# Explainable Artificial Intelligence (XAI) for the Prediction of Diabetes Management: An Ensemble Approach

Rita Ganguly[1], Dharmpal Singh[2]
Department of Computer Applications, Dr. B. C. Roy Engineering College,
Fulljhore, Durgapur, West Bengal 713206, India[1]
Department of Computer Science and Engineering, JIS University,
81, Nilgunj Road, Agarpara, Kolkata, West Bengal 700109, India[2]

***Abstract*—Machine learning determines patterns from data to expedite the process of decision making. Fact-based decisions and data-driven decisions are specified by the industry specialist. Due to the continuous growth of machine language models in healthcare, they are breeding continuous complexity and black boxes in ML models. To make the ML model crystal clear and authentically explainable, AI accession came in prevalence. This research scrutinizes the explainable AI and capabilities in the Indian healthcare system to detect diabetes. LIME and SHAP are two libraries and packages that are used to implement explainable AI. The intimated base amalgamates the local and global interpretable methods, which enhances the crystallinity of the complex model and obtains intuition into the equity from the complex model. Moreover, the obtained intuition could also boost clinical data scientists to plan a more felicitous composition of computer-aided diagnosis. Importance of XAI to forecast stubborn disease. In this case, of stubborn diabetes, the correlation between plasma versus insulin, age versus pregnancies, class (diabetic and nondiabetic) versus plasma glucose persisted with a strong relationship. The PIDD (PIMA Indian Diabetic Data set) with the SHAP value is used for concise dependency, and LIME is applicable when anchors and importance of features are both required simultaneously. Dependency plots help physicians visualize independent relationships with predicted disease. To identify dependencies of different attributes, a correlation heatmap is used. From an academic perspective, XAI is very indispensable to mature in the near future. To estimate the presentation of other applicable data set correspondence studies are very much apprenticed.**

***Keywords—Explainable Artificial Intelligence (XAI); diabetes; interpretability; machine learning; chronic disease management***

## I. Introduction

In developing countries such as India, one of the top ten causes of death related to the disease is diabetes. The researchers are working on the project to find the earliest way to detect the disease at a very early stage. Type 1 and type 2 are the two main types of diabetes. Type 1 diabetes causes a shortage of insulin, which affects increases in blood sugar levels. Whenever humans eat any food that is broken down into glucose, blood sugar in the blood stream increases [1]. The pancreas cell emancipates insulin as the revolt of blood sugar levels, and as a result, it furnishes us to zeal for every day's task [2]. If the cell ceases to produce insulin in the body, it may cause extravagant blood sugar to be sustained in the bloodstream. Significant health issues, such as blurred vision, kidney disease and heart disease, may be sustained as a result of the excessive presence of blood sugar. The symptoms of diabetes are delayed healing, itching, weaknesses, stiffness of muscle, polydipsia and blurred vision [3]. Millions of deaths are caused by diabetes, which is a metabolic condition throughout the year with various health issues. Approximately 84 million to 238 million diabetic cases will be diagnosed within 2030 around the world, which will impose a consequence load on the health care system [4, 5].

Glucose level changes are the main reason for diabetes. Maintaining a healthy lifestyle, a balanced diet and regular medical check-ups are the most preventive ways to restrict diabetes. Based on laboratory tests and disease symptoms, an intelligent system might be an important tool for diabetes prevention and detection.

Patient trust is the main dispute of an AI system. Without any explanation and without any reason to provide the output of the system is an opaque AI system. Especially in the health care system, the machine provides the output without any explanation, which makes it very difficult for the patient to believe the machine properly. To overcome this situation, explainable AI is rapidly used today to diagnose the disease fairly and correctly without any errors. Artificial intelligence adds a layer where the output can be defined clearly, which is known as explainable AI.

Interpretability as well as transparency can be escalated through XAI in the medical field. There is no legal right, which is reflected in the literature. XAI enhances the exposure of service by obliging end users to rely on the decision that AI makes correct decisions. The scope of AI is to make faster and correct decisions regarding the patients' diagnosis and treatment and make it most trustworthy. The main goal of XAI is to generate a high-performance level explainable model. XAI will provide a detailed description of the AI technique, which helps to make AI algorithms more illuminate and translucent by depending on diagnosis with treatment protocols, drug research, medicine patterns, tests, etc.

Addressing XAI, the two important features are interpretability and explainability. The interpretability defines how the output will be changed when the corresponding input parameters are changed. The term explainability defines a clear conception about a modern model, including certain assumptions made by the model and why.

This research aims to furnish interpretability of ML models and boost the performance of prediction to detect diabetes. The contributions of this work are as follows:

- Explainable AI now a day provides us with a "Black Box" nature, which releases more transparency and accuracy in the detection of disease and heightens the confidence of users.
- Shapley Additive Explanations (SHAP) and Local Interpretability Model-Agnostic Explainable (LIME) have been used widely in decision processes.
- A detailed analysis in the SHAP and LIME framework has been used to divulge significant perceptions to make a complete decision process considering all risk factors.

The remaining elements of the research discussion are organized as follows: briefs of similar papers are in Section II that use different categorizations of XAI technologies to forecast chronic disease. Section III describes the methodology used. The implementation of XAI techniques discussed in Section IV and in Section V describes the XAI upgrading of the health care system. The research effort is concluded in Section VI.

## II. Literature Review

To diagnose diabetes, various types of machine learning algorithms have been used in [6,7-12]. In [9], PIDD missing values are replaced by mean values. A decision tree classifier is used for the classification. For the same data set, the researcher use different classification algorithms in [10], and the support vector machine provided the best result compared to other classification algorithms.

In another work with the PIDD data set [11], the models were developed using three different machine learning algorithms where SVM provided 70% accuracy with an accuracy of 80 of the trained data set. The same data set provides 78.9% accuracy using theXGBoost algorithm [8] by Tiwari and Singh.

The imbalanced classes and lack of preprocessing in the data set resulted in poor accuracy in these studies [13, 14, 15, 16]. In Fitriyaniet al. [17], these constraints are overcome.

In [18], the researcher developed a commission model amalgamation model to detect heart disease, which is an improved version of [6]. In [19], training set a split for detecting diabetes that scored high accuracy.

For several reasons [20], academic research findings and fruitful applications in medical practice have a remarkable distance. The most recent methodologies and techniques are not reliable by the physician [21, 22]. Black box methodology is not accepted by most medical practitioners due to its lack of explanation [23]. With the sacrifice of higher accuracy [24], clinical ML keeps away from complex models [25-27].

Due to the lack of explainability and biased accuracy, diabetes detection [28-30] and progression have become challenging. This research provides interpretability of the machine learning model and boosts the prediction performance using cross-validation, which helps to predict disease more reliably, accurately and effectively.

## III. Methodology

Fig. 1 demonstrates the workflow of the overall model. PIMA Indian Diabetic Data set (PIDD), which contains eight attributes with values that the system has made use of. Then, the data are reprocessed to remove any missing or void values. Next, for training validation and testing, the data set was separated. For the implementation of separation, random sampling was used. As a result, the training and testing division will be unbalanced. To eliminate this problem, stratified sampling was used with a training size of 20% and a training validation size of 80%. After that, different machine learning models were put into action using the Scikit-Learn package. Next, the results were evaluated and explained with the LIME package and SHAP tool to make a complete decision process considering all risk factors. Table I describes the statistical description of the PIMA database.

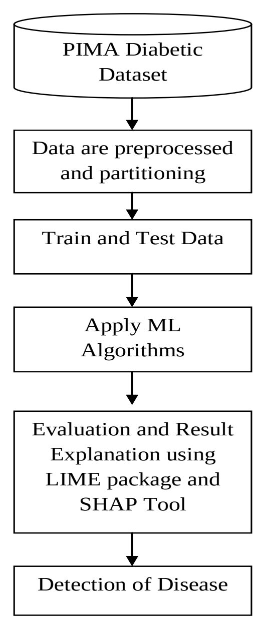


Fig. 1. Workflow of the overall diabetes model.

TABLE II. STATISTICAL DESCRIPTION OF DATABASE

| | Preg | Plas | Pres | Skin | Insu | Mass | Pedi | Age | Class |
|---|---|---|---|---|---|---|---|---|---|
| **Count** | 768.00 | 768.00 | 768.00 | 768.00 | 768.00 | 768.00 | 768.00 | 768.00 | 768.00 |
| **Mean** | 3.8451 | 120.8945 | 69.1055 | 20.5365 | 79.7995 | 31.9926 | 0.4719 | 33.24 | 0.3490 |
| **Std** | 3.3696 | 31.9726 | 19.3558 | 15.9522 | 115.2440 | 7.8842 | 0.3313 | 11.7602 | 0.4770 |
| **Min** | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0780 | 21.0000 | 0.0000 |
| **25%** | 1.00 | 99.0000 | 62.0000 | 0.0000 | 0.0000 | 27.3000 | 0.2438 | 24.0000 | 0.0000 |
| **50%** | 3.00 | 117.0000 | 72.0000 | 23.0000 | 30.5000 | 32.0000 | 0.3725 | 29.0000 | 0.0000 |
| **75%** | 6.00 | 140.2500 | 80.0000 | 32.0000 | 127.2500 | 36.6000 | 0.6263 | 41.0000 | 1.0000 |
| **Max** | 17.00 | 199.0000 | 122.0000 | 99.0000 | 846.0000 | 67.1000 | 2.4200 | 81.0000 | 1.0000 |

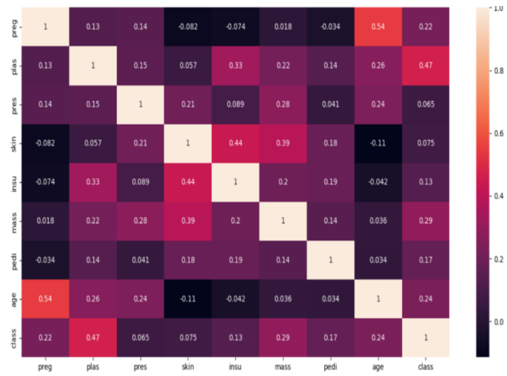

Fig. 2. Correlation coefficient matrix of diabetes.

All diabetes feature correlation coefficient matrices are displayed in Fig. 2. The calculation is based on the linear relationship of measures and features; -1 to + 1, the correlation coefficient value ranged. Values closer to 0 indicate a weak relationship, and values higher than 0 indicate a strong relationship. The correlations between plasma versus insulin, age versus pregnancy, and class (diabetic and nondiabetic) versus plasma glucose were strong.

## A. *Data Description*

An auspicious disease is diabetes with its barrier. Mentioning the present health care system, it is very important to detect the disease properly. Table II explains the dataset description of proposed machine learning system. The present PIDD(https://www.kaggle.com/datasets/kumargh/pimaindians diabetescsv) is a tabular data set containing 768 data points with two classes (diabetic and nondiabetic). In AI-based solutions in the area of healthcare, XAI is a censorious tool for predicting disease more accurately. The personalized description of XAI solves constraints such as the proper interpretations of model features, performances, explanations of the data set, and knowledge of the model with associations of trained data.

TABLE III. DATA SET DESCRIPTION OF PROPOSED MACHINE LEARNING SYSTEM

| Attributes Abbreviation | Attributes | Attributes Types |
|---|---|---|
| Preg | Pregnancies | int 64 |
| Plas | Plasma Glucose | int 64 |
| Pres | Diastolic Blood Pressure | int 64 |
| Skin | Triceps Skin Fold Tickness | int 64 |
| Insu | Insulin(2Hrs serum insulin) | int 64 |
| Mass | Body Mass Index | int 64 |
| Pedi | Diabetic Pedigree Function | int 64 |
| Age | Person's Age | int 64 |

The characteristics of the XAI technique can be explained by LIME, SHAP and personal dependency plots with the diabetic data set in Fig. 3. The scatter plot in Fig. 4 represents the classifications of 768 patients into two classes, diabetes, which is class 1, and nondiabetes, which is class 0, according to the two important features plasma and insulin. It can be inferred that most of the patients belong to Class 0 nondiabetes. From this plot, low plasma and high insulin are less likely to be predicted as diabetes.

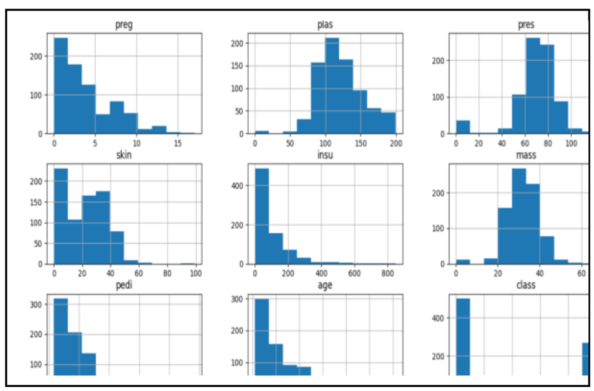


Fig. 3. Histogram of data set features.

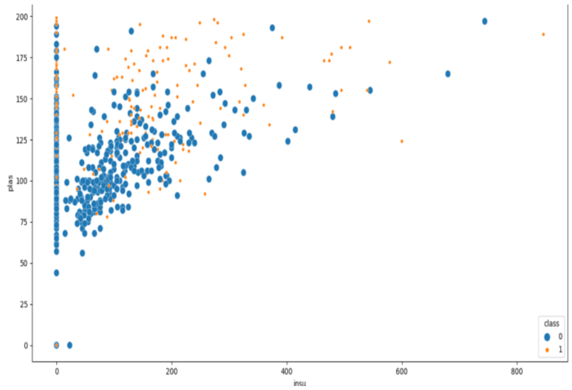

Fig. 4. Scatter plot of two classes.

## IV. Implementation of XAI Techniques

XAI process accomplishment helps researchers understand the system properly and expedites the process of diagnosing chronic disease properly with supportive results.

### A. Platform and Language

XAI techniques can be easily implemented by using the Python programming language, which is an object-oriented, asynchronous interpretive high-level programming language with lower maintenance cost. ML with AI is very much in the Python language for the following reasons:

- Substantial libraries

Actualization of AI algorithms with ML is very difficult to implement, cumbersome and time-consuming. Python libraries and frameworks are used in programming and reduce the complexity of the program, which makes it simpler and more flexible.

- Compact and Elementary

Python acknowledges the programmers to develop a solution that is trustworthy instead of highlighting the practical.

### B. SHAP (SHapley Additive Explanations)

In XAI techniques, SHAP is a very functional tool that can be used to evaluate the precise value of an attribute in forecasting. The SHAP values represent the difference between the actual forecasting of the result and the average forecasting of the model. The altar in the forecasting of the predicted model is represented by the SHAP values for each and every attribute when forcing the condition on a feature.

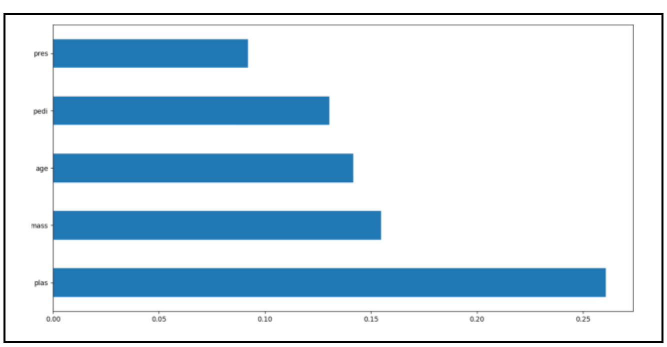

Fig. 5. Global variable importance.

In Fig. 5, it is clearly observed that the plasma glucose variable has the highest influence in the model, followed by mass and age and Fig. 6, describes the SHAP values with all attributes with two classes.

Fig. 7 clearly depicts how individual attributes subscribe to the average or standard model forecasting on a global level. The y-axis on the right side represents whether the relevant value of the feature is low or high. The instances of data are represented by each dot sign.

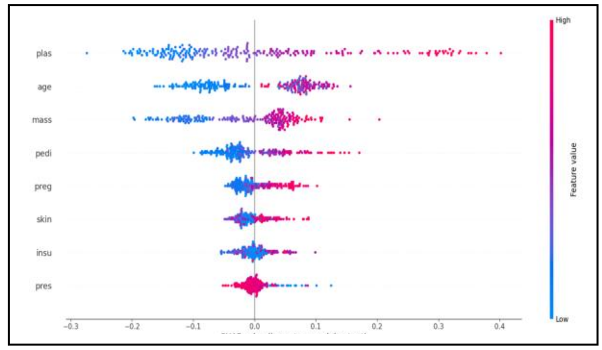

Fig. 6. Mean SHAP value.

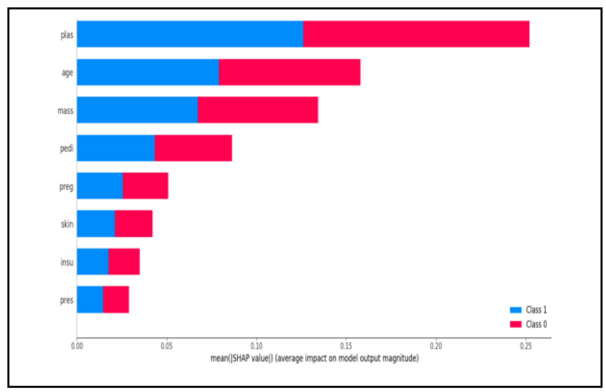

Fig. 7. SHAP value of the model.

### C. LIME (Local Interpretable Model-Agnostic Explanations)

What ML model is currently being used by the programmer that will be explained by a tool is known as LIME. LIME endeavours to embrace the model by making differences in input data and scrutinizing the revised prediction in altered input data. The questionnaires below are the key ingredients of LIME.

- Which attribute steers to predict a particular specimen of data and changes the entire output as an alteration?
- Why did this anticipation take place?

The Fig. 8 describes here the class names that tested positive and tested negative and fit the explainer on the training data set using the lime tabular explainer and perform the explanation on the 8th instance in the test data, Fig. 9. The output describes the local LIME model intercept of 0.3580, and LIME model prediction is 0.3960 (Prediction local). The original random forest model prediction is 0.3051 and with the prediction explanations of intercept 0.3644 and prediction local is 0.3892.

```
Intercept 0.35801780662587024
Prediction_local [0.39604192]
Right: 0.30517363636363637
```

Fig. 8. Local interpretable model-agnostic for class diabetes +ve.

```
Intercept 0.3644016123406618
Prediction_local [0.38925742]
Right: 0.30517363636363637
```

Fig. 9. Prediction explanation.

## V. USING XAI UPGRADATION OF HEALTH CARE SYSTEM

- Conventional Health Care System

This healthcare system provides care and proper planning for the treatment of chronic diseases. This system provides a plan for proper care and treatment and accomplishes the goal of boosting the care of patients. It may consist of several important components as follows:

Information system based on clinical data

Organizational health system

Treatment system design

Right decision support system

Self-decorum action

Resource management of the community

- Obstacles with Traditional Healthcare Systems

It may consist of several important components -

Mistakes and lapses in clinical reports

Financial obstacles of patients

Lack of self-caring patients

Insufficient knowledge and a lack of proper training of health workers

- Usage of Explainability

XAI may be implemented in the following ways with auspicious technology for the management of chronic health care.

By using ML models, we can minimize errors and improve the accuracy precision.

Chronic healthcare management can be properly managed by a decision support system, as all the components are explainable.

Explainable AI optimizes the cost and provides relevant inferences and predictions with proper explanations for accurate results.

Explainable AI also suggests hypothetical information that notifies about the applicable alteration in a feature for accuracy.

Therefore, it is necessary, such as in India, to build an XAI system in hospitals for the proper diagnosis of chronic diseases such as diabetes.

## VI. CONCLUSION

Importance of XAI is that XAI helps to predict the diagnosis of chronic disease. In the case of diabetic pregnancies, plasma glucose and age have been identified as the most important features based on the Pima Indian Diabetes Data set (PIDD), and the SHAP value is used for concise dependencies and visualizations of trends. LIME is applicable when anchors and importance of features are both required simultaneously. A correlation heatmap helps to identify dependencies of different attributes. Dependency plots help doctors visualize independent relationships with predicted classes.

The present work explains the utility of the XAI technique with only 768 limited clinical data. In future research wishing to work with more clinical data in different ways, XAI may be used for automatic diagnosis.

From an academic perspective, XAI works are far from the authentic medical framework, which is essential for development in the near future. Correspondence studies are necessary to estimate the presentation of recommended proposals in other applicable data sets. To a greater extent, refinement in the accomplishment of the model and explainability will endeavor using various algorithms in machine learning to furnish different types of accumulation models.

## REFERENCES

[1] V.Garc´ıa, L Aznarte. Shapley additive explanations for no2 forecasting. Ecological Informatics, 2020:56.

[2] A. Dhurandhar, Y. Chen, R. Luss, C Tu, P Ting, K Shanmugam, P. Das Explanations based on the missing: Towards contrastive explanations with pertinent negatives. Advances in Neural Information Processing Systems, 2018;592–603.

[3] A.V Looveren, J Klaise. Interpretable counterfactual explanations guided by prototypes. 2020.

[4] S.M Lundberg, S.I Lee. A unified approach to interpreting the model predictions. Advances in neural information processing systems, 2017: 4765–4774.

[5] F Jiang,Y. Jiang, H Zhi, Y.Dong,H Li, S Ma, Y Wang, Q Dong, H Shen, Y Wang. Artificial intelligence in healthcare: past, present and future. Stroke and Vascular Neurology, 2017; 2 (4): 230–243.

[6] U Pawar,DO'shea, S. Rea, R. O'Reilly. Explainable AI in healthcare. Reasonable Explainability for Regulating AI in Health. June 2020.

[7] H. B Kibria; A. Matin; N Jahan; S.Isla A Comparative Study with Different Machine Learning Algorithms for Diabetes Disease Prediction. In Proceedings of the 2021 18th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), Mexico City, Mexico, 10–12 November 2021; pp. 1–8. [CrossRef].

[8] P.Tiwari; V. Singh, Diabetes disease prediction using significant attribute selection and classification approach. J. Phys. Conf. Ser. 2021, 1714, 012013. [CrossRef].

[9] V.Chang; J.Bailey; Q.A.Xu; Z.Sun, Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. Neural Comput. Appl. 2022, 1–17. [CrossRef] [PubMed].

[10] W.Chen; S.Chen; H.Zhang; T.Wu, A hybrid prediction model for type 2 diabetes using K-means and decision tree. In Proceedings of the 2017 8thIEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 24–26 November 2017; pp. 386–390. [CrossRef].

[11] A.Mir; S.N. Dhage, Diabetes Disease Prediction Using Machine Learning on Big Data of Healthcare. In Proceedings of the 2018 Fourth International Conference on Computing Communication Control and

Automation (ICCUBEA), Pune, India, 16–18 August 2018; pp. 1–6. [CrossRef].

[12] T.Sangien; T.Bhat; M.S.Khan, Diabetes Disease Prediction Using Classification Algorithms. In Internet of Things and Its Applications; Springer: Singapore, 2022; pp. 185–197. [CrossRef].

[13] J.Ramesh; R.Aburukba; A.A. Sagahyroon,remote healthcare monitoring framework for diabetes prediction using machine learning. Health Technol. Lett. 2021, 8, 45–57. [CrossRef] [PubMed].

[14] S. Kumari; D.Kumar; M.Mittal, An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. Int. J. Cogn. Comput. Eng. 2021, 2, 40–46. [CrossRef].

[15] U.Ahmed; G.F. Issa; M.A Khan;S. Aftab; R.A.T. Said; T.M. Ghazal, M. Ahmad, Prediction of Diabetes Empowered With Fused Machine Learning. IEEE Access 2022, 10, 8529–8538. [CrossRef].

[16] J. Abdollahi; B.Nouri-Moghaddam,Hybrid stacked ensemble combined with genetic algorithms for diabetes prediction. Iran J. Comput. Sci. 2022, 1–16. [CrossRef].

[17] N.L Fitriyani; M.Syafrudin; G.Alfian; J.Rhee,Development of Disease Prediction Model Based on Ensemble Learning Approach for Diabetes and Hypertension. IEEE Access 2019, 7, 144777–144789. [CrossRef].

[18] H.B Kibria; A.Matin, An Efficient Machine Learning-Based Decision-Level Fusion Model to Predict Cardiovascular Disease. In International Conference on Intelligent Computing & Optimization; Springer: Berlin/Heidelberg, Germany, 2021; pp. 1097–1110. [CrossRef].

[19] M.S Santos; J.P Soares; P. Abreu; H.Araujo; J. Santos, Cross-Validation for Imbalanced Data sets: Avoiding Overoptimistic and Overfitting Approaches [Research Frontier]. IEEE Comput. Intell. Mag. 2018, 13, 59–76. [CrossRef].

[20] M. Bucholc ; X.Ding; H.Wang; D.H. Glass; H.Wang; G.Prasad; L.P Maguire; A.J Bjourson; P.L McClean; S.Todd; A. et al,A practical computerized decision support system for predicting the severity of Alzheimer's disease of an individual. Expert Syst. Appl. 2019, 130, 157– 171. [CrossRef].

[21] D.Das; J.Ito; T. Kadowaki; K. Tsuda, An interpretable machine learning model for diagnosis of Alzheimer's disease. PeerJ 2019, 7, e6543. [CrossRef].

[22] J. Burrell, How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data Soc. 2016, 3. [CrossRef].

[23] A.Adadi; M.Berrada, Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access 2018, 6, 52138–52160. [CrossRef].

[24] M.S.Lundberg; B. Nair; M.S. Vavilala; M. Horibe; M.J. Eisses; T. Adams; D.E. Liston; D.K.-W Low; S.-F.Newman; J. Kim; et al. Explainable machinelearning predictions for the prevention of hypoxaemia during surgery. Nat. Biomed. Eng. 2018, 2, 749–760. [CrossRef] [PubMed].

[25] M.S. Kamal, A. Northcote, L. Chowdhury, N. Dey, R.G.Crespo & E. Herrera-E.Viedma, (2021). Alzheimer's patient analysis using image and gene expression data and explainable-AI to present associated genes.IEEE Transactions on Instrumentation and Measurement, 70, 1-7.

[26] M.S. Kamal, N. Dey, L. Chowdhury,S.I. Hasan & K.C. Santosh, (2022). Explainable AI for Glaucoma Prediction Analysis to Understand Risk Factors in Treatment Planning.IEEE Transactions on Instrumentation and Measurement, 71, 1-9.

[27] S. Majumder & N. Dey, (2022). Explainable Artificial Intelligence (XAI) for Knowledge Management (KM). In AI-empowered Knowledge Management (pp. 101-104). Springer, Singapore.

[28] S.E. Zohora, S, Chakraborty, A.M. Khan & N. Dey, (2016, March). Detection of exudates in diabetic retinopathy: a review. In 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) (pp. 2063-2068). IEEE.

[29] L. Cao, N. Dey, A.S. Ashour, S. Fong, R.S. Sherratt,L. Wu & F. Shi (2020). Diabetic plantar pressure analysis using image fusion.Multimedia Tools and Applications, 79(15), 11213-11236.

[30] M. Gospodinov, E. Gospodinova, I. Domuschiev, N.Dey & A. Ashour, (2016). Nonlinear analysis of heart rate variability in type 2 diabetic patients. Fractal Geom Nonlinear Anal Med Biol, 1(14), 0-0.