Explainable Artificial Intelligence (XAI) for Population Health Management – An Appraisal

Het Naik, Priyanka Goradia, Vomini Desai, Yukta Desai, and Muralikrishna Iyyanki

Abstract — This study explores Explainable Artificial Intelligence (XAI) in general and then talked about its potential use for the India Healthcare system. It also demonstrated some XAI techniques on a diabetes dataset with an aim to show practical implementation and implore the readers to think about more application areas. However, there are certain limitations of the technology which are highlighted along with the future scope in the discussion.

Keywords — Artificial Intelligence (AI), Chronic disease management, Decision Support System, Explainable Artificial Intelligence (XAI).

I. POPULATION HEALTH MANAGEMENT

Population health management is progressing as a valuable approach to improve health and promote the well-being of people. This approach creates scope for people to see organized healthcare interventions including those with chronic conditions such as diabetes, asthma etc. It will promote mental and physical health awareness. The country is currently dealing with a triple burden of disease, including an unfinished agenda of infectious diseases, the issue of noncommunicable diseases (NCDs), which are connected to lifestyle changes, and the introduction of novel pathogens that cause epidemics and pandemics. Furthermore, the health-care infrastructure is currently overburdened and needs to be upgraded to meet these problems in the future and population health management is identified as an alternate feasible and viable method. Population health management is being designed to facilitate stratification of patients with precision by leveraging health insights and data analytics. Thus, PHM at its best is scoping to prioritize health care management interventions by integrating tools of data analytics and cloud infrastructure to match low to high-risk patients, to clinical treatment based on their need. This facilitates control of chronic diseases with appropriate data analytics to model the diseases and prescribe interventions as well as treatment.

A. Current Healthcare Scenario in India

The rise of chronic diseases has seen rapid growth in India. Due to diabetes, the total disability-adjusted life years have increased from 1990 to 2016, by 170%. Between 1998 and 2016, many of the people especially women having less access to education were found to suffer more from obesity. Fee-for-service structures, a scarcity of skilled health workers with knowledge of non-communicable disease prevention, a lack of coordination within multidisciplinary care teams, irregular review and follow-ups, limited engagement between patients and health-care providers, and underdeveloped decision support systems are all factors that have hampered the optimal delivery of chronic care. Given these restrictions, the immediate focus should be possible interventions for changing the health system by integrating appropriate methods including technological actions to provide quality health care. The quality of healthcare systems depends on availability and costs of medical professionals and tools. Some of the related issues identified are unequal distribution of skilled human resources, low public spending on healthcare by government, irrational use and spiraling cost of drugs, weak governance and accountability, delay in diagnosis of Illnesses, high out-of-pocket expenditure and not having equal accessibility to all categories of people. Thus, there is a definite need to ramp up the healthcare systems to combat the majority of the issues. Most of these issues can be better combated with the aid of AI/ML technologies.

B. AI and Explainable AI (XAI)

There are established public health management methods and policies for communicable diseases. Any task of disease diagnosis, prescription and or modeling requires a large amount of data for real time monitoring of health risk alerts and health predictions. In this process, AI in general has been found to be effective in extracting useful information from the large patient population data. The traditional goals of AI research include classification, prediction, reasoning, knowledge representation etc. In fact, the benefits of AI in healthcare have been extensively discussed in medical literature. AI can learn features from a large volume of population health data using sophisticated algorithms and then use the obtained results to aid in clinical practice where-ever healthcare is necessary. There has been advancement in technology in the medical sectors and Artificial Intelligence and Machine learning seems to be the most promising field. Automated analysis of medical information has proved to be beneficial for doctors as well as patients. Despite all these advantages, after reviewing the concerns and the necessities

Submitter

Het Naik, Pandit Deendayal Energy University, Gandhinagar, Gujarat, India.
(e-mail: het.nce18sot.pdpu.ac.in)

Priyanka Goradia, Pandit Deendayal Energy University, Gandhinagar, Gujarat, India.
(e-mail: priyanka.gce18sot.pdpu.ac.in)

Vomini Desai, Pandit Deendayal Energy University, Gandhinagar, Gujarat, India.
(e-mail: vomini.dict18sot.pdpu.ac.in)

Yukta Desai, Pandit Deendayal Energy University, Gandhinagar, Gujarat, India.
(e-mail: yukta.dict18sot.pdpu.ac.in)

Muralikrishna Iyyanki, Former Director R&D JNTU, UC Berkeley - India Academic Programs Director, India.
(e-mail: iyyankigm@gmail.com)

DOI: http://dx.doi.org/10.24018/ejce.2021.5.6.368

Vol 5 | Issue 6 | December 2021
related to the handling of health data with specific reference
to the transparency, it has been identified that Explainable AI
which is a type of AI, is more suitable for this health data
analytics. XAI is an added layer of artificial intelligence (AI)
where the output can be understood better. We can increase
transparency as well as interpretability in medical fields
through Explainable AI. The laymen can also interpret it. In
healthcare, the trust between medical experts and patients is
crucial to ensure a successful implementation of AI. The
existing literature shows that even if there is no legal right or
regulatory need, XAI may improve the user experience of a
product or service by assisting end users in trusting that the
AI is making sound judgments. Inside the healthcare domain,
the moral difficulty of transparency related to AI and the
dearth of consideration inside the black-box operation of AI
structures leaves unexplainable models. XAI tries to address
this issue. The goal of XAI is to develop a set of approaches
that generate more explainable models while retaining high
performance levels. XAI has scope to aid in faster decision
making for diagnosis & treatment of disease and make the
predictions more trustworthy.

XAI has scope to play a key role in the sustainable
integration of AI into medicine. An important role of XAI
will be to provide the transparency required by physicians to
trust AI systems. The major goal of health related XAI
applications is to investigate the links between preventative
or treatment methods and patient outcomes.

Drug research, treatment protocol creation, Diagnoses,
personalized medicine, and patient monitoring and care are
the major areas in the healthcare domain where AI systems
are currently being used. XAI will provide explanations for
the AI techniques used, making the Machine learning
algorithms transparent and elucidated.

For addressing these issues, XAI based system can be
proposed where additional layers are added to the testing and
production phase. These additional layers provide
interpretabiltiy and explainability. Interpretation helps in
understanding the cause and effect in AI systems. It helps to
understand how the output changes when a change is made in
input or algorithmic parameters. Explainability on the other
hand helps in understanding why a model predicted certain
predictions.

Figures 2 and 3 have examples of the outputs from the AI
and XAI model respectively for a diabetes dataset.

It can be seen from Figure 3 by an XAI model using
Anchors (discussed in the further sections). This explains
additional information as to why the outputs generated are
more understandable and interpretable.

II. OBJECTIVES OF THE STUDY

The main objectives of the study are:
1. To develop a decision support system to address clinical
diagnosis and treatment measures for chronic diseases using
XAI.
2. To use different XAI techniques, identify features in the
chronic disease dataset which have a greater impact on the
likelihood of a person being diagnosed with it.
3. To facilitate better decision making using XAI.
4. Impact of feature variations on the probability of
diagnosing the patient with the disease and how it affects the
severity of the disease after being diagnosed with it.
5. Impact of feature variations on the probability that the
patient has the disease and how it impacts the severity of the
disease post diagnosis.
6. To identify XAI techniques for specific chronic disease
feature-based analysis.
7. And finally, to make AI predictions more reliable for
chronic diseases.

III. UNDERSTANDING AND APPLICATION OF XAI

A. Scope

The current research aims to use XAI - Explainable AI
techniques as already stated for development of a healthcare
information and decision support system applicable to
chronic diseases.
1. What has been done.
2. What is being done now.
3. What will be done next.
4. To reveal the knowledge on which the actions are based.
These would enable the following:
1. The confirmation of current knowledge.
2. The challenge of existing knowledge.
3. The generation of new assumptions.

In the context of the present analysis, the following
benefits of XAI are observed:
1. When an explanation is needed to make informed
decisions and fairness is paramount (However this needs
further support with more data).
2. XAI can address issues in a more rigorous manner,
particularly decisions which are far reaching.
3. When the cost of a mistake is high leading to unwanted financial costs, uplifted health risks, and personal trauma (such as malignant tumor misclassification).
4. When a new hypothesis is put together by AI systems, it must be verified by subject matter experts.
5. For compliance purposes which promises the “right to an explanation” when user data goes through the automated systems.

B. Features of the Proposed System

This study highlights the entire process of application of XAI for one chronic disease i.e., diabetes, which is taken up for a detailed analysis of XAI application. The application of XAI techniques can be extended to various diseases by following similar methods of data processing. Techniques like Feature Importance and Permutation Importance are used to study the importance of the parameters or features of a particular disease by a doctor. Partial Dependence plots (PDP), Individual Conditional Expectation (ICE), and Accumulated Local Effects (ALE) will help in the effect of change of a feature on the model, when all the other features are constant. This will help in analyzing the type of relation that a feature has with the disease, i.e., a linear or curvilinear. Questions like why a particular modification was made and what particular features have led to the following predictions are answered by the techniques LIME and SHAP. Techniques such as Counterfactuals will provide the possible changes required in the parameters to achieve the goal, i.e., to treat a particular disease. Anchors help in providing explanations for the particular outcome, whose features have led to the following outcome. The Contrastive Explanation Method (CEM) provides insights upon the unwanted and preferable features as the model is explained. The following techniques will decrease the time, thus providing relevant and fast treatment to patients. It will assist doctors in analyzing the numerous parameters that must be evaluated for a certain ailment, as well as important disease forecasts, so that doctors may deliver the best possible treatment to the patient. The following are the salient features of the proposed decision support system which comprise three core elements:

1. Robust Data Platform — A data platform for gathering data from various different sources is must as the predictive models need one source of truth across the system for required accuracy. This will also help in preventing data silos and in ensuring that the machine learning models are using the most accurate and latest data.

2. Expertise in Predictive Analytics and Knowledge of the Institution — Understanding institution-specific and case specific knowledge along with an understanding of healthcare systems are as important as having a strong grasp on the predictive analysis that the model learning models use. Granular understanding of a healthcare institute’s procedures and operations can assist the project manager in tailoring the system to their unique demands in order to optimize its worth.

3. Technology of Machine Learning — The platform on which the predictive model is built can be inhouse or popular cloud services. A data pipeline is critical for connecting the data source and the predictive analysis platforms for seamless flow between the two, resulting in faster findings with no delays. Specific hardware requirements for machine learning models: A GPU cluster for multi-GPU computing or a cloud service subscription which provides infrastructure and software as a service.

C. Description of the Data

Diabetes is an opportune disease with complications and there is a large wealth of data available. Its diagnosis needs to be more accurate looking at the present healthcare system. The present dataset is a tabular dataset which belongs to 1000 patients and covers three classes (Diabetic, Non-Diabetic, and Predicted-Diabetic). 70% of the available dataset has been used for training purposes and the remaining data has been used for testing. XAI is a critical tool for assisting medical practitioners in successfully understanding and evaluating AI-based solutions in the healthcare area. The subjective nature of XAI in the medical setting presents challenges such as associating a trained model’s knowledge with medical features, providing explanations for the underlying medical dataset, and understanding how the presence or absence of some medical features’ information affects a model’s performance and feature interpretation.

The data was collected from the Iraqi society as claimed to be acquired from their Laboratory of Medical City Hospital and the Specialized Center for Endocrinology. Patients’ data was extracted to construct the diabetes dataset. Medical information and laboratory analysis make up the data. The data that have been entered initially into the system has the following attributes/features:

1. Age;
2. Gender;
3. Body Mass Index (BMI);
4. Creatinine ratio (Cr);
5. Urea ;
6. Cholesterol (Chol);
7. LDL;
8. VLDL;
9. Triglycerides (TG);
10. HDL Cholesterol;
11. HbA1c;
12. Class (the patient’s diabetes disease class may be Diabetic, Non-Diabetic, or Predict-Diabetic).

Some of the crucial characteristics, such as blood pressure, were not considered in the dataset which could contribute to the doctors’ ability to make a confident diagnosis. These factors may have an impact on the present predictions.

The feature based XAI techniques such as Shapley, LIME, Partial Dependence Plots etc. are applied for data treatment and analysis which are further explained under the techniques implemented in the diabetes dataset. The example based XAI techniques such as Anchors and Counterfactuals are used for the demonstration purpose. A ML model using XGBoost is developed. XGBoost is used to construct gradient-boosted decision trees that are optimized for speed and performance. The present model gives an accuracy of 98%.

The scatter plot in Fig. 4 represents the classification of 1000 patients into three classes: Non-diabetic (Class 0), Diabetic (Class 1), and Predict Diabetic (Class 2) according to the two most important features: BMI and HbA1c. From the plot, it can be inferred that most of the patients belong to class 1, i.e., they are diabetic. We can see from the plot that
those with a low BMI and HbA1c are less likely to be diagnosed with diabetes.

Fig. 4. Scatter plot of three classes.

Fig. 5. Histogram for dataset features.

IV. XAI TECHNIQUES IMPLEMENTATION

Implementation of the given XAI techniques below will help the reader understand that a decision support system that helps in clinical diagnosis of chronic diseases can be developed with the support of results from XAI techniques as peruse-case.

A. Language and Platform Used:

Python is used as the programming language for implementation of XAI techniques. Python is an object-oriented high-level programming language with a dynamically semantic, interpreted syntax. Python's concise, easy-to-learn syntax emphasizes readability, which lowers software maintenance costs. Modules and packages supported by Python foster program modularity and code reuse. Python makes the best programming language for AI and Machine Learning applications due to two reasons:
Machine learning and AI are based on complex algorithms and flexible workflows, while Python's simplicity allows developers to design dependable solutions. Instead of focusing on the technical subtleties of the language, developers can devote all their attention to solving an ML problem.

2) Extensive libraries and Frameworks

Implementing AI and machine learning algorithms can be difficult and time-consuming. Python frameworks and libraries are used by programmers to reduce development time. Python has many libraries for artificial intelligence and machine learning because of its robust technology stack. Use of Google Colab for implementation provides cloud-based CPU and GPU computational power as per the requirement for training and testing machine learning models and frameworks.

B. Techniques Implemented on Diabetes Dataset

1) Histogram

Histograms help us in getting a count of the no. of observations with a particular feature value. They help identify the nature of data by showing us patterns in the frequency plots.

For example, in Fig. 5, the no. of instances having BMI value 30 is the highest. However, the number of occurrences with different BMI values is not insignificant, implying that the data is roughly evenly distributed over all BMI values. There is not a significant difference in the number of cases with a high BMI value vs those with a low BMI value.

2) Feature Importance

Feature importance indicates how significant a feature is to the model. A feature is “essential” if modifying its values increases model error, because the model relies on the feature for prediction in this instance. A feature is considered “unimportant” if rearranging its values results in the same model error since the feature was disregarded by the model during prediction.

Fig. 6 shows the importance of features according to our dataset. HbA1c is the most influencing factor followed by BMI, age and so on. It can also be predicted for various instances which in turn can be more interpretable and useful.

Fig. 7 shows an instance where a person is diabetic, and the feature contributions are at non-diabetes, diabetes, and predicted diabetes. Here it can be seen that a feature like HbA1c is a feature that is negatively contributing to pushing the model towards non-diabetic and positively pushing the model to predict the model towards diabetic.

3) Permutation Importance

Permutation importance gives the possible variations that can happen according to the dataset. BMI will have 0.0083 positive or negative variation where gender will not have any variation because it can be either male or female. As seen from the graphs in Fig. 9, class 0 is affected by BMI negatively, class 1 is affected positively and class 2 negatively. In all the cases, after a certain point in the graphs, further increase in the BMI value has no impact on any of the diabetes classes. Fig. 10 shows a cluster representation of the same graphs. The yellow highlighted lines are the final plots.

2) PDP Interact

PDP interact shows how two features are related to each other, or how they interact with each other. By studying the heat maps that represent these PDP interactions, the nature of dependency of two features can be found out.

In Fig. 11, the graphs show the relationship between BMI and HbA1c for all the three diabetes classes. With the understanding that 1 means linear dependence and 0 means no dependence at all, for classes 0 and 2 the two features have no dependence for most values. However, for class 1, they have linear dependence for most of the values. This tells us that BMI and HbA1c values affect each other linearly and positively for class 1.
Fig. 7. Feature Importance.

PDP for feature "BMI"

Number of unique grid points: 10

Fig. 9. Partial Dependency Plots.
Fig. 10. Cluster representation of Partial Dependency Plots.

ICE Plot of BMI on Diabetes Class

Fig. 12. ICE Plot.
PDP interact for "HbA1c" and "BMI"

Number of unique grid points: (HbA1c: 10, BMI: 10)

Fig. 11. PDP Interact Maps.

Fig. 13. Local Interpretable Model-Agnostic Expectations Code and Output.
4) Local Interpretable Model-Agnostic Explanations

LIME is a tool for explaining what machine learning models predict or evaluate. LIME attempts to comprehend the model by altering the input data and examining the changes in predictions as a result of the change in input data. LIME addresses questions like why this prediction was made or which factors lead to such forecasts by changing a single instance of data and monitoring and analyzing the changes in the output as a result of the modifications.

The left portion of Fig. 13 displays the model's prediction probabilities for a specific case for the classes:

0: Non-Diabetic (0%);
1: Diabetic (100%);
2: Predict Diabetic (0%).

The middle portion here gives the weights of the features which impacts the prediction probabilities. For instance, if the BMI (body mass index) value, in the range of 0 to 26 contributes towards not having diabetes (it may be predicted though). The right portion indicates the actual value of a particular feature for a specific local instance.

5) SHapley Additive exPlanations

The SHAP value is a useful tool, like LIME, in which interpretations measure the impact of a certain value of a feature on the prediction.

The change in the predicted model prediction when imposing conditions on that feature is represented by the SHAP values for each feature. The SHAP value represents the contribution of each feature to explaining the difference between the average model prediction and the actual prediction of the instance.

This is a highly useful view in Fig. 14 that displays how each characteristic contributes to the average model prediction on a global level. On the right side, the y-axis indicates whether the respective value of a feature is low or high. Because each dot represents one data instance, a cluster of dots shows that the SHAP value is present in a large number of instances in the data.

6) ALE

Accumulated local effects show how features impact a machine learning model's prediction on average. ALE plots are a quicker and more accurate alternative to partial dependency graphs (PDPs). However, if the features are heavily connected, PDP may include implausible feature value combinations in the average prediction computation due to independent manipulation of feature values across the marginal distribution. When features have a significant connection, trust in the PDP explanation is diminished.

ALE, as opposed to PDP, controls feature correlations by averaging and aggregating the difference in predictions throughout the conditional distribution, isolating the effects of the particular feature. This comes at the cost of needing a larger number of observations and a nearly uniform distribution of those observations in order to appropriately compute the conditional distribution. The resultant ALE explanation is concentrated on the feature's mean effect, so that the major feature effect is compared to the data's average forecast.

As seen in Fig. 15 graphs, the dependence of individual features with respect to the outcome. As an instance, the ALE plot for BMI shows that the prediction is being pushed to a person being diabetic when the BMI is more than around 26.

7) Anchors

Anchors help to detect faulty models and thus help in making machine learning more productively viable, because there are less chances that a faulty model will get deployed. They help minimize the gap between risks associated with trusting machine learning in health care and the opportunities created by it. Anchors are very precise explanations in the form of human readable IF-THEN rules that help in describing which values caused the model's outcome to take a particular value for one specific instance. All in all, they make the system’s decisions more trustworthy by providing a clear coverage as they state exactly to what all instances they apply.

Fig. 16 depicts the Anchor method being applied on an instance where it predicts that the person is non-diabetic. A single anchor, HbA1c > 6.50, is generated which shows that it is sufficient from this feature to conclude that the person is non-diabetic.
8) Counterfactuals

The contradictory or counterfactual description of the event or situation Y is presented as “If X had not appeared, Y would not have occurred” (Descriptive Machine Translation). In the context of machine learning classification, X represents the pattern of interest and Y represents the predicted label of the model.

The purpose of explaining a is to point out any X that resembles the initial state of the X but leads to a different prediction of Y. Humans are very good at thinking about contradictory words, like asking what should have been done differently to get different results. For this reason, using counterfactual models in machine learning predictions is a potent way to generate comprehensible meanings for humans.

From Fig. 17, the prediction of the person changed from being diabetic (1) to being non-diabetic (0) by slight modifications in the features like BMI, TG and HbA1c which can be done by physical and medical techniques.

Person is Non diabetic
Anchor generated feature(s) ['HbA1c > 6.50']

Original prediction: 1
Counterfactual prediction: 0
Change HbA1c from 6.8 to 5.250233436169516
Change TG from 2.5 to 2.2413257497111942
Change BMI from 32.0 to 25.38997659853263

C. Identified Limitations

The temporary identified limitations related to the availability of data and the tools used for the study are as follows.

1. Lack of large-scale clinical data.
2. Many features such as blood pressure which contribute to a person’s diabetic condition haven’t been considered due to the limited scope of data.
3. Data sources are case studies that focus on specific cohorts of patients (Although this is an inherent risk of sampling, it raises the likelihood of us catching a patient population that is not typical of the wider population.).
4. Explainability is a less accessible alternative for inventors due to opportunity costs such as poorer accuracy and breach of trade secrets.

V. IMPROVING THE CHRONIC HEALTHCARE MODEL USING XAI

A. Traditional Chronic Healthcare Model

The chronic care model is a commonly used paradigm for planning and providing care for persons suffering from chronic diseases. The model ensures that care is provided within the context of primary care, working with a plan to integrate the patient, provider and plan needed to achieve the overall goal of improving care for chronic diseases.

The chronic care model consists of six components:

1. Health System/Organizational Support: Creation of an organization, culture and processes that encourages safe health care.
   (i) Encourage open and methodical treatment of errors and quality concerns in order to enhance care.
   (ii) Incentives for providing high-quality treatment.
   (iii) Create agreements that enable care coordination both within and between organizations.

2. Clinical Information Systems: To support efficient and effective care, organize patient and population data.
   (i) Timely reminders for routine care.
   (ii) Individual patient care is supported.
   (iii) Information is shared between providers and patients for coordination.
   (iv) Performance of the practice team and healthcare system is monitored.

3. Delivery System Design: Make certain that clinical treatment and self-management assistance are provided in a timely and efficient way.
   (i) Planned interactions must accompany evidence-based care.
   (ii) Complex patients should be provided with clinical case management.
   (iii) Regular follow-up by the healthcare team is ensured.
   (iv) Care to the patients' needs to be provided keeping in mind their cultural background.

4. Decision Support: Encourage the use of clinical treatment based on scientific evidence and patient preferences.
   (i) Incorporate evidence-based guidelines into daily clinical practice.
   (ii) Patients should be given evidence-based instructions and information to encourage them to participate.
   (iii) Primary care and expertise should be integrated.

5. Self-Management Support: Patients should be equipped and ready to manage their own health and treatment.
   (i) Stress the importance of the patient's participation in their own health care.
   (ii) Assessment, goal setting, action planning, issue resolution, and follow-up are all successful self-management support tactics.
   (iii) Organize internal and community resources to assist patients with continuous self-management.

6. Community Resources: Mobilize community resources to meet the needs of patients.
   (i) Encourage patients to participate in community-based activities that are effective.
   (ii) Form ties with community groups to support and develop solutions that fill gaps in important services.
   (iii) Promote policies that will improve patient care.

B. Problems with the Traditional Model

Before discussing how explainable AI and the chronic healthcare model can work in synergy, some major limitations of the chronic healthcare model must be understood. Below are some identified limitations:

1. Health professionals might be working in isolation and may have insufficient knowledge, low motivation, and even economic concerns.

2. There may be patients who are difficult to motivate; they may not be proactively taking part in engaging with the self-management support component of the model.

3. Errors and omissions may be present in clinical information systems.

4. The six components require integrated applications and empowered interactions to optimize their benefits to the healthcare system.

5. Each chronic healthcare model is case specific, and what is applicable for developed countries may not be remotely applicable for developing countries like India. Further investigation of case specificness relating to the Indian scenarios requires an increase in collaborative efforts from various stakeholders. This will require calculatedly huge financial support.

C. Explainability for Improvement

Following are the ways in which XAI can be used as an added advantageous technology for chronic healthcare models:

1. The proposed Decision Support System by the very name is for decision support thus helping the professionals who are not as knowledgeable as required. These professionals would be provided with the right insights and tools to discover more which would keep them motivated.

2. Although there is no doubt that chronic diseases have a negative impact on the healthcare system and individuals across the country, analytics technologies may be able to help doctors anticipate bad patient outcomes. Explainable AI not only will provide reasons or features that have led to certain predictions but will also suggest counterfactuals which inform about the relevant changes required in a feature for improvement. Patient engagement with understandable analytics can help them become proactively involved in their health improvement journey.

3. Machine learning models improve accuracy and reliability with more data and time. System errors and omissions can be minimized as more context specific data is made available.

4. Integration of the components of the chronic healthcare model can be made possible with the help of the proposed DSS as the key to all the components is explainable.

5. For complex chronic diseases, the hope is that machine learning could help provide some additional understanding which is provided by Explainable Artificial Intelligence at an
optimized cost. It provides relevant inferences and explanations for the predictions which helps doctors to better analyze the results, making it suitable for the Indian environment.

Thus, the proposed XAI-DSS framework can be a base on which a chronic healthcare model can be built for government hospitals, which are in dire need of such models.

VI. CONCLUSION

Highlighting the usefulness of XAI for the Indian healthcare scenario, following are the conclusions:

1. XAI techniques can be the basis of a decision support system that helps in clinical diagnosis of chronic diseases.
2. For different chronic diseases, feature identification and drawing feature importance is made possible with the application of XAI techniques. In the case of diabetes, according to the model, HbA1c and BMI have been identified as the two most important features as per the conventional practice.
3. Dependency plots help to study the trends of specific features. This helps the doctors to visualize independent relationships of these features with the prediction classes. Counterfactuals help in studying the impact of feature variations and thus in identifying minimum changes required to align with the prescribed health standards.
4. Based on the diabetes dataset, the following techniques were found to be relevant for this case of application:
   (i) ALE, PDP and PDP Interact which help in visualizing individual feature dependencies.
   (ii) Correlation heatmaps which helps in identifying dependence between different features.
   (iii) Permutation importance which gives the range of variations for each feature.
   (iv) Anchors help in identifying the minimum values for each feature for a particular class prediction.
   (v) LIME is used when both feature importance and anchors are needed simultaneously.
   (vi) SHAP is used for concise visualization of trends and dependencies.
5. This analysis helps in understanding the importance and use of XAI techniques in making AI predictions more reliable and efficient.

VII. FUTURE SCOPE

The present study is carried out to demonstrate the utility of XAI technique with limited clinical data. When more and more clinical data is openly available, the model can further be validated in different ways in which XAI can be used for automated diagnosis. The model can be trained again, say every fortnight because new data keeps coming and the model gets trained, and its accuracy keeps on getting improved.

REFERENCES

[1] Dharuradhar A, Chen Y, Luss R, Tu C, Ting P, Shannugam K, Das P. Explanations based on the missing: Towards contrastive explanations with pertinent negatives. Advances in Neural Information Processing Systems, 2018;592–603.
[2] García V, Aznarte L. Shapley additive explanations for no2 forecasting. Ecological Informatics, 2020;56.
[3] Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y, Dong Q, Shen H, Wang Y. Artificial intelligence in healthcare: past, present and future. Stroke and Vascular Neurology, 2017; 2 (4): 230–243.
[4] Looveren A.V, Klaise J. Interpretable counterfactual explanations guided by prototypes. 2020.
[5] Lundberg S.M, Lee S.I. A unified approach to interpreting model predictions. Advances in neural information processing systems, 2017: 4765–4774.
[6] Pawar U, O’shea D, Rea S, O’Reilly R. Explainable AI in healthcare. Reasonable Explainability for Regulating AI in Health. June 2020.
[7] Ribeiro M.T, Singh S, Guestrin C. Anchors: High-precision model-agnostic Explanations. AAAI Conference on Artificial Intelligence (AAAI), 2018.
[8] Sundararajan M, Taly A, Yan A. X-agnostic attribution for deep networks. arXiv preprint arXiv.2017;1703.01365.
[9] Thampi A. Interpretable AI: Building explainable machine learning systems. Manning Publications, 2020.
[10] Wachter S, Mittelstadt B, Russell C.Counterfactual explanations without opening the black box: Automated decisions and the gdpr. 2018.
[11] Alex A., Freitas. Comprehensible Classification Models: A Position Paper. SIGKDD Explor. Newsl. 2014; 15:1–10.
[12] Friedman J. Greedy function approximation: a gradient boosting machine. Annals of statistics, 2001; 1189–1232.
[13] Friedman J, Popescu B. Predictive learning via rule ensemble. The Annals of Applied Statistics 2, 2008; 3: 916–954.
[14] Gade K, Geyik S, Kenthapadi K, Mithal V, Taly A. Explainable AI in Industry. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (Anchorage, AK, USA) (KDD '19) Association for Computing Machinery.2019; 3203–3204.
[15] Garcia M, a José L. Aznarte. Shapley additive explanations for NO2 forecasting. Ecological Informatics, 2020;56(101039).
[16] Patel V, Mazumdar-Shaw K, Kang G, Das P, Khanna T. Reimagining India's health system: a Lancet Citizens’ Commission. The Lancet, 2021;397(10283):427-1430.
[17] Kausik A, Raman A. The new data-driven enterprise architecture for e-healthcare: Lessons from the Indian public sector. Government Information Quarterly, 2015;32(1): 63-74.
[18] Dharuradhar D, Goswami M, Kumar G. Impact of Trust and Privacy Concerns on Technology Acceptance in Healthcare: An Indian Perspective. International Journal of Medical Informatics, 2020;41(104164): ISSN 1386-5056.

Het Naik is a final year undergraduate student currently pursuing a Bachelor of Technology in computer engineering from Pandit Deendayal Energy University, Gandhinagar, India. He has a deep interest in Explainable AI and has published two research papers on arXiv titled “Explainable AI for ICT: System and Software Architecture” has been accepted in ACM CSUR (ACM Computing Surveys). His book chapter titled “Explainable AI meets healthcare: A study on heart disease dataset” and “Towards Designing Computer Vision-based Explainable-AI Solution: A Use Case of Livestock Mart Industry”. His journal paper titled “A Survey on Programming Frameworks for Explainable AI: State of the Art” has been accepted in ACM CSUR (ACM Computing Surveys). His book chapter titled “Explainable AI for ICT: System and Software Architecture” has been accepted under the book titled “Recent Advancements in ICT Infrastructure and Applications”. He worked as a Machine Learning Engineering Intern at Larsen & Toubro Defence. He was an Engineering Fellow at Smart Village and Applications”. He worked as a Machine Learning Engineering Intern at Larsen & Toubro Defence. He was an Engineering Fellow at Smart Village Movement, an initiative by U.C. Berkeley. He is the Co-founder of Seatizen, an AI-based start-up incubated under the Government of Gujarat. He was the semi-finalist in Smart India Hackathon 2020. Mr. Naik is a member of the Computer Society of India. He is also the technical head of its University Student Chapter. He is the AI/ML Lead for Google Developer Student Clubs, PDEU.

Priyanka Goradia is a final year undergraduate student currently pursuing a Bachelor of Technology in computer engineering from School of Technology, Pandit Deendayal Energy University, Gandhinagar, India. She has co-authored two machine learning based papers titled, “Covid-19 Vaccine Tweets Sentiment Analysis and Topic Modelling for Public Opinion Mining May 2021” & “Deep Convolution Neural...
Networks for Scene Understanding: A Study of Semantic Segmentation Models” that have been accepted for 2021 International Conference on Artificial Intelligence and Machine Learning and are to be published in IEEE Digital Library.

She worked as a Machine Learning Intern at Verzeo and as an Associate Technical Consultant at OpsHub. She was an Engineering Fellow at Smart Village Movement, an initiative by U.C. Berkeley. She is the Co-founder of Seatizen, an AI+X based startup incubated under the Government of Gujarat.

She was the semi-finalist in Smart India Hackathon 2020 and a part of the winning team of the university level hackathon that developed a concept map for an Intelligent Traffic Control System. She has been a part of the university level Robotics club’s technical team and an executive member of the Codechef student chapter. She has an innate interest in Psychology. She is currently an undergraduate researcher for Motor Imagery Classification using Brain Computer Interface at the university.

Vomini Desai is a final year undergraduate student currently pursuing a Bachelor of Technology in Information and Communication Technology from School of Technology, Pandit Deendayal Energy University, Gandhinagar, India. She worked as an Intern at Gujarat Energy Research & Management Institute (GERMI). She was an Engineering Fellow at Smart Village Movement, an initiative by U.C. Berkeley. Ms. Vomini was the Documentation Head of Encode: The Computer Science Club of PDEU.

Yukta Desai is a final year undergraduate student currently pursuing a Bachelor of Technology in Information and Communication Technology from School of Technology, Pandit Deendayal Energy University, Gandhinagar, India. She worked as a Machine Learning Engineering Intern at Larsen & Toubro Defence. She was an Engineering Fellow at Smart Village Movement, an initiative by U.C. Berkeley. She was the semi-finalist in Smart India Hackathon 2020. Ms. Yukta is a member of the Google Developer Student Clubs, PDEU.

Prof. Dr. Muralikrishna Iyyanki, PhD from the Premier Institute -Indian Institute of Science (IISc), Bangalore, was Dr Raja Ramanna Distinguished Fellow at the Defense Research and Development Organization (DRDO), India, and the National Coordinator for Geospatial Public Health, which is National Networking Government of India Project. He was the Editor and Member of the Editorial board of several S&T Journals. His present research focuses are on Open innovation, Artificial intelligence and geospatial data analytics and geospatial population and public health management. He was the Professor and Founder Head of the Centre for Spatial Information Technology (CSIT) at Jawaharlal Nehru Technological University (1990-2005), and Director of Research and Development Centre (2005-2008). He has served as a guest scientist at German Space Research Institute (DLR) and GKSS Research Centre.