Review on Explainable AI by using LIME and SHAP Models for Healthcare Domain

Abujar S. Shaikh
NBNSTIC, Pune, India

Rahul M. Samant
NBNSTIC, Pune, India

Kshitij S. Patil
SCOE, Pune, India

Nilesh R. Patil
SCOE, Pune, India,

Aarya R. Mirkale
NBNSTIC, Pune, India

ABSTRACT
In the dynamic realm of healthcare research and the burgeoning utilization of artificial intelligence (AI), multiple research studies converge to accentuate the immense potential and persistent hurdles facing AI systems. At its core, artificial intelligence seeks to emulate human intelligence, enabling the performance of tasks, pattern recognition, and outcome prediction through the assimilation of data from diverse sources. Its far-reaching applications encompass autonomous driving, e-commerce recommendations, fin-tech, natural language comprehension, and healthcare, with the latter domain undergoing significant transformation. Historically, healthcare leaned heavily on rule-based methodologies rooted in curated medical knowledge. However, the landscape has evolved considerably, with the emergence of machine learning algorithms like deep learning, capable of comprehending intricate interplays within medical data. These algorithms have demonstrated exceptional performance in healthcare applications. Yet, a critical impediment lingers: the enigma of explainability. Despite their prowess, certain AI algorithms struggle to gain full acceptance in practical clinical environments due to their lack of interpretability. In response to this challenge, Explainable Artificial Intelligence (XAI) has risen as a pivotal solution. XAI functions as a conduit for elucidating the inner workings of AI algorithms, shedding light on their decision-making processes, behaviors, and actions. This newfound transparency fosters trust among healthcare professionals, enabling them to judiciously apply predictive models in real-world healthcare scenarios, rather than passively adhering to algorithmic predictions. Nonetheless, the journey toward rendering XAI genuinely effective in clinical settings remains ongoing, a testament to the intricate nature of medical knowledge and the multifaceted challenges it presents. In summation, this research paper underscores the importance of XAI in the domain of healthcare. It emphasizes the necessity for transparency and interpretability to fully harness the potential of AI systems while navigating the intricate landscape of medical practice, thus heralding a transformative era in healthcare research and delivery.

Keywords
Healthcare Domain, LIME, SHAP

1. INTRODUCTION
AI is already having a significant impact on healthcare, with applications in areas such as medical diagnosis [9] [10], image analysis, drug discovery, and patient care [11]. However, the lack of transparency and explainability of AI models is a major barrier to their widespread adoption in healthcare. [12] XAI is a field of research that aims to make AI models more understandable to humans. XAI techniques can be used to paramount explain how an AI model made a particular prediction, or to identify the most important features that influenced the prediction. The Imperative of Explainable AI (XAI) for Trust and Transparency: In numerous applications, the ability to provide an explanation for how an AI-derived answer was reached holds paramount importance for establishing trust and transparency. A prime example of such an application is within the medical field, where the certainty of conclusions is of utmost significance. For instance, in cases involving the analysis of CT scan images, doctors need to understand how AI algorithms arrived at their assessments of whether an individual is suffering from a disease. It is essential to acknowledge that AI-based systems are not infallible, and gaining insight into the decision-making process not only instills trust but can also prevent life-threatening errors. In certain other domains, such as law and order, the need extends beyond the "what" to encompass the "wh" questions—questions pertaining to "why," "when," "where," and more. Conventional AI systems have historically struggled to provide satisfactory responses to these crucial inquiries. This indispensable requirement for explainability has given rise to a burgeoning field of AI research known as Explainable AI (XAI). As illustrated in Figure 1, XAI introduces a transformative dimension to AI, enabling it to address the "wh" questions that have remained elusive within traditional AI. In essence, XAI represents a paradigm shift in the AI landscape, promising enhanced transparency and interpretability in AI-driven decision-making. It not only equips medical professionals with the means to comprehend AI-generated diagnoses but also empowers decision-makers across various domains with insights into the rationale behind AI-derived answers. This newfound transparency not only fosters trust in AI systems but also significantly reduces the risk of critical errors in life-altering situations.

In conclusion, the advent of Explainable AI (XAI) is pivotal for applications where trust, transparency, and the ability to answer complex "wh" questions are paramount. By bridging the gap between AI's decision-making processes and human comprehension, XAI brings forth a new era in AI capabilities, ensuring that AI systems not only provide answers but also elucidate how those answers were reached.

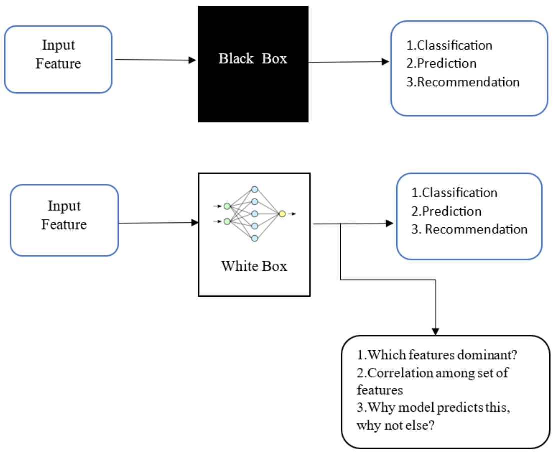

Fig. 1 XAI VS ML

The Significance of Explainability in Intrinsic AI Models: In the realm of artificial intelligence (AI), the concept of "explainability" emerges as a fundamental need and expectation, contributing to the transparency of AI decision-making processes. This imperative not only fosters a rational approach to implementing actions driven by AI but also empowers end-users to comprehend the underlying mechanisms. In rudimentary AI applications, such as symptom-based health diagnosis, achieving explainability is relatively straightforward. However, as the pursuit of human-level accuracy intensifies, researchers and scientists have crafted increasingly intricate algorithms. Notably, the adoption of neural networks and deep learning for decision-making purposes introduces a level of elusiveness and opacity to the decision-making process, rendering it less interpretable. In the pursuit of heightened accuracy and performance, AI models have grown in complexity. Neural networks and deep learning algorithms, renowned for their remarkable predictive capabilities, excel in solving intricate problems but at the cost of transparency. Their intricate web of interconnected nodes and weights makes it challenging to elucidate the reasoning behind their decisions. Despite the allure of heightened accuracy, this lack of transparency can be a double-edged sword. While these advanced AI models may deliver exceptional results, they often do so without providing clear insights into how they arrived at their conclusions. This inherent opacity not only diminishes the user's ability to comprehend the decision-making process but also raises concerns about bias, fairness, and accountability. In light of these challenges, there is a growing imperative to bridge the gap between accuracy and transparency in AI models. The development of methods and techniques to enhance the explainability of complex algorithms, such as neural networks, is becoming a critical area of research. By making the decision-making process more interpretable, we can bolster users' trust in AI systems and ensure that AI-driven actions are not only accurate but also justifiable and accountable. In conclusion, explainability stands as a pivotal factor in the realm of AI, rendering intricate AI models more transparent and understandable. While the pursuit of accuracy drives the development of complex algorithms like neural networks, achieving a balance between accuracy and explainability is essential to maximize the potential benefits of AI while maintaining transparency, fairness, and accountability. The Necessity of Explainable AI: Ensuring Fairness and Legal Accountability: Explainable AI (XAI) has become imperative due to the potential for skewed or biased model outcomes, which can result in wrongful identification of innocent visitors or employees based on certain weighted features present in the training data. Instances of such misidentification have far-reaching legal implications ,particularly in surveillance systems, where transparency is a critical factor before branding someone as a criminal or suspect. Companies responsible for AI-enabled surveillance systems must be prepared to provide justifications in a court of law. In cases where individuals are subjected to public humiliation and search by security forces, profound legal consequences may follow, affecting both the government and airport authorities. These incidents underscore the pressing need for AI systems to be explainable and accountable for their decisions. Such transparency not only safeguards the rights of individuals but also ensures the integrity and legality of surveillance systems. Techniques Accompanied by XAI:Feature-Based Techniques: Our research focuses on feature-based model explainability techniques. You've mentioned several techniques, including Permutation Feature Importance, Partial Dependence Plots
(PDPs), Individual Conditional Expectation (ICE) plots, Accumulated Local Effects (ALE)
Plot, Global surrogate models, LIME, and SHAP.LEDGMENTS

Our thanks to the experts who have contributed towards development of the template.

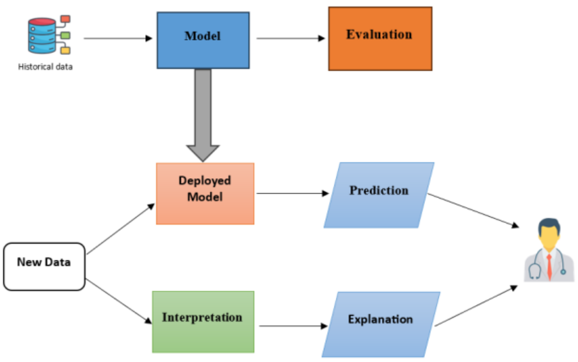

Fig. 2 ML Life-Cycle in Conjunction with XAI

2. LITERATURE SURVEY

[13] is a comprehensive and accessible resource for understanding and implementing interpretable machine learning models. Molnar's work is widely recognized for its depth and clarity, providing a valuable guide for both beginners and experienced practitioners in the field of machine learning.

[14] presents a pioneering and highly influential approach to model interpretability in machine learning. It introduces the Local Interpretable Model-Agnostic Explanations (LIME) framework, which aims to provide explanations for the predictions of any black-box classifier.

[15] dedicated to demystifying deep learning models, which are often considered "black boxes," and providing insights into making these models interpretable and transparent.

[16] this work delves into the challenges, significance, and methods of ensuring that machine learning models are not just accurate, but also interpretable and transparent when deployed in practical settings.

[17] offers a comprehensive overview of the field of Explainable Artificial Intelligence (XAI) with a particular

focus on how XAI intersects with the challenges and opportunities presented by big data, provide a detailed analysis of XAI techniques and their application to large and complex datasets.

[18] with a primary focus on the need for transparency, interpretability, and explainability in AI systems deployed in healthcare settings.

[19] provides an insightful and accessible introduction to the world of explainable deep learning. serves as a valuable guide for those new to the field of deep learning and looking to make complex deep neural networks more transparent and interpretable.

[20] addresses the growing importance of making artificial intelligence (AI) systems transparent and interpretable in the context of Industry 4.0. This transformative era of industrial automation and data exchange demands that AI systems provide clear explanations for their decisions, rather than functioning as "black-box" models.

3. BACKGROUND

LIME, which stands for Local Interpretable Model-Agnostic Explanations, is a technique designed to shed light on the inner workings of machine learning models, particularly complex and opaque ones. It does this by simplified, interpretable models around individual data points to explain their predictions. When you want to understand why a specific prediction was made by your model, LIME comes into play. To do this, LIME selects a specific data instance for explanation and then generates a dataset of similar instances by perturbing its features. It queries the original model for predictions on these perturbed instances and fits a simple linear model to this data, considering the feature changes and corresponding predictions. The resulting linear model helps reveal how each feature contributed to the prediction for that specific instance. LIME's strength lies in its model-agnostic nature, as it can be applied to any machine learning model without requiring knowledge of the model's internal architecture. It provides valuable local, instance-specific explanations, making it a valuable tool for understanding the rationale behind individual predictions in various applications, including healthcare and finance.

3.1 Shapley Additive Explanations (SHAP)

SHAP, or Shapley Additive Explanations, is a sophisticated method for interpreting the predictions of machine learning models. It borrows its name from cooperative game theory, where Shapley values are used to fairly distribute rewards among participants in a collaborative game. In the realm of machine learning, SHAP provides a systematic way to understand the contribution of each feature to model predictions. It accomplishes this by computing the "Shapley values" for each feature, quantifying their impact on predictions.

SHAP's key strength lies in its ability to offer both local and global interpretability. On a local level, it explains why a particular prediction was made for a specific data point by dissecting the contributions of individual features. Globally, it provides an overview of feature importance across the entire dataset, helping to identify which features consistently drive model predictions. This interpretability tool is invaluable in various fields, including healthcare, finance, and recommendation systems, as it empowers users to comprehend and trust complex machine learning models by shedding light on the factors influencing their decisions.

3.2 XAI AND HEALTHCARE

XAI and healthcare is an effective combo of digital technology. In the trend of AI-based diagnosis systems, trust over AI-based conclusions is a matter of serious concern. If-else diagnosis models are inherently explainable because they consist of feature value sets and will assign a score based on a feature value of an instance case of health diagnosis. If-else based explainable medical diagnosis systems are well suited for external symptomatic disease diagnosis. For example, if the patient is already having a past history of respiratory illness and cough then there are higher chances of having asthma. Deep learning is a very important tool for accurate medical diagnosis but its black-box approach for prediction and conclusion makes it restricted for certain critical areas of human medical science.

3.2.1. Explainability methods for XAI-Healthcare

There are two types of methods for an explanation of medical imaging. One method is attribution based and another method is based on perturbation. In the attribution-based methods, one needs to determine the contribution and weight of each feature. LIME is an attribution-based approach for medical image diagnosis. The success of attribution-based explanation is based on the generality of assigned weights for a given prediction or conclusion at the end of the model. Perturbation can be achieved by masking or editing certain input features and observations are recorded as model start training using forward pass and backward pass. In this approach, input features are changed to observe the impact on final prediction at the end of the last layer in the neural networks. This is similar to the sensitivity analysis performed in parametric control system models. This continuous observation makes the XAI practitioner justify different predictions at the end of the neural network.

3.2.2. XAI for Health care applications

Using post hoc analysis we can understand there is a certain amount of overfitting available due to certain features. CNN is a tool used for accurate Image classification. Features-based classification of Alzheimer's using CNN gives robust classification and accuracy. Guided backpropagation (GBP), LRP, and DeepShap are useful for brain imaging and classification.

3.2.3 EXPLAINABILITY & EFFECTIVE CONTESTABILITY

Providing patients with an explanation of their AI generated diagnosis may be guided by different principles and interests. A maximal interpretation of the requirement of explainability would require that an explanation should spell out why the diagnosis was the scientifically best possible explanation of the set of signs, symptoms and indicators. A minimal interpretation would require a statement to the effect that the diagnosis was arrived at by a machine on the basis of health data. But which of these interpretations should guide the explanation of AI driven diagnostics? The GDPR provides some – yet again rather vague – guidance on this issue. Article 22 states that in those cases, where a data subject may legitimately be subjected to automated decision-making including profiling, the data controller should safeguard the data subject's right "to express his or her point of view and to contest the decision". 12 (Although the GDPR is specific to the EU, other legal systems also contain mechanisms by which patients can contest health care decisions). Interpreting the requirement of explainability along these lines, it would imply that an explanation of a diagnosis must allow an individual to contest the diagnosis. The bare right to contest is, however, empty. Just having the right to say 'I disagree with this decision and contest it' does not help the data subject. Only a right to effective contestation is worth having, i.e. a right to contest a decision through a demand for an adequate explanation. What is needed here is a more substantial notion of contestability in relation to AI driven diagnostics. A notion that it is embedded in a wider ethical framework of individual rights and interests in relation to diagnosis. This approach to contestability is distinct from the 'Causability' approach recently developed by Holzinger et al (Holzinger et al 2019, 2020).21,22 Holzinger et al define causality as a relation between an expert user and an AI system where causability is "… the extent to which an explanation of a statement to a user achieves a specified level of causal understanding with effectiveness, efficiency and satisfaction in a specified context of use." (Holzinger et al 2020). This explicitly requires a causal model of the diagnostic reasoning, or in the case of machine learning a mapping from the machine model to a causal model understandable by the human expert. in two ways. We focus on contestability by the patient and on contestability without assuming or defining any specific requirements for explainability in advance. The potential bias of AI diagnostics Second, an individual should be able to contest potential bias in AI diagnostics. AI diagnostics may be biased due to bias in the training data or in the prior human categorization of the training data. 25–28 Bias may lead to discrimination defined as unfair differential treatment if it causes unwarranted differences in diagnostic patterns for particular individuals or groups.29–31 Prior to the clinical deployment of an AI system potential bias in AI decision-making may be tested for by applying the AI model to relevantly different datasets. Prior testing in this way requires a set list of known 'triggers' of discrimination, e.g. gender, age, ethnicity etc. It may also be possible to test for bias in an individual case, e.g. by using a suitably modified set of input data to investigate whether the system provides the same result. Finally, indications of biased system diagnostics may also come from 'counterfactually' testing a system against previous or alternative systems or simply against HCP diagnostics.32 In both cases the testing cannot be exhaustive, but it will clearly be a way of minimizing the risk of harmful and discriminatory bias. Individuals have a right to protect themselves against discrimination, and therefore should be granted a right to contest bias in AI diagnostics. Exercising the right to contest bias requires that individuals have access to information about 1) the character of the dataset on which the model is built, 2) how the data were categorized by humans, and 3) the character and level of testing the AI model has undergone. In some cases where an initial general claim of potentially relevant bias can be made out following disclosure of these three elements, an individual would also have a right to have bias investigated at the individual level.

4. IMPLICATIONS OF CONTESTABILITY

4.1. Contestability and the medical encounter

The right to contest and the correlative duty to provide the information needed for effective contestation does not imply that every patient should be provided with all of the information needed for contestability in all of the four dimensions. Most patients are probably unlikely to want to contest the advice provided by the AI system to the HCP, and will be satisfied with the explanation of their diagnosis provided by the HCP. The right to contest does, however, generate one duty that is relevant whenever an AI system has provided advice, that is the duty to inform that patient that AI advice has been provided and used by the HCP. Contestability does, however, create duties for developers of AI systems and for organizations purchasing and using such systems. Developers and user organizations have to be able to provide all the elements of information outlined above if a patient contests the AI advice. And user organizations have a further duty to train their employees to provide this information to patients and help them understand whether there is a justifiable basis for their contestation of the AI advice.

4.2. Contestability requirements apply to both AI and HCPs

Contestability requirements apply to AI and HCP diagnostics alike. There seems to be no relevant difference between AI and HCP diagnostics that would justify double standards.45 HCPs can also be biased, make mistakes, or not work optimally with colleagues or AI systems. And, HCPs are arguably also 'black boxes'. The exact reasoning of HCPs – every aspect of it – cannot be fully replicated, scrutinized and simulated. Only key factors behind their diagnostics may be reconstructed. A set of contestability requirements for HCP diagnostics is therefore also needed. The contestability requirements cannot, however, be the same. It simply does not make sense to require information about the 'training data' for a HCP. Contestability requirements must reflect how a diagnostic system – whether it be a HCP or an AI system – is trained and processes data in the diagnostic context. In short, contestability requirements must concern types of information that it makes sense to require in relation to a specific diagnostic setup. Developing contestability requirements for HCPs is beyond the scope of this paper.

4.3. Contestability requirements do not impede performance

A key concern in the literature on explainable AI is the potential trade-off between diagnostic performance and explainability. It has been suggested that AI decision-making should be understood as simulatability, i.e. it should be possible for a human to take the input data together with the parameters of the model and in reasonable time step through every calculation required to produce a prediction". A requirement of simulatability would imply that all sufficiently complex AI models, including some of the best performing types of machine learning such as deep learning models, are unexplainable. Hence, requiring explainability as simulatability would be at the cost of performance.

4.4 Drawbacks of XAI

Explainable Artificial Intelligence (XAI) has shown great promise in the healthcare domain by helping healthcare professionals better understand and trust AI-driven decisions. However, it also comes with several drawbacks and challenges: Complexity of Medical Data: Healthcare data is often complex, heterogeneous, and high-dimensional. XAI techniques may struggle to provide comprehensible explanations for models trained on such data, limiting their effectiveness. Trade-off Between Accuracy and Explainability: There is often a trade-off between model accuracy and explainability. Highly accurate models like deep learning neural networks may be less interpretable, making it challenging to strike the right balance between performance and transparency.Black Box Models: Many AI models used in healthcare, such as deep learning models, are inherently black-box in nature, making it difficult to provide meaningful explanations. Techniques like LIME (Local Interpretable Model-agnostic Explanations) or SHAP (SHapley Additive exPlanations) can help, but they may not always yield intuitive or trustworthy explanations. Data Privacy Concerns: Healthcare data is highly sensitive and governed by strict privacy regulations, such as HIPAA in the United States or GDPR in Europe. Generating interpretable explanations while preserving patient privacy can be challenging. High Stakes and Safety: In healthcare, the stakes are often life and death. AI models, even if they provide explanations, need to be thoroughly validated and tested for safety and accuracy. Over Reliance on AI recommendations without critical assessment could lead to serious consequences. Human Bias and Subjectivity: The process of defining what constitutes a good explanation is subjective and can vary among individuals. There is a risk of introducing human bias into the explanation generation process. Scalability: Creating explainable AI models that scale across different healthcare applications, such as diagnostics, treatment recommendation, and drug discovery, can be a significant challenge. What works well for one application may not work as effectively for another. Interoperability: Integrating XAI systems into existing healthcare infrastructure and electronic health record (EHR) systems can be complex. Ensuring that explanations can be easily understood and utilized by healthcare professionals within their existing workflows is crucial. Education and Training: Healthcare professionals need to be trained in understanding and trusting AI-driven decisions. This can be time-consuming and require substantial effort. Cost and Resources: Developing and maintaining XAI systems can be costly and resource-intensive, which can be a barrier for smaller healthcare organizations with limited budgets. Limited Generalization: Some XAI methods may provide explanations that are specific to the dataset they were trained on and may not generalize well to different patient populations or healthcare settings. Dynamic and Evolving Nature of Medicine: Healthcare is a rapidly evolving field, with new research and treatment guidelines emerging constantly. XAI models may struggle to adapt to these changes in real-time. Despite these drawbacks, XAI holds great potential in healthcare for improving diagnostic accuracy, treatment recommendations, and patient outcomes. However, addressing these challenges and continuously refining XAI techniques is essential to ensure safe and effective integration into the healthcare domain.

5. PROPOSED SYSTEM FOR EXPLAINABLE AI (XAI) IN HEALTHCARE

System Overview:

Our proposed XAI system aims to enhance the transparency and interpretability of AI-driven healthcare applications, such as diagnostic decision support and patient monitoring. By integrating explainable AI techniques, the system will provide healthcare professionals, patients, and regulatory authorities with clear and understandable insights into the AI-driven decision-making process.

Components of the System

XAI Algorithms: We will incorporate state-of-the-art XAI methods, such as rule-based systems, model-agnostic techniques, and visual explanations, to make AI models more interpretable.
User Interface: A user-friendly interface will be designed to present XAI-generated explanations in a comprehensible manner, using visuals and natural language descriptions. Data Privacy and Security: Robust data privacy measures will be implemented to ensure compliance with regulations like HIPAA, preserving patient confidentiality. Feedback Loop: The system will include mechanisms for collecting user feedback, which can be used to continuously improve the XAI models.

Positive Points

Enhanced Trust: The system will foster trust among healthcare professionals and patients by providing clear explanations of AI-generated recommendations and decisions. Regulatory Compliance: It will help in meeting stringent regulatory requirements, which often demand transparency and accountability in healthcare AI applications. Improved Decision-Making: Healthcare providers can make more informed decisions by understanding how AI algorithms arrived at specific recommendations, ultimately improving patient care. Ethical Considerations: The system will address ethical concerns, including bias and fairness, by making these issues more transparent and easier to mitigate. User-Friendly: The user interface, designed with natural language explanations and visuals, ensures that healthcare professionals and patients can easily understand the AI's reasoning.

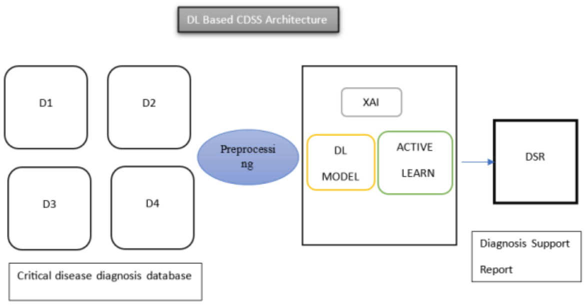

Fig. 3 Proposed Architecture

6. CONCLUSION AND FUTURE SCOPE

The utilization of sharp and lime in our research paper has proven to be an effective strategy for addressing the complexities of Explainable AI in healthcare. By drawing distinct connections between various concepts, we have demonstrated the potential for extending the applications of this technology. This approach not only enhances the clarity of our findings but also highlights the paths for further research and development in the realm of AI-driven disease prediction and explanation. The integration of Explainable AI in healthcare to predict and provide explanations for disease detection is a promising and transformative approach. This technology has the potential to enhance diagnostic accuracy, improve patient outcomes, and build trust among healthcare professionals and patients.However, to fully realize the potential of Explainable AI in healthcare, it is crucial to address ethical and privacy concerns, ensure transparency in the algorithms, and engage in continuous research and development to refine these systems. By doing so, we can harness the full potential of Explainable AI in healthcare, ultimately revolutionizing disease detection and patient care. This interconnected approach not only benefits healthcare professionals and patients but also paves the way for further advancements in the field. As we move forward, we can push the boundaries of what is achievable, ultimately leading to a brighter future for AI-driven disease prediction and explanation in the realm of healthcare.

7. REFERENCES

[1] Prashant Gohel 1, Priyanka Singh 1, And Manoranjan Mohanty 2 ,Explainable Ai: current status and future directions(2021)

[2] Ploug T, Holm S, The four dimensions of contestable AI diagnostics- A patient-centric approach to explainable AI, Artificial Intelligence In Medicine(2020)

[3] Chaddad, A. Peng, J.; Xu, J. Bouridane, A Survey of Explainable AI Techniques in Healthcare. Sensors 2023.

[4] Christopher C. Yang, Explainable Artificial Intelligence for Predictive Modeling in Healthcare(2022)

[5] Devam Dave, Het Naik, Smiti Singhal, and Pankesh Patel, Explainable AI meets Healthcare: A Study onHeart Disease Dataset(2020)

[6] Senthilkumar Mohan , Chandrasegar ThirumalaiI, And Gautam Srivastava. Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques(2019)

[7] Guoguang Rong, Arnaldo Mendez, Elie Bou Assi, Bo Zhao, Mohamad Sawan, Artificial Intelligence in Healthcare: Review and Prediction Case Studies(2020)

[8] Tim Hulsen, Explainable Artificial Intelligence (XAI) in Healthcare(2023)

[9] Samant, Rahul, and Srikantha Rao. "A study on Comparative Performance of SVM Classifier Models with Kernel Functions in Prediction of Hypertension." International Journal of Computer Science and Information Technologies 4.6 (2013): 818-821.

[10] Pradnyesh Kadam Shikha Yadav B Abhishek R. Patel Anusha Vollal and Rahul M Samant. MoodyPlayer: A Mood based Music Player. International Journal of Computer Applications 141(4):21-25, May 2016.

[11] Rahul Samant and Srikantha Rao. Article: Evaluation of Artificial Neural Networks in Prediction of Essential Hypertension. International Journal of Computer Applications 81(12):34-38, November 2013

[12] Rahul Samant, Srikantha Rao, Performance of Alternate Structures of Artificial Neural Networks in Prediction Of Essential Hypertension, International Journal of Advanced Technology & Engineering Research (IJATER)Volume 3, Issue 6, Nov. 2013 ISSN No: 2250-3536 pp:22-27

[13] Christoph Molnar, "Interpretable Machine Learning: A Guide for Making Black Box Models Explainable"(2019)

[14] Marco Tulio Ribeiro, et al., "Why Should I Trust You? Explaining the Predictions of Any Classifier

[15] Rudzicz, A. (2018). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1808.00064.

[16] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608

[17] Zhang, X., Wu, L., & Li, S. (2020). A survey of explainable artificial intelligence (XAI) from a big data perspective. European Journal of Operational Research.

[18] Das, A., & Zhang, L. (2018). Explainable AI for healthcare. Ar Xiv preprint arXiv:1812.10464.

[19] Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep inside convolutional networks: Visualizing image classification models and saliency maps. arXiv preprint arXiv:1312.6034.

[19] Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep inside convolutional networks: Visualizing image classification models and saliency maps. arXiv preprint arXiv:1312.6034.

[20] Andreassen, T. (2021). Explainable AI in Industry 4.0: From Black-Box Models to White-Box Insights and Transparency. In Proceedings of the International Joint Conference on Neural Networks (IJCNN).

[21] R M. Samant, etl The effect of Noise in Automatic Text Classification, Proceedings of the international conference and workshops on emerging trends in technologies, pp 557-558, Feb 2011 https://doi.org/10.1145/1980022.1980142