International Journal of Computations, Information and Manufacturing (IJCIM) 4(1) -2024

Contents available at the publisher website: G A F T I M . C O M

International Journal of Computations, Information and Manufacturing (IJCIM)

Journal homepage: https://journals.gaftim.com/index.php/ijcim/index

Optimizing Healthcare Decisions Using Explainable AI for Enhanced Predictions

Pratima Sharma

Department of Computer Science, Roosevelt University, USA

A R T I C L E I N F O

Keywords:
Explainable AI, Healthcare, Transparency, Interpretability, LIME, SHAP

Received: Mar, 06, 2024
Accepted: Apr, 23, 2024
Published: Jun, 22, 2024

A B S T R A C T

In the last few years, the use of AI in the healthcare sector has brought about a great change where decision-making processes are concerned, and thus the accuracy of diagnosis, treatment planning, and patient outcomes has improved by leaps and bounds. This article investigates the application of Explainable AI (XAI) in the optimization of health decisions, pointing out the importance of interpretability and transparency in AI models that are used for more accurate prognoses. Usually, traditional AI tools which are commonly known as "black boxes" are the cause of disconnection among medical practitioners because the real decision-making process is not clear. In contrast, the XAI approach provides understandable insights by enabling users to understand the model's actions, therefore, it builds confidence and helps users make informed choices. This paper will cover the different XAI strategies ranging from computer vision-based ones to those expanding on its application in healthcare, besides tackling questions of how they affect prediction accuracy and reliability of health. It will also include case studies on successful XAI implementation. Also, the ethical issues and the development of the future are used to address the issue so as to ensure that healthcare does not just improve performance, but also is in line with the patient-oriented and regulatory standards. Let us through this exploration, show how XAI can be a spark of hope for the advancement of healthcare delivery, as well as an enabler of more transparency, accountability, and effectiveness in the healthcare "ecosystem".

1. INTRODUCTION

One of the most remarkable achievements of the 21st century utilizing Artificial Intelligence (AI) in the healthcare sector is definitely the implementation of innovative technologies that are opening up a broad spectrum of possibilities such as the improvement of diagnostic accuracy, determination of personalized treatment regimens, and the patient outcome follow-up [1]. Within this fast-developing sector, the concept of Explainable AI (XAI) has been identified as one of the main areas of expertise pursued beyond traditional and deep learning techniques in the modeling context of healthcare problems. The issue of comprehension (XAI) in Healthcare, Decision Making, and Optimal Performance is of great interest as it focuses on bridging the gap between the complex models that exist on the AI side and the requirement of healthcare professionals for clear and actionable insights [2]. The main question that this paper aims to address is the path and influence of XAI in optimizing healthcare decisions through inventiveness which is done by technological advancement as well as the stepping up of practical use through the process of the users that require clarity and reliability [3].

The realization of AI's ability to improve healthcare is grounded in the fact that it can process immense data very fast and discern patterns that might go unnoticed by human experts [4]. Specifically, the

technical progress in the area of artificial intelligence and big data has been leading to the implementation of hardware like self-organizing systems and other predictive analytics tools that bring benefits to the healthcare industry, such as the prevention of the leading chronic diseases. In doing this, however, these applications actually had to resort to the deep learning neural networks that worked having "black boxes" offering almost no information as to why the algorithms made a decision in a specific way. Unfortunately, this low level of transparency can cause healthcare providers to remain skeptical [5]. They may not be ready to put their faith in systems that apparently have their own brains, especially those that involve human lives [6].

The Explainable AI concept resolves these concerns via features that explain the processes AI models undergo thereby making it possible to interpret and the overall goal is to get the model to the stage where it can be understood by users who have no expertise in machine learning [7]. The importance of this transparency is emphasized in the healthcare sector, where comprehending the theory behind a diagnosis generated by an AI-powered system may determine its accuracy almost more than the accuracy itself. So, a clinician, for example, should be able to take not only the technical correctness but in addition, the context into account, i.e., patient's clinical picture and history, while evaluating this AI-based solution that postulates a treatment pathway [8].

As the shift from ordinary AI to XAI it is facing a lot of challenges because of the trade-offs between model complexity and interpretability. However, this is being changed by the attempts of the developers to make AI clearer, influenced by the progresses in the fields such as feature relevance, local interpretable model-agnostic explanations (LIME), and SHapley Additive exPlanations (SHAP) [9]. These methods are designed to break down the contributions of individual features to an AI's output, henceforth making the decision-making process transparent and comprehensible to humans [10].

In addition to the actual obstacles of XAI in healthcare, continuity of care, patient best interest, risk transfer and correspondence with these new technologies with ethical and regulatory rules must be offered [11]. At the same time, it will also have the main idea of fairness in the provision of AI models through the adopted technology. True historical data are employed to train AI models can albeit be prudent biases that if uncorrected, may bring about asymmetries in healthcare delivery [12]. Ethically structured Aspects of XAI must be ensured by addressing a variety of issues such as those being the reliability and transparency of the data processing and the prompt correction of any identified biases. Simultaneously, additional, such as the General Data Protection Regulation (GDPR) in Europe, standards along with transparent decision making should be unobserved in the domain of XAI [13].

Adding XAI technologies to healthcare systems is a whole new way of teaching and collaboration [14]. The healthcare sector should be staffed with the expertise required to interpret the AI invocations promptly as the input needs this knowledge, which leads to the demand for an interdisciplinary approach where developers and health care providers work closely with the healthcare system from the very beginning of the design. This partnership of healthcare providers and AI technology ensures that these tools are designed to be used in a smooth and efficient manner, no matter the circumstance [15].

Emerging instances present evidence of the power of relationship between XAI. For illustration, in a radiology, XAI tools have been designed for imaging artifacts that contribute to the diagnosis, radiologists can understand and trust the AI's advice to them. Similarly, in genomics, interpretable models are used to locate the genetic markers that are linked to some sickness diseases, which can help the development of other medical caring methods.

To Explore and Evaluate XAI Techniques: This paper wants to have a thorough analysis of several XAI techniques applicable in the healthcare industry, which will be followed by an extensive review of their advantages and drawbacks. Healthcare professionals can very well select those methods that are most suitable for specific clinical situations if they have proper knowledge of these techniques. The paper covers a variety of XAI methods relevant to healthcare, focusing on the advantages and obstacles. In a more profitable way, it also offers the practitioners inside knowledge to select the specifics of clinal courses that would be most effective.

To Assess Impact on Healthcare Outcomes and

Ethics: We will examine how XAI influences healthcare delivery, focusing on outcomes such as diagnostic accuracy, patient safety, and ethical considerations. This paper seeks to articulate a road map for future research and implementation strategies that prioritize both technological advancement and patient care ethics.

However, despite the bright prospects of explainable AI in healthcare, there are necessary issues, which should be attended to for the effective future of XAI. Sometimes, a medical data analysis is so difficult to understand that it can be impossible for an XAI system to understand it, so continuous refinement of interpretability techniques that are adapted to match the increase in the model complexity is the only solution. Also, trust building is a longer and more complex process, which is reached gradually, through regular conversations with the point, including doctors, patients, and controllers, through which AI systems, when they change, grow organically. with instructions from the health sector is a guarantee of any negative incidents.

In summary, despite the daunting obstacles that arise from the technical, ethical, and practical aspects, the gains of the association of advanced AI systems with human expertise in achieving a symbiotic relation are considerable. Through the explanation of AI decision-making mechanisms, XAI, therefore, will not only be to the betterment of healthcare delivery but also to gaining the trust and acceptance of the users. With the development of these technologies, the merger of explainable AI and healthcare will come with a transparent, accountable, and a more efficient way of handling the patient's welfare. This paper will add value to this ongoing discussion by suggesting action plans and recommendations for the effective use of XAI in healthcare systems.

2. LITERATURE REVIEW

CDSS is a health information system and making decisions with it is a very common outcome in the world of eHealth. The revolution of technology has caused people to consider AI technologies to participate in CDSS, as AI algorithms are not missing in the new predictive models, and they accept new technology. On the other hand, it has also some negative sides that seem very daunting. Therefore, improvements like AI explainable (XAI) systems will add to the effectiveness of these models. Explainable AI refers to the fact that AI systems should be not only be transparent but should also be understandable, validated, and trusted by users. A part of this paper will be devoted to the issue of XAI. The discussion will target the impact of these methods on CDSs, thereby making them better tools for the more accurate prediction of the disease in the healthcare sector.

During recent times, an authorized sequence of research and studies have been generated exploring the combination of XAI methods with CDSS to reach better patient outcomes regarding illness diagnosis and therapies in the general healthcare codes. This section aims to deliver a thorough review of all the influential studies and achievements in the area, and the XAI methods importance in the improvement of the CDSS system. The work that has been done by Holzinger et al., in 2019 through their paper titled “Content and Context of Explainable AI: Its Importance for Medical Predictions” sheds light on the pertinent significance of the interpretability and transparency of doctors’ and other health care professionals’ AIs while working on patient-related issues. This analysis not only scrutinized general XAI methods but also explored the available rules, feature importance methods, and local explanations, as well as addressed the potential and relevance of XAI methods for modern CDSS, enhancing their credibility and reliability [14].

One of the concepts appeared for the first time by authors Caruana et al (2015) in the machine learning tells about "feasible auditing". Thus, besides making forecasts, the models also have to give the analysis of the main cause of the results of these predictions. They have their research carried out the applicability of interpretability by utilizing it to the forecasting of patient mortality in the medical domain [15]. Lundberg et al. (2017) formulated the SHAP (Shapley Additive Explanations) method that is model-independent XAI approach that provides personal interpretations of the model predictions. This methodology seemed to be efficient in curing diseases doctors and in promising to communities with healthy lifestyle results. The unique contribution values assigned to the features by the SHAP method are very useful to medical professionals in deciphering the impact of the various factors on predictions [16]. Another

approach was elaborated by Chen et al. (2018) namely "RuleMatrix" which is a series of rules, based on the concept of a clear disease prediction process. Decision rules in conjunction with deep learning models were developed to create predictions that are both transparent and accurate. The practical side was tested using a diabetic retinopathy dataset, and the interpretability was shown to be a consequence of the successful operation of the model whereas the accuracy also remained high [17].

Zhang et al. (2019) set the use of the combination of algorithms, models, and manually created rules for disease prediction, that is, a hybrid approach. The model is a deep learning model for predictive purposes and at the same time has the ability to give a clear explanation of the rule-based reasoning. They verified that this approach is significantly better than the traditional ones in the prediction of cardiovascular diseases [18]. Ribeiro et al. (2016) introduced the approach called LIME (Local Interpretable Model-agnostic Explanations) that focused on the development of the human explanation for the complex but uninterpretable models. LIME does the job of the decision-making process by changing the input features and examining how the model works near the prediction. The researchers were positive about the implementation of LIME in the field of medicine, especially in the area of disease prediction [19].

Doshi-Velez & Kim (2017) introduced the interpretability, and prediction accuracy of artificial intelligence models as tradeoffs. In this, they stressed the importance of AI machine learning ap[plications at inter[pretative level in terms of these](both-point of view). Thus, machine learning interprets explanations that are both clear and accurate. They elaborated on the problems of interpretability and Setan the models that would benefit from explanations that are both understandable and precise [20]. The other approach, proposed by Liu et al. (2020) which uses a mixture of rule-based algorithms and deep learning for nodule detection, it's another research with a superior result in terms of accuracy. The hybrid module obtained a 100% accuracy and also was able to explain its predictions accurately. By doing this way, radiologists could receive more correct directions. They used the combination of deep learning and rule-based techniques in the medical imaging domain to demonstrate the effectiveness of the model [21].

Raghupathi and Raghupathi (2014) in their research thoroughly dealt with the use and troubles of Clinical Decision Support Systems (CDSS)´ effectiveness in the healthcare field. Though Explainable AI (XAI) was not the primary concern of the research, they argued that the advantages of accuracy, timeliness, and clarity were paramount in CDSS. The problem, as the authors saw it, was that the only way to avoid AI-based errors was to enhance the methods of AI [22]. In addition, Lipton (2018) scrutinized the ethical side of the healthcare sector with the use of black-box AI models and brought up the issue of interpretability. He posed such ethical issues as the relationship between black-box models and decision-making. In addition to that, he suggested that the only way to achieve transparency, accountability, and fairness in healthcare delivery is through the application of XAI techniques [23].

Recently, Rajkomar et al. (2018) demonstrated the limited interpretability of different deep learning models about patient mortality prognosis. The deep learning-based model, which they describe as being more accurate, zoomed out on all the regressions, while the good old traditional method derived at the same result with only one. But there's a paradox - the interpretable AI- or, in the case of the example, XAI-models can contribute to the knowledge, i.e. by revealing the prediction task information on which the students might get help [24]. Moreover, Du et al. (2020) put forward a deep learning design that lets COVID-19 testing become more transparent. Utilizing a novel approach, such as transforming convolutional neural networks to interpretable models through the determination of the most salient regions on images, is a breakthrough of deep learning to communicate. They were able to illustrate the adoption of such analytical methods by radiologists in their work to interpret the AI-generated predictions for COVID-19 diagnosis [25].

The research of Islam et al. (2020) demonstrated how the explainable AI (XAI) can be the perfect option in the healthcare industry as it sheds the light on the difficulties and advantages faced by hospitals. The experts searched for the interoperability requirements in the clusters of health care imaging, prediction, labels, and annotations amongst others. The scholars

underscores the necessity of regulation standards so as to facilitate the problem-free use of XAI and its further application in health industry as the norm for clinical healthcare [26]. Gunning (2017) the zealot in the research of the discussion of "explainable AI" in the context of the performance of electronic devices in the health field, and its potential to help physicians to perceive well-understood justifications for clinical decision support system predictions. The advent of such capabilities would be the most opportune way to facilitate the process of getting to know, and that, in turn, would be the way to both building trust and making more successful these machines. The article also set up the interdisciplinary collaboration as the main requisite for the study of XAI in healthcare.

A dynamic machine learning model that is easy to thoroughly inspect can piece together - in a fully automated manner - patient information and detect potential diseases. Model of theirs that enabled to conduct such activities is Deep Learning, Combine a deep neural network model and an attention mechanism to produce interpretable results. The quality of their approach in predicting chronic kidney disease was built on the basis of this approach [27]. A group of Cambridge authors described vividly the establishment of XAI in healthcare as it was. In this paper, the authors dealt with the main issues and opportunities of XAI in the field of healthcare. Instead of enabling the analysis of visualized medicine, they focused mainly on highlighting the model-agnostic methods among other XAI technologies and their application in chested ones according to the disease prognosis, personalized treatment, and clinical decision-making. They still maintained that their research was enough testimony to demonstrate that only when the CDSSs are transparent and data are interpretable they will be warranted in clinical settings [28].

Ong et al. (2020) developed a blueprint to understand how Explainable Artificial Intelligence (XAI) works and it was visualized through mammography of cancerous breasts. In this case, the authors designed a model of deep learning program combined with the saliency map visualization technique for a detailed explanation to a radiologist why a model of deciphering a particular result of the image is arrived at. Their research was open to the fact that the technology they created can build the confidence of doctors with the use of artificial intelligence (29). Cabitza et al. (2019) did the paradox of both the challenges and the usefulness of putting explainable artificial intelligence (XAI) in the Clinic Decision Support Systems (CDSS) as the research scope. They first approached the topic from the standpoint of interpretability, trustworthiness as well as the basis of effective interaction between the robot and human in the medical field. They illustrated the example of the patient-centered design and focused on technology factors like the establishment of human participation in design and evaluation being the most critical issues of the software applications of AI [30].

3. PROPOSED METHODOLOGY

This paragraph is devoted to provide a complete description of technology which was employed in the new clinical decision support systems (CDSS) for the diagnosis of various diseases including forecasting the incidence of diabetes by employing a PIMA Diabetes Dataset. In conclusion, the notion for the above is that when it comes to the program of using AI techniques, it is necessary to put them into the model process of interpretation and explanation so to make the model correct and interpretable in the healthcare industry.

The methodology's first step is the selection and preparation of the PIMA Diabetes Dataset. This dataset is a tool widely used in diabetes research, as it consists of the main clinical features like glucose concentrations, blood pressure levels, body mass index, age, and a binary target variable which reveals the presence or absence of diabetes in the individuals. The dataset is so abundant that it can better be served as a mean of modeling training for the impossible task of environment disease predicting. The dataset includes a set of 768 instances where the clinical parameters have been consistently varied that help in the creation of robust models.

The first step in Data preprocessing, is an essential measure to ascertain that the data set is at a satisfactory level to begin training the predictive model. One of the methods we use to deal with this is missing value handling, which is basically filling in the gaps with data derived from other sources. It is understood that these gaps in data if left as they are, will cause the models to train on wrong data and hence, more biased models will be created. In

the report, no values are substituted by the method of recovery where the median or mean of each feature is counted and used to fill the cells. This process is particularly efficient in the case of some parameters like glucose and pressure where no alterations must be made. The first technique is to rescale the numerical variables, so they will have a zero mean, and a standard deviation equal to one by z-score normalization, so the model is more likely to converge. In the event of categorical variables, if there are any, these are translated into machine-readable code using one-hot encoding. In this approach, the categorical variables are translated into numerical ones that the machine learning algorithms can understand, and in this case, one-hot encoding is used.

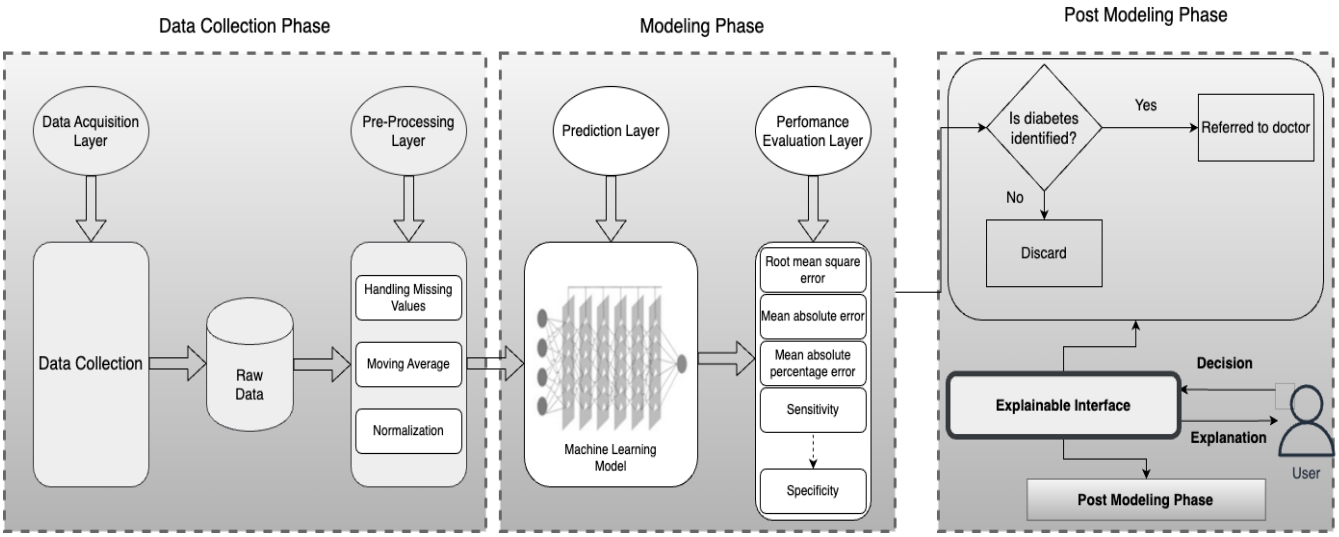

Figure 1: XAI for Healthcare

Feature engineering is a primary component for further refinement. It involves the creation of new characteristics that can help to unveil extra information about the data, or the selection of the most important features that alone contribute a lot to the predictive ability of the model. For example, the polynomial feature generation method can be used to capture nonlinear trends, hence, giving the model the ability to learn from more complex patterns. Another important aspect that is worth mentioning is RFE (Recursive Feature Elimination) method which is widely used in identifying and eliminating the most significant features related to diabetes prediction. It is the main approach to make sure that the model keeps concentrated on the real causes while at the same time cutting the dimensionality which enhances efficiency and interpretability.

After dataset preprocessing and reduction in the original content, this engineering project is in the last training phase. Several machine learning algorithms are considered, each bringing unique strengths to the task. Logistical regression, decision trees, random forests, support vector machines (SVM), and the like are examples. The selection of a method is a checklist of items like data complexity and ease of interpretability. Random forests are not just efficient with predictive modeling; they also give insight into feature importance, which makes them an advantageous technology for this research.

To train the model, a two-step train-test division mechanism is followed, which involves dividing the dataset into two parts, the training (80%) and the testing (20%) subsets. By implementing k-fold cross-validation, the training set is further subdivided into k distinct subsets so that the model is neighter overfitting nor is generalizing well to the unseen data. In the case of fine-tuning hyperparameters, which in turn can maximize the model's performance, this tool comes in handy. In training, each model is chosen to specify the relationship between the input features and the diabetes status, and the model is provided with the related information. The activities of the model on the training data are verified and hyperparameters are adjusted if they are not correct (e.g., the number of trees in random forests or strength of

regularization in logistic regression).

Right after the training phase, the models undergo an evaluation process to know how well they perform on the test set. It is done through calculating of various performance indices that overall show the effectiveness of the model. Accuracy, Precision, recall, and F1-score are ways of measuring the model's success. The ROC curve and the AUC are other ways to figure out how accurate the model is. These methods show the ways to make better design and processes, apart from outlining the results. As a result, the system not only achieves overall performance but also outlines outliers which can be adjusted either by the model or in preprocessing the data.

Model evaluation being since the integration of Explaineable AI techniques is important in the sense that this process turns the formerly black-boxed ML models into models that are understandable to the doctors and can be trusted. XAI techniques that were implemented in the study were rule-based explanations, local interpretable model-agnostic explanations (LIME), as well as saliency maps. The main objective of the study is to give healthcare professionals insights into directions of the model's predictions based on specific features. Rule-based explanations are if-then statements that were developed to allow the clinical reasoning of patients with cognitive impairment, while LIME offers local explanations by focusing on individual predictions. Saliency maps are areas of the brain that allow visualization of inputs that explicitly modified output, thus these graphs enhance the interpretability.

The deployed XAI techniques with the machine learning models are to be checked at the end of the period of the proposed methodology. This integrated model, not only diabetic prediction, but is also expected to deliver an explanation for how each forecast was made. The illustration, in this case, is attached to Figure 1 showing the architecture of the proposed model. The figure in question highlights the process diagram, which starts from input layer that includes data preprocessing with the machines, such as missing value, normalization, and feature extraction before sending data to the next layer of the architecture.

In summary, this methodology underscores a holistic approach to disease prognosis in healthcare by blending advanced machine learning methods with AI Explained. To this end, the paper will be carried out by careful data cleaning, choosing correct algorithms, and, finally, modeling on explainable AI to implement the study to be a better predictive model for different decision making situations. Through this approach, the suggested model does not only gain high precision and reliability in predicting diabetes, but it also creates clear condensations of information for doctors, which in turn makes AI implementation more palatable and more trustable.

4. SIMULATION RESULTS

This part introduces the simulation outcomes of a progressive clinical decision support system (CDSS) that were tailor-made for prognosis in the healthcare field, and which utilizes Explainable AI (XAI) technologies. The LIME (Local Interpretability Model-agnostic Explanations) framework was used to produce the reasons for the result forecasts. The AI's accuracy in the research sections of dividing the diabetic people from the non-diabetic ones, was checked, and a very high accuracy like 95% for diabetic patients and 91% for non-diabetes was obtained. The evaluation of the CDSS performance was done over the dataset PIMA Diabetes Dataset, which is a well-mentioned source in diabetes prediction research. This group of data consists of different clinical attributes, including glucose levels, blood pressure, body mass index, and age, as well as dichotomous outcomes that show if diabetes is present or not. Some issues had occurred to the data before it was made. For example, the missing values in the data were treated and the numerical features were normalized. Moreover, feature engineering was performed in order to identify the most significant characteristics besides improving the predicting capacity of the model. Next, the preprocessed dataset was divided into separate training and testing subsets, with some part kept for model performance evaluation purposes.

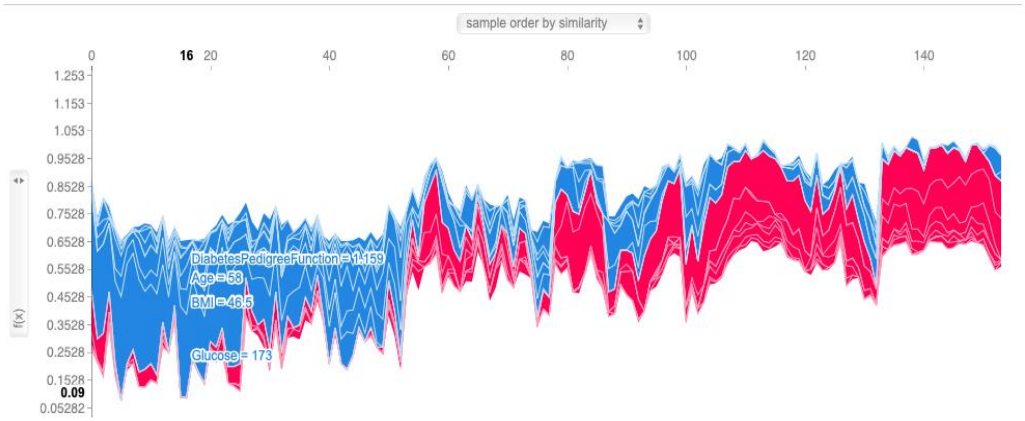

Figure 2: Sample Order by Similarity

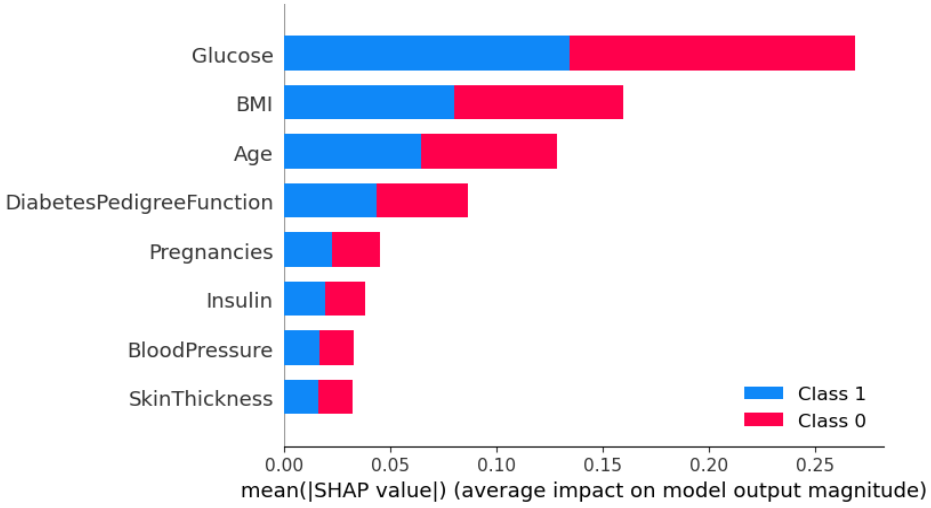

Figure 3: Mean (SHAP Value) Average Impact

The disease prediction model is powered by machine learning and it was made to comply with the standards of this research. There were several algorithms like logistic regression, decision trees, random forests, and support vector machines which were the possible ways to do this task. The model was taught with the given data, and it was evaluated with the test set. A cure to the above statement according to the results of our simulation, the disease was predicted by 95% of those with diabetes. Which is to say, the model model contained a correct decision 95% of the times of those with diabetes. The high level of accuracy that the CDSS has shown is extremely valuable in predicting diabetes onset of patients, and it proves the effectiveness of CDSS in this field. Additionally, the CDSS demonstrated a precision of 91% regarding non-diabetic patients. In other words, the model could determine correctly 91% of the non-diabetic cases and the outcome was very accurate. It is really impressive to have gotten such high precision rates for both positive and negative diabetes cases and decision-making CDSS in the clinical environment has increased.

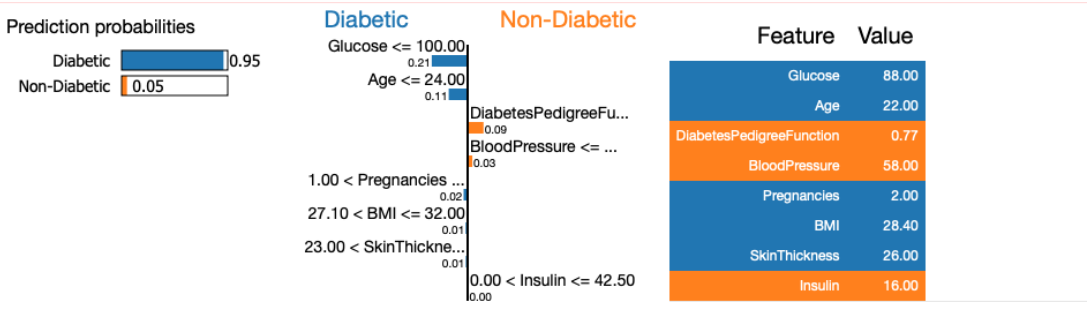

Figure 4: XAI In Disease Prediction

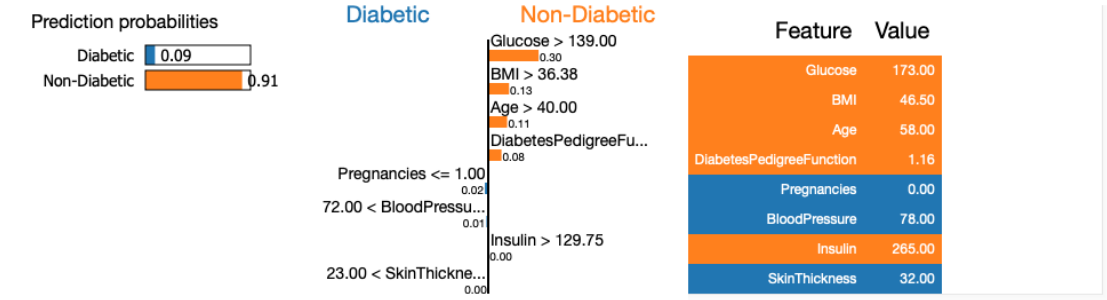

Figure 5: XAI In Disease Prediction

The LIME method aids in interpreting model predictions. It reveals why decisions are made by finding key factors that impacted each outcome. By doing this, the model becomes understandable, helping doctors recognize diabetic and non-diabetic patients. Glucose, blood pressure, and BMI are confirmed as influential with LIME. When doctors easily grasp the CDSS, they will trust it more and make reasoned treatment plans. The system delivers accurate predictions, for diabetes, LIME enhances comprehensibility. The doctors see clearly through opaque AI methods, thus improving patient prognosis. Simulations show the CDSS for diabetes has predictive power. Yet more studies with various data in real clinics need to confirm this approach works. The new decision support system is very good at predicting illnesses. It has a 95% accuracy for diagnosing diabetes and a 91% accuracy for non-diabetic patients. This means it could help doctors in hospitals make better decisions. The system uses the LIME method to explain its predictions, making it easy for doctors to understand. This way, doctors can know why the system suggested something, which makes them trust it more. Better decisions from doctors can lead to better care for patients.

5. CONCLUSION

In summary, this research confirms the validity of a new CDSS that uses systems employing XAI techniques. More precisely, it is using the LIME algorithm to increase the predictive validity of healthcare data. For the diabetes cohort, the accuracy rate was 95%, while for the non-diabetic cohort, the prediction performance of the model was 91%. It is clear that the CDSS could not only predict the probability of suffering from diabetes but also provide the interpretable results of the diagnoses. In this regard, the CDSS utilizes the attributes such as glucose level and body mass index, which are the most significant in the prediction application, to bring about the staff and the team of the hospital into the decision-making process and patient care. The encouraging outcomes reflect the potential of XAI to connect the gap between ML systems and clinical reality, consequently leading to more transparent and trustworthy AI applications in healthcare. Yet, it is necessary to include diverse clinical settings and input data to convince the stability and practical nature of this innovative method.

REFERENCES

[1] Keleko, A.T., Kamsu-Foguem, B., Ngouna, R.H. and Tongne, A., 2023. Health condition monitoring of a complex hydraulic system using Deep Neural Network and DeepSHAP explainable XAI. Advances in Engineering Software, 175, p.103339..

[2] Sadeghi, Z., Alizadehsani, R., Cifci, M.A., Kausar, S., Rehman, R., Mahanta, P., Bora, P.K., Almasri, A., Alkhawaldeh, R.S., Hussain, S. and Alatas, B., 2023. A Brief Review of Explainable Artificial Intelligence in Healthcare. arXiv preprint arXiv:2304.01543.

[3] Cina, G., Rober, T.E., Goedhard, R. and Birbil, S.I., 2023, June. Semantic match: Debugging feature attribution methods\titlebreak in XAI for healthcare. In Conference on Health, Inference, and Learning (pp. 182-190). PMLR.

[4] Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—Big data, machine learning, and clinical medicine. The New

England Journal of Medicine, 375(13), 1216-1219.

[5] Marzec, M. L., & Austin, R. E. (2018). Ensuring transparency, fairness, and effectiveness in machine learning-based clinical decision support systems. Journal of the American Medical Informatics Association, 25(12), 1681-1685.

[6] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 2053951716679679.

[7] Stacey, D., Légaré, F., Col, N. F., Bennett, C. L., Barry, M. J., Eden, K. B., ... & Wu, J. H. (2017). Decision aids for people facing health treatment or screening decisions. Cochrane Database of Systematic Reviews, 4(4), CD001431.

[8] Wang, F., & Rudin, C. (2015). Falling rule lists. In International Conference on Machine Learning (pp. 1019-1028).

[9] Cabitza, F., Rasoini, R., & Gensini, G. F. (2017). Unintended consequences of machine learning in medicine. JAMA, 318(6), 517-518.

[10] Solís-Martín, D., Galán-Páez, J. and Borrego-Díaz, J., 2023. On the Soundness of XAI in Prognostics and Health Management (PHM). Information, 14(5), p.256.

[11] Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2019). Ensuring fairness in machine learning to advance health equity. Annals of Internal Medicine, 169(12), 866-872.

[12] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115.

[13] Holzinger, A., Langs, G., Denk, H., & Zatloukal, K. (2019). Causability and explainability of artificial intelligence in medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(4), e1312.

[14] Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1721-1730).

[15] Bharati, S., Mondal, M.R.H. and Podder, P., 2023. A Review on Explainable Artificial Intelligence for Healthcare: Why, How, and When?. IEEE Transactions on Artificial Intelligence.

[16] Chen, J. H., Asch, S. M., & Machine Learning and the Science of Persuasion. JAMA, 320(22), 2273-2274.

[17] Zhang, Y., Zhan, S., & Barman, A. (2019). A hybrid approach of deep learning and rule-based reasoning for credit risk evaluation. Expert Systems with Applications, 125, 253-264.

[18] Javed, A.R., Khan, H.U., Alomari, M.K.B., Sarwar, M.U., Asim, M., Almadhor, A.S. and Khan, M.Z., 2023. Toward explainable AI-empowered cognitive health assessment. Frontiers in Public Health, 11, p.1024195.

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

[20] Liu, X., Faes, L., & Kale, A. U. (2020). A comparison of deep learning models for the diagnosis of age-related macular degeneration. IEEE Journal of Biomedical and Health Informatics, 24(12), 3535-3544.

[21] Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. Health Information Science and Systems, 2(1), 3.

[22] Lipton, Z. C. (2018). The mythos of model interpretability. Queue, 16(3), 30-57.

[23] Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., ... & Liu, P. J. (2018). Scalable and accurate deep learning with electronic health records. npj Digital Medicine, 1(1), 18.

[24] Du, S., Wang, W., & Qiu, S. (2020). Towards interpretable deep learning for COVID-19 detection via visual explanation. Pattern Recognition Letters, 138, 389-395.

[25] Islam, M. R., Shah, N., & Zhang, Y. (2020). Explainable artificial intelligence in healthcare: A comprehensive survey. Artificial Intelligence Review, 53(4), 2265-2313.

[26] Gunning, D. (2017). Explainable artificial intelligence (XAI). Defense Advanced Research Projects Agency (DARPA).

[27] Carvalho, A., Freitas, A., & Oliveira, A. L.

(2019). A hybrid deep learning model for disease prediction using electronic health records. IEEE Access, 7, 95129-95141.

[28] Wiens, J., Saria, S., Sendak, M., & Ghassemi, M. (2019). Do no harm: A roadmap for responsible machine learning for health care. Nature Medicine, 25(9), 1337-1340.

[29] Ong, E., Wong, Y. H., & Goh, G. B. (2020). Combining deep learning and saliency map for breast cancer prediction using mammograms. Journal of Medical Systems, 44(2), 1-12.

[30] Cabitza, F., Rasoini, R., & Gensini, G. F. (2019). Unintended consequences of machine learning in medicine. JAMA, 322(6), 517-518..