bharathvarshana R, Shalini P, editors. Deep Learning-Based Models for Predicting and Interpreting Ischemic Stroke with Improved Concentration. *2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*; 2025: IEEE.
58. Khanapur S, Bharadwaj CB, Bhardwaj R, Nayak JS, editors. An Approach for XAI Visualizations for Explainability of Alzheimer's Detection. *2024 2nd International Conference on Networking, Embedded and Wireless Systems (ICNEWS)*; 2024: IEEE.
59. Hasan Saif F, Al-Andoli MN, Bejuri WMYW. Explainable AI for Alzheimer detection: A review of current methods and applications. *Applied Sciences*. 2024;14(22):10121.
60. Sigcha L, Borzì L, Amato F, Rechichi I, Ramos-Romero C, Cárdenas A, et al.

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

Deep learning and wearable sensors for the diagnosis and monitoring of Parkinson's disease: A systematic review. *Expert Systems with Applications*. 2023;229:120541.

61. Gurmesssa DK, Jimma W. A comprehensive evaluation of explainable Artificial Intelligence techniques in stroke diagnosis: A systematic review. *Cogent Engineering*. 2023;10(2):2273088.
62. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. *Journal of biomedical informatics*. 2021;113:103655.
63. Ferrario A, Loi M, editors. How explainability contributes to trust in AI. *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency*; 2022.
64. Gupta N. Explainable AI for Regulatory Compliance in Financial and Healthcare Sectors: A comprehensive review. *ResearchGate*, March. 2025.
65. Raz A, Heinrichs B, Avnoon N, Eyal G, Inbar Y. Prediction and explainability in AI: Striking a new balance? *Big Data & Society*. 2024;11(1):20539517241235871.
66. González-Alday R, García-Cuesta E, Kulikowski CA, Maojo V. A scoping review on the progress, applicability, and future of explainable artificial intelligence in medicine. *Applied Sciences*. 2023;13(19):10778.