A Minor Project Synopsis on

The Healthcare Sector: Evolution from Blackbox models to SHAP and LIME in Explainable AI

Submitted to Manipal University, Jaipur

Towards the partial fulfilment for the Award of the Degree of

BACHELOR OF TECHNOLOGY

In Computers Science and Engineering

2022-2023

By

Name of the Candidate: Ritvik Chawla

Registration Number: 209301056

MANIPAL UNIVERSITY
JAIPUR

Under the guidance of

Name of the Supervisor: Dr Rishi Gupta

Department of Computer Science and Engineering

School of Computer Science and Engineering

Manipal University Jaipur

Jaipur, Rajasthan

1. **Introduction:** (should not exceed 1 page)

Artificial intelligence (AI) models have been used to automate decision-making in a variety of industries, from business to more important ones that significantly influence people's lives, like healthcare. There is a growing movement seeking to construct medically comprehensible AI systems, even though a large percentage of these suggested AI systems are regarded as black box models with no explainability. If people can understand how an AI system arrived at its conclusion, the system is said to be explainable. There is a discussion of several XAI-driven healthcare techniques and how well they performed in the present research. The current paper discusses the development tools utilised in local and global post hoc explainability as well as the many explainability methodologies relevant to logical, statistics, and operational explainability. The Local Interpretable Model-Agnostic Explanations and Shapley Additive Explanations are used to enforce the explainability of the artificially intelligent system in the medical sector for a stronger comprehension of the internal functionalities of the classic AI models and the similarity among the number of features that impacts the decision of the model.

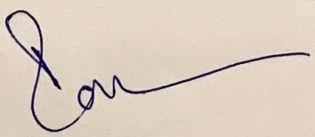
The state-of-the-art XAI-based methods of today and those of the future are documented on scientific studies in many implementation elements, encompassing research difficulties and model constraints. It is addressed how XAI might be used in the healthcare industry for everything from disease detection to prior illness prediction. Along with several explainability tools, the parameters considered in assessing the model's explainability are offered. For better comprehension, three scenarios regarding using XAI in the healthcare industry are included with their results. The potential of XAI in healthcare will help researchers gain insight into the subject.

2. **Motivation:** (should not exceed 1 page) **(For the research project only)**

I have undertaken this research project as I am interested in Machine Learning, Data Science and Artificial Intelligence. I have been learning and exploring these domains since my fresher year at Manipal University, Jaipur and I feel I can research more in these fields through this opportunity in my 6th-semester minor project. A major industry where explainable AI is being applied today is the healthcare sector, where extensive research is being done for finding and refining techniques with more explainability. I want to research this domain since I have the skillset required to understand the practical functionalities of the explainable AI models and what could lie ahead. This will help me access further opportunities in the domain of explainable AI and will get me to explore the ongoing advancements in the medical sector since in the coming time, the health index will play a major role to check any community or nation's development. Wherever people are healthy, there will be progress and to meet such standards, technology needs to advance. Therefore, explainable AI and its toolkits will play an enormous role in the rising development of such standards.
In today's world, where 'Information is power' and 'Time is money', this field of research can uncover answers which will ultimately provide both information and time to people and organizations.

3. **Project Objective:** (with cons and pros of existing methods in tabular form)

The current study will examine various explainable perspectives, tools, and research reports that would help people in explaining the models employed in the health sector. When it comes to medical implementations, there is a meaningful relationship between the efficiency of treatment or therapy and the acknowledgement of the approach used.The following is an overview of the studies' total contributions:



- Explaining the several domains in which XAI models and systems can be included making the models noticeable.
- Discuss the different systems, such as explainable models based on LIME and SHAP.
- Talk about the many toolkits that are offered to make the model explainable under different explainable variables.
- Present the numerous explainability categories connected to the decision models in intricate detail.
- The study would be easier to understand as healthcare-related case studies and statistical analyses are presented.

PROS	CONS
• the LIME's basic operating premise is to evaluate the model for local transparency and understandability. To assess the local transparency of the model, the characteristics used throughout the prediction phase are crucial. The local explainability does, however, increase the prediction process' vigilance.	• While the process's vigilance does increase in a LIME models basic operation premise, it may or may not fit the model globally.
• Utilizing XAI technology, gene expression variations are being studied. Since rule-based techniques are particularly well adapted for empirical confirmation of the predictions generated from gene analysis, XAI was developed as a solution to this problem.	• Numerous reliable methods have been utilised as "black boxes," providing no information on the usage of certain evaluation, classification, and prediction techniques. Although consumers' potential to use the tools or apps that these models equip may not be immediately impacted by this lack of transparency, professionals may nevertheless be able to understand their structure.
• To approximate the coherence and accuracy of the local model, the Shapy values in the SHAP model provide a distinct additive feature set. Both model-specific and model-independent justifications may be used with SHAP successfully.	• For a comprehensive understanding of diverse challenges, collecting and examining useful but outdated functionalities may be necessary. In several disciplines of healthcare informatics, determining the kind of illness and future model risk analysis requires the application of data collecting, preprocessing, preparation, modelling, and visualisation.
• In order to replace the originally utilised linear models with more complicated and conceivably more reliable models, interpretable artificial intelligent models and methodologies for error handling, description, and	

fairness may aid in the spread of and trust in newer or more stringent machine learning techniques.	

4. **Methodology/ Planning of work:** (Methodology will include the steps to be followed to achieve the objective of the project during the project development. Include Gantt chart, flow chart, entity, and relation diagram with specific timelines)

    1. The paper will start with first defining the problem statement related to the topic chosen.
    2. Next, we need to explain the current and prerequisite knowledge about the problem statement we are trying to solve. This includes the theoretical knowledge about the previous black box models along with the innovations which took place with them and all problem statements they helped tackle.
    3. We shall also include flow charts and diagrams which help the reader understand the concepts we are trying to explain as well as the features of the problem statement through such visuals.
    4. A few examples are:

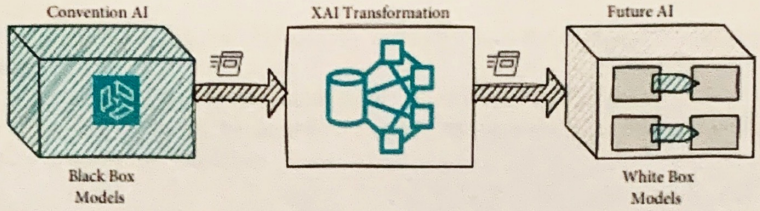
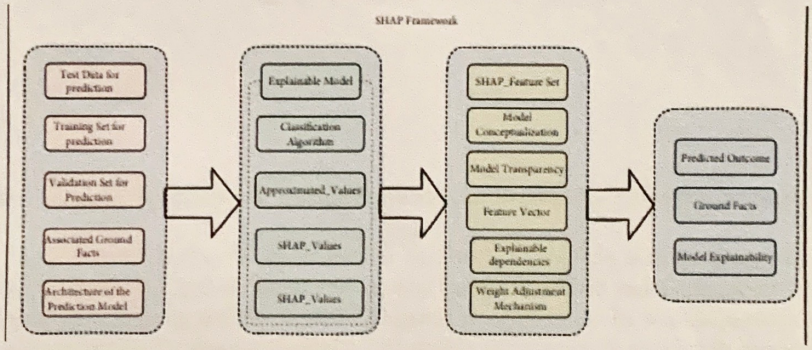


Image representing the transformation of XAI



SHAP framework for the explainable model.

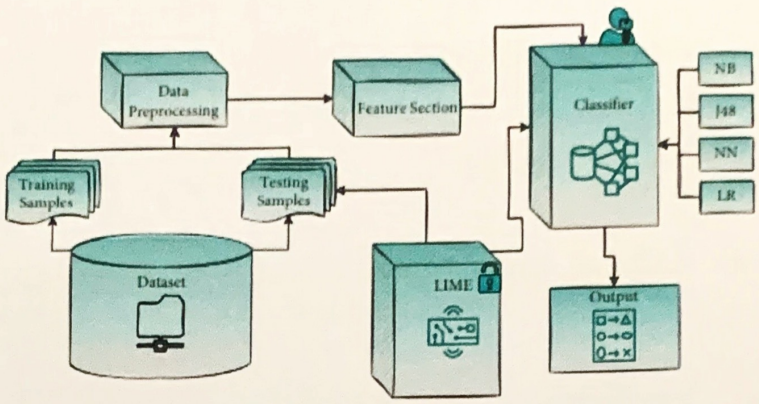
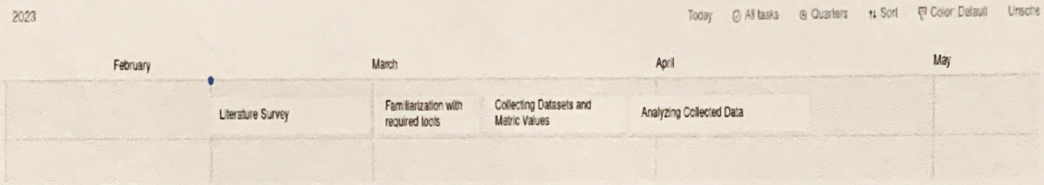


Image representing the block diagram of the LIME-based prediction model.

5. Try out a minor project on certain platforms like Google Collaboratory, Kaggle, Jupyter Notebook, etc where we can get results and outputs corresponding to the problem statement.
6. Credit the sites using the bibliography through which the content for your research came from.
7. Upload the project on GitHub and get the research paper published.
8. A Gantt chart is shown below which explains the time duration that will be followed in order to cover this research.



5. **Facilities required for proposed work:** (Software/Hardware required for the development of the project.)
   - Google Collaboratory/ Kaggle/ Jupyter Notebook: Used for a minor project which will help explain the XAI concepts and toolkits being researched and discussed in this paper. Machine Learning and Artificial Intelligence algorithms can be implemented on such platforms to get certain statistical conclusions and parameters which will help consolidate the research.

6. **Bibliography/References:**

1) Parvathaneni NagaSrinivasu , 1N. Sandhya , 1RutvijH. Jhaveri , 2 andRoshaniRaut13 June 2022 “From Blackbox to Explainable AI in Healthcare: Existing Tools and Case Studies”

2) 1,2 Ugochukwu Onwudebelu 1LIPN, Université Sorbonne Paris Nord, 99, avenue jeanbaptisteclément, 93430 Villetaneuse, Paris & 2Department of Computer Science Alex-Ekwueme Federal University of Ndufu-Alike Ikwo (AE-FUNAI), Abakaliki Ebonyi State, Nigeria, September 2021 “The Third Wave of AI: Fraud Detection via Knowledge Graphs

within Explainable AI"

3) SERHIY KANDUL, University of Zurich, Switzerland VINCENT MICHELI, University of Geneva, Switzerland JULIANE BECK, University of St. Gallen, Switzerland MARKUS KNEER, University of Zurich, Switzerland THOMAS BURRI, University of St. Gallen, Switzerland FRANÇOIS FLEURET, University of Geneva, Switzerland MARKUS CHRISTEN, University of Zurich, Switzerland, January 2023, "Explainable AI: A review of the empirical literature"

4) Dr. Tom KrausLeneGanschowMarlene EisenträgerDr. Steffen Wischmann, April 2022, "Explainable AI - Requirements, Use Cases and Solutions"