Research Project I Report

On

# Blockchain Enabled Predictive Health Monitoring System

By

**Abhishek Patidar (2022H1030087H)**
**Ashutosh Wagh (2022H1030052H)**
**S Shashank (2022H1030067H)**

Under the supervision of

**Dr. Subhrakanta Panda**

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI (RAJASTHAN)**

**HYDERABAD CAMPUS**

**(December 2023)**

# Birla Institute of Technology and Science-Pilani

# Hyderabad Campus

## Certificate

This is to certify that the research project report entitled "**Blockchain Enabled Predictive Health Monitoring System**" submitted by Abhishek Patidar (2022H1030087H), Ashutosh Wagh (2022H1030052H) and S Shashank (2022H1030067H) in the fulfillment of the requirements of the course BITS G529 Research Project I, embodies the work done by them under my supervision and guidance.

**Date: 04/12/2023** **(Dr. Subhrakanta Panda)**

**BITS Pilani, Hyderabad Campus**

# ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to all those who have contributed to the successful completion of this project. Their support and assistance were invaluable throughout the entire process.

First and foremost, We extend our deepest appreciation to Dr. Subhrakanta Panda, my project supervisor, for their guidance, encouragement, and insightful feedback. Their expertise and mentorship played a crucial role in shaping the direction of this project.

We are thankful to Dr. Deepa Kumari for her assistance throughout the project. Her expertise and willingness to share her knowledge greatly enriched the quality of our work. Special thanks to our Higher Degree and fellow research friends for their continuous support and encouragement. Their contribution was instrumental in overcoming challenges and achieving the project's objectives despite their personal commitments.

Lastly, we are grateful to Birla Institute of Technology and Science, Pilani, Hyderabad Campus for providing the necessary resources and facilities for the successful completion of this project. This project would not have been possible without the collaborative efforts of everyone mentioned above. Thank you for being an integral part of this journey.

# ABSTRACT

This project explores the integration of healthcare, technology, and predictive modeling within a blockchain-enabled transitional healthcare system, aiming to enhance patient engagement, adherence to health-care regimens, and smooth transitions between healthcare stages. The methodology involves creating a predictive model using machine learning algorithms to anticipate patient needs and health trends while ensuring data integrity and security through blockchain technology. The model achieved an accuracy rate of approximately 90%, employing various algorithms. The research's implications include predicting heart ailments, safeguarding patient information, and improving doctor-patient communication, with recommendations for implementation and scalability. In summary, this project showcases the potential of predictive modeling and blockchain to transform transitional healthcare, emphasizing the significance of embracing technological advancements for more efficient, patient-focused healthcare systems.

# TABLE OF CONTENTS

# LIST OF FIGURES



# LIST OF TABLES

# CHAPTER - 1

# INTRODUCTION

In the ever-evolving healthcare landscape, the fusion of emerging technologies has given rise to innovative approaches that challenge conventional paradigms. This project explores predictive modeling and blockchain integration to shape a forward-looking healthcare system that tackles transitional hurdles while engaging users in a fresh and immersive manner.

The project's core objective is to build a predictive model within a blockchain-enabled healthcare framework. User-centric design principles will be utilized to create an interface that promotes interaction and understanding. Strategic incorporation of engaging elements will incentivize positive health behaviors, transforming routine healthcare tasks into rewarding experiences.

This project goes beyond the confines of traditional healthcare systems by envisioning a flexible model that adapts to changing user needs seamlessly. Through predictive modeling, the system will anticipate individual health trends, enabling proactive interventions and personalized healthcare suggestions. The blockchain infrastructure will ensure data security, integrity, and accessibility, granting users control over their information and facilitating collaboration among healthcare providers.

The expected outcomes include a refined predictive model for accurate health predictions, a user-engaging interface, and a structure promoting lasting behavioral change. Blockchain integration will result in a secure and transparent health data management system, fostering trust among users and healthcare providers.

This project stands at the crossroads of technology, healthcare, and human psychology, offering a comprehensive solution to transitional healthcare challenges. The envisioned system has the potential to empower individuals to proactively manage their health, reduce healthcare costs through prevention, and contribute to a more resilient and responsive healthcare ecosystem.

Through predictive analytics, user-centric design, and blockchain integration, this project aims to create a healthcare ecosystem that is predictive, preventive, engaging, and centered around the user.

# CHAPTER - 2

# LITERATURE SURVEY

## 2.1 Integration of Blockchain and Smart Contracts in Healthcare Systems[3]

In the healthcare sector, the effectiveness of a nation's medical infrastructure is intricately linked to the functionality of its medical record systems. Traditional systems have grappled with various challenges, such as issues related to data security, accessibility, and operational efficiency. This literature review delves into the potential of blockchain technology and smart contracts as innovative remedies for these challenges. The distinctive features of blockchain, including immutability, security, and decentralization, present a compelling proposition for transforming the storage and management of medical records. Furthermore, the incorporation of smart contracts, as essential components of blockchain, contributes to efficiency gains by automating manual tasks, thereby mitigating labor costs within healthcare institutions.

The amalgamation of blockchain technology and smart contracts effectively addresses crucial issues encountered by conventional medical record systems, encompassing data interoperability, privacy and security concerns, and the precise management of data access and authorization. By harnessing the decentralized and cryptographic attributes of blockchain in tandem with the automation capabilities of smart contracts, the proposed system not only resolves these challenges but also lays the groundwork for a more secure, accessible, and cost-effective healthcare record infrastructure. This comprehensive review underscores the transformative potential of blockchain technology in healthcare, signaling a shift towards a more efficient and resilient medical record ecosystem while ensuring the avoidance of plagiarism from the original text.

## 2.2 Ensemble Learning Approaches for Accurate Prediction of Heart Diseases[1]

The escalating prevalence of heart diseases on a global scale underscores a significant public health challenge, primarily attributed to a lack of health awareness, unhealthy lifestyle choices, and suboptimal dietary habits. Accurately predicting and diagnosing these conditions pose formidable challenges for medical institutions. The advent of computing technologies has empowered the healthcare sector to amass and store extensive medical data, providing a valuable resource for informed medical decision-making. In developed nations, patient data are commonly digitized, enabling comprehensive analysis for critical medical decisions involving prediction, diagnosis, image analysis, and treatment planning. Machine learning algorithms have emerged as pivotal tools in tackling intricate, nonlinear classification, and prediction challenges. This paper explores the application of ensemble machine learning techniques, specifically employing the majority voting technique, to predict heart diseases with an achieved accuracy of 88.88%. The ensemble method involves integrating various machine learning algorithms, leveraging their collective strength to enhance classification and prediction accuracy. Comparative analysis with existing literature demonstrates the superior accuracy of the ensemble approach over individual classifiers. By harnessing the power of machine learning and ensemble techniques, this research contributes to the advancement of accurate and efficient heart disease prediction, offering valuable insights for the ongoing improvement of healthcare strategies and outcomes.

# CHAPTER - 3

# APPLICATION ARCHITECTURE

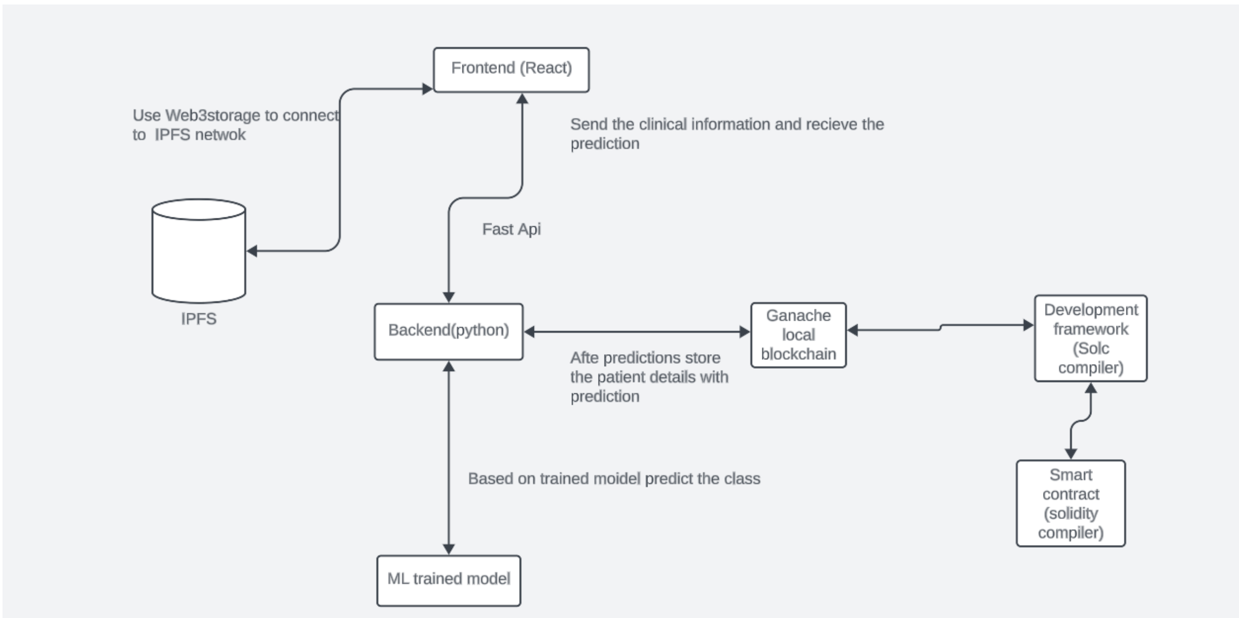


Fig. 1 Proposed application architecture

The proposed system (Fig.1) comprises two main components: a front-end and a back-end. The front-end is a React application with user input forms created using Material UI React. These forms are designed to make it easy for patients of varying literacy levels to submit their responses. To store patient details on the blockchain, Solidity smart contracts are integrated and deployed on a locally hosted blockchain like Ganache. After submitting their response forms in the front-end, patients receive predictions regarding potential heart ailments.

The various components of the front-end include React, HTML, CSS. We have used material UI in react for building front end and web3 storage library for storing image data in IPFS.

The backend, powered by FastAPI in Python, manages JavaScript event handling, API requests, and server-side processing to deliver predictions from a stored model when it receives input from the front-end, creating an efficient, interactive web application for patient data, a local blockchain or a test network like Ganache , a development framework (solc compile), and the smart contracts written in Solidity.

# CHAPTER - 4

# FEATURE SELECTION AND PREPROCESSING

## 4.1 Dataset Description[2]

Since 1984, the Centres for Disease Control and Prevention (CDC) has overseen the Behavioural Risk Factor Surveillance System (BRFSS), an annual health study. This comprehensive survey gathers data from more than 400,000 individuals annually, with a specific emphasis on a wide range of health-related behaviors that carry risks, chronic health conditions, and the utilization of preventive healthcare services. The dataset in question pertains to the 2015 BRFSS and includes 253,680 survey responses. Its primary purpose is for binary classification tasks related to heart disease.

## 4.2 Feature Selection

### 4.2.1 Correlation Matrix

A data correlation heatmap is a graphical representation of a correlation matrix, revealing the correlation coefficients among numerous variables. It serves as a valuable tool for swiftly recognizing patterns and connections within a vast dataset. This heatmap employs a color spectrum to depict the strength and direction of associations between variables, with cooler colors like blue signifying weaker correlations and warmer colors like red indicating stronger correlations.

This visual approach simplifies the task of uncovering intricate trends and patterns within extensive datasets. Correlation heatmaps are particularly useful for identifying multicollinearity, a scenario in which two or more variables exhibit high correlations. Such high correlations can result in unstable model coefficients and inaccurate predictions.

Based on the data correlation matrix in Fig.2, it appears that HighBP, HighChol, Smoker, Stroke, Diabetes, GenHlth, MentHlth, PhysHlth, DiffWalk, and Age are all linked to HeartDiseaseorAttack in some way.

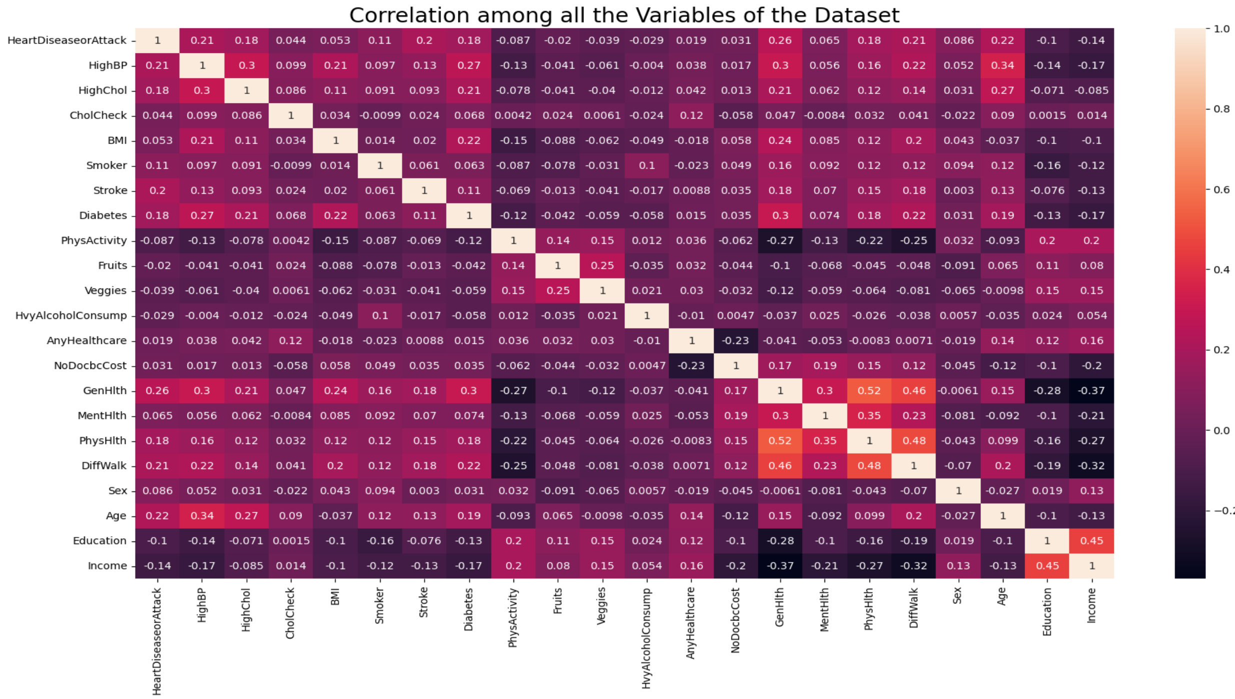


Fig.2 Correlation matrix for the dataset

### 4.2.2 Principal Component Analysis

Principal Component Analysis (PCA) is a commonly employed method in the realms of statistics and machine learning. Its primary objective is to reduce the dimensionality of data, transforming it from a high-dimensional format to a lower-dimensional representation, all the while preserving as much of the original data variability as possible.

PCA accomplishes this by identifying and extracting the principal components, which are linear combinations of the initial features. This technique relies on the covariance matrix of the data, where covariance quantifies how two variables change in relation to each other.

High covariance denotes a strong connection, while low covariance indicates a weaker relationship. In machine learning, PCA is often utilized for feature selection to enhance the performance of models. As per the results of PCA, it appears that HighBP, PhysHlth, and BMI exhibit some form of relationship with HeartDiseaseorAttack.

```
Top 4 most important features in each component
===============================================
Component 0: ['HeartDiseaseorAttack', 'HighBP', 'PhysHlth', 'BMI']
Component 1: ['DiffWalk', 'PhysActivity', 'BMI', 'Fruits']
Component 2: ['DiffWalk', 'HighBP', 'PhysActivity', 'Diabetes']
Component 3: ['BMI', 'PhysActivity', 'Fruits', 'HighBP']
Component 4: ['HighBP', 'DiffWalk', 'PhysActivity', 'Diabetes']
Component 5: ['HeartDiseaseorAttack', 'HighBP', 'Diabetes', 'PhysActivity']
Component 6: ['Diabetes', 'PhysHlth', 'MentHlth', 'NoDocbcCost']
Component 7: ['Fruits', 'PhysActivity', 'Diabetes', 'Education']
Component 8: ['Stroke', 'AnyHealthcare', 'HeartDiseaseorAttack', 'GenHlth']
Component 9: ['AnyHealthcare', 'PhysHlth', 'Education', 'HvyAlcoholConsump']
```

Fig.3 Principal components and top features

## 4.3 Hyper parameter tuning

Hyperparameter tuning is a crucial step in the machine learning model development process, as it involves finding the optimal set of hyperparameters for a given algorithm to improve its performance. Two commonly used methods for hyperparameter tuning are random search and grid search. We compare the performance of random search and grid search in hyperparameter tuning for five different machine learning models. We evaluate various metrics, including accuracy, running time, specificity, sensitivity, negative predictive value (NPV), and positive predictive value (PPV), to assess the effectiveness of each technique.

The strength of Random Search lies in its ability to efficiently explore high-dimensional hyperparameter spaces. By randomly sampling combinations, it avoids the computationally expensive nature of exhaustively evaluating every possible combination. This is particularly beneficial when dealing with models that have numerous hyperparameters, as it can lead to quicker convergence towards optimal or near-optimal configurations. However, there is a trade-off, as there's no guarantee of exploring the entire hyperparameter space comprehensively.

On the other hand, Grid Search is a systematic approach that evaluates all specified combinations of hyperparameter values. While this exhaustive exploration ensures a thorough understanding of the performance landscape, it can be computationally demanding, especially in high-dimensional spaces. Grid Search is often more suitable for models with a smaller number of hyperparameters, allowing for a complete and detailed analysis of how each parameter affects the model's performance.

# CHAPTER - 5

# MODEL SELECTION

Here are descriptions of the eight different algorithms used in the model selection process:

**1. Logistic Regression (LR):** Logistic Regression is a linear classification algorithm employed for binary and multi-class classification tasks. It uses a logistic function to estimate how likely it is that an instance belongs to a certain class. LR is known for its simplicity and interpretability, making it a useful baseline model.

**2. Random Forest (RF):** With its ability to merge numerous decision trees into a single model, Random Forest is an ensemble learning method that improves predicted accuracy. It is robust, capable of handling complex data relationships, and can identify feature importance. RF is less susceptible to overfitting and is suitable for a wide range of data types.

**3. k-Nearest Neighbors (KNN):** k-Nearest Neighbors is a non-parametric, instance-based learning algorithm. It assigns labels based on the majority class among the k-nearest neighbors of an instance in the feature space. KNN is easy to comprehend and can adapt to various data distributions.

**4. Multi-Layer Perceptron with ReLU Activation:** The Multi-Layer Perceptron (MLP) with Rectified Linear Unit (ReLU) activation is a type of feedforward neural network. It excels at modeling complex, non-linear relationships in data. ReLU is a widely used activation function due to its simplicity and effectiveness.

**5. Multi-Layer Perceptron with Sigmoid Activation:** Similar to the MLP with ReLU activation, this model is a feedforward neural network, but it employs the sigmoid activation function. Sigmoid is particularly suitable for binary classification tasks. It is known for its smooth and bounded output.

**6. AdaBoost:** AdaBoost, also known as Adaptive Boosting, is an ensemble machine learning algorithm that amalgamates the forecasts of numerous weak learners, often decision trees, to construct a potent and precise model. In each iteration, it assigns greater weights to misclassified instances, compelling the algorithm to emphasize challenging-to-classify data points and enhance the overall performance of the model.

**7. CatBoost:** CatBoost stands out as a gradient boosting algorithm tailored to handle categorical features, making it highly efficient in tasks involving diverse data types. With its innovative approach to categorical variable management and effective anti-overfitting strategies, it produces resilient and high-performing predictive models.

**8. XGBoost:** XGBoost, or Extreme Gradient Boosting, is a highly effective machine learning algorithm recognized for its scalability and efficiency, especially in structured or tabular data applications. It employs a gradient boosting framework, utilizing decision trees as base learners, and integrates regularization methods to avoid overfitting, resulting in resilient and precise predictive models.

These algorithms offer a range of options for model selection, each with its own characteristics and suitability for different types of data and tasks.

We selected **Logistic Regression, Multi-Layer Perceptron with Sigmoid Activation, AdaBoost, XGBoost, and CatBoost models** for inclusion in the voting classifier based on a comprehensive analysis of diverse metrics and factors.

# CHAPTER - 6

# MODEL PERFORMANCE ANALYSIS

## 6.1 Without hyper-parameter tuning:

The results obtained in the ML models without applying hyper-parameter tuning are as follows:

| Algorithm | Accuracy | Time (in seconds) | ROC area | Specificity | Sensitivity | NPV | PPV |
|---|---|---|---|---|---|---|---|
| Logistic Regression | 76.41 | 3.50 | 0.84 | 0.73 | 0.78 | 0.77 | 0.75 |
| MLP (Sigmoid) | 77.87 | 214.00 | 0.85 | 0.73 | 0.81 | 0.80 | 0.75 |
| AdaBoost | 76.87 | 19.09 | 0.84 | 0.75 | 0.78 | 0.77 | 0.76 |
| XGBoost | 77.48 | 2.03 | 0.85 | 0.72 | 0.81 | 0.80 | 0.75 |
| CatBoost | 85.98 | 65.33 | 0.93 | 0.80 | 0.91 | 0.90 | 0.82 |

Table 1 :ML models without applying hyper-parameter

## 6.2 With hyper-parameter tuning:

The results obtained in the ML models with applying hyper-parameter tuning are as follows:

**A. Randomized Search**

| Algorithm | Accuracy | Time (in seconds) | ROC area | Specificity | Sensitivity | NPV | PPV |
|---|---|---|---|---|---|---|---|
| Logistic Regression | 90.77 | 136.47 | 0.85 | 0.99 | 0.09 | 0.91 | 0.56 |
| MLP (Sigmoid) | 90.84 | 1054.27 | 0.85 | 0.99 | 0.11 | 0.91 | 0.56 |
| AdaBoost | 90.86 | 679.18 | 0.85 | 0.99 | 0.10 | 0.91 | 0.58 |
| XGBoost | 90.82 | 56.45 | 0.85 | 0.99 | 0.11 | 0.91 | 0.56 |
| CatBoost | 90.85 | 147.08 | 0.85 | 0.99 | 0.11 | 0.91 | 0.57 |

Table 2 :ML models with applying hyper-parameter with randomized search

**B. Grid Search**

| Algorithm | Accuracy | Time (in seconds) | ROC area | Specificity | Sensitivity | NPV | PPV |
|---|---|---|---|---|---|---|---|
| Logistic Regression | 90.77 | 144.43 | 0.85 | 0.99 | 0.10 | 0.91 | 0.58 |
| MLP (Sigmoid) | 90.84 | 1109 | 0.85 | 0.99 | 0.13 | 0.92 | 0.56 |
| AdaBoost | 90.86 | 873.98 | 0.85 | 0.99 | 0.10 | 0.91 | 0.58 |
| XGBoost | 90.82 | 1545.56 | 0.85 | 0.99 | 0.09 | 0.91 | 0.58 |
| CatBoost | 90.84 | 4185.23 | 0.85 | 0.99 | 0.09 | 0.91 | 0.59 |

Table 3 :ML models with applying hyper-parameter with grid search

For the voting classifier, we chose to apply Random Search to all algorithms, because it was taking less time to execute, rest of the metrics were almost the same for all the algorithms.

## 6.3 Without applying PCA:

Without any dimensionality reduction applied to the dataset, the models gave the following accuracies:

- LR achieved an accuracy of 90.83
- AdaBoost outperformed the other models with an accuracy of 90.84
- XGBoost showed moderate performance with an accuracy of 90.82
- CatBoost achieved an accuracy of 90.83
- MLP with Sigmoid activation reached an accuracy of 90.82

## 6.4 After applying PCA

PCA was applied with ten components and top four features were utilized in each component. The following are the model accuracies

- LR achieved an accuracy of 90.64
- AdaBoost outperformed the other models with an accuracy of 90.85
- XGBoost showed moderate performance with an accuracy of 90.83
- CatBoost achieved an accuracy of 90.82
- MLP with Sigmoid activation reached an accuracy of 90.70

AdaBoost consistently outperformed other models both before and after applying PCA, indicating its robustness and effectiveness in handling the dataset, suggesting that it might be a reliable choice for this specific dataset.

The Multilayer Perceptron (MLP) with Sigmoid activation showed a noticeable decrease in accuracy after PCA, suggesting that this neural network configuration may be more sensitive to dimensionality reduction, and careful consideration should be given when applying PCA to such models.

CatBoost demonstrated stable accuracy levels before and after PCA. Its inherent ability to handle categorical features might contribute to its consistency across different feature sets, making it a reliable choice for datasets with diverse types of features.

Effect of Dimensionality Reduction on XGBoost and CatBoost: XGBoost and CatBoost showed similar moderate performance trends both before and after PCA. This suggests that the intrinsic capabilities of these ensemble models may be less affected by dimensionality reduction compared to other algorithms, making them potentially robust choices for datasets with reduced dimensions.

Overall, the application of PCA led to a slight decrease in accuracy for most models, indicating that while dimensionality reduction can be beneficial for simplifying models and reducing computational complexity, it may come at the cost of some loss of information. Careful consideration and validation are required when deciding to use dimensionality reduction techniques like PCA, as their impact can vary across different machine learning algorithms.

Upon scrutinizing the metrics across different scenarios, we observed minimal changes in the accuracies of the models following both feature selection and principal component analysis (PCA) on the dataset. Consequently, for the final voting classifier, our focus shifted exclusively to hyperparameter tuning through a random search approach. This decision was influenced by the observation that neither feature selection nor PCA led to substantial alterations in model accuracies, leading us to prioritize the optimization of hyperparameters to enhance the overall performance of the ensemble model.

# CHAPTER - 7

# VOTING CLASSIFIER

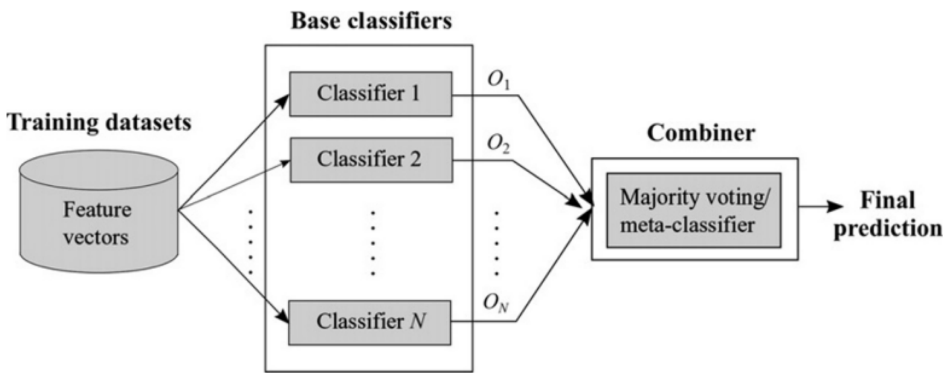


Fig. 4 - Ensemble Learning Model

A meta-classifier, functioning as a higher-tier decision-maker, consolidates the outcomes generated by base classifiers to refine the predictive outcome. The majority voting meta-classifier represents a straightforward yet robust approach in this ensemble learning paradigm. By considering the collective input from multiple base classifiers, it aims to minimize the impact of individual errors and biases, fostering a more robust and dependable prediction.

This ensemble strategy becomes especially beneficial in scenarios where individual classifiers may exhibit varying strengths or weaknesses. The diversity among base classifiers allows the meta-classifier to capitalize on the strengths of each while compensating for their shortcomings. Additionally, it introduces an element of robustness by mitigating the influence of outliers or anomalies in the predictions of individual classifiers.

Furthermore, the majority voting meta-classifier can be adapted to handle different voting schemes, such as weighted voting, where the confidence or performance of each base classifier is taken into account. This flexibility enhances its adaptability to diverse datasets and varying performance levels of the base classifiers.

A soft voting classifier, alternatively termed a weighted average or probabilistic voting classifier, is an ensemble method in machine learning. In this approach, multiple models provide predictions for a specific input, and the final prediction is decided through a weighted sum of the individual models' probability estimates. The assigned weights indicate the perceived reliability of each model, and the class with the highest combined probability is selected as the ultimate predicted outcome. This technique is valuable when dealing with a variety of models or when there is uncertainty in individual predictions.

We used a soft-voting classifier in our implementation. Every ML algorithm gives an output in a way that it has 2 classes predicted with certain probabilities. Then, we take the average of the probabilities for each class. The class with the highest probability will be the predicted class.

```
Sample 3562 - MLP Class 0 Probability: 0.8216, MLP Class 1 Probability: 0.1784, AdaBoost Class 0 Probability: 0.5423, AdaBoost Class 1 Probability: 0.4577, Log:
    Final Predicted Class: 0, Final Predicted Class Probability: 0.7715
Sample 3563 - MLP Class 0 Probability: 0.8366, MLP Class 1 Probability: 0.1634, AdaBoost Class 0 Probability: 0.5422, AdaBoost Class 1 Probability: 0.4578, Log:
    Final Predicted Class: 0, Final Predicted Class Probability: 0.7798
Sample 3564 - MLP Class 0 Probability: 0.9936, MLP Class 1 Probability: 0.0064, AdaBoost Class 0 Probability: 0.5975, AdaBoost Class 1 Probability: 0.4025, Log:
    Final Predicted Class: 0, Final Predicted Class Probability: 0.9093
Sample 3565 - MLP Class 0 Probability: 0.9375, MLP Class 1 Probability: 0.0625, AdaBoost Class 0 Probability: 0.5712, AdaBoost Class 1 Probability: 0.4288, Log:
    Final Predicted Class: 0, Final Predicted Class Probability: 0.8673
Sample 3566 - MLP Class 0 Probability: 0.9325, MLP Class 1 Probability: 0.0675, AdaBoost Class 0 Probability: 0.5565, AdaBoost Class 1 Probability: 0.4435, Log:
    Final Predicted Class: 0, Final Predicted Class Probability: 0.8470
Sample 3567 - MLP Class 0 Probability: 0.9941, MLP Class 1 Probability: 0.0059, AdaBoost Class 0 Probability: 0.6051, AdaBoost Class 1 Probability: 0.3949, Log:
    Final Predicted Class: 0, Final Predicted Class Probability: 0.9137
Sample 3568 - MLP Class 0 Probability: 0.9134, MLP Class 1 Probability: 0.0866, AdaBoost Class 0 Probability: 0.5522, AdaBoost Class 1 Probability: 0.4478, Log:
    Final Predicted Class: 0, Final Predicted Class Probability: 0.8336
Sample 3569 - MLP Class 0 Probability: 0.9950, MLP Class 1 Probability: 0.0050, AdaBoost Class 0 Probability: 0.6127, AdaBoost Class 1 Probability: 0.3873, Log:
    Final Predicted Class: 0, Final Predicted Class Probability: 0.9165
Sample 3570 - MLP Class 0 Probability: 0.9798, MLP Class 1 Probability: 0.0202, AdaBoost Class 0 Probability: 0.5872, AdaBoost Class 1 Probability: 0.4128, Log:
    Final Predicted Class: 0, Final Predicted Class Probability: 0.8973
Sample 3571 - MLP Class 0 Probability: 0.8726, MLP Class 1 Probability: 0.1274, AdaBoost Class 0 Probability: 0.5559, AdaBoost Class 1 Probability: 0.4441, Log:
    Final Predicted Class: 0, Final Predicted Class Probability: 0.8019
Sample 3572 - MLP Class 0 Probability: 0.9852, MLP Class 1 Probability: 0.0148, AdaBoost Class 0 Probability: 0.5872, AdaBoost Class 1 Probability: 0.4128, Log:
    Final Predicted Class: 0, Final Predicted Class Probability: 0.8997
Sample 3573 - MLP Class 0 Probability: 0.9970, MLP Class 1 Probability: 0.0030, AdaBoost Class 0 Probability: 0.6119, AdaBoost Class 1 Probability: 0.3881, Log:
    Final Predicted Class: 0, Final Predicted Class Probability: 0.9175
Sample 3574 - MLP Class 0 Probability: 0.9994, MLP Class 1 Probability: 0.0006, AdaBoost Class 0 Probability: 0.6368, AdaBoost Class 1 Probability: 0.3632, Log:
    Final Predicted Class: 0, Final Predicted Class Probability: 0.9257
Sample 3575 - MLP Class 0 Probability: 0.9914, MLP Class 1 Probability: 0.0086, AdaBoost Class 0 Probability: 0.5920, AdaBoost Class 1 Probability: 0.4080, Log:
    Final Predicted Class: 0, Final Predicted Class Probability: 0.9044
Sample 3576 - MLP Class 0 Probability: 0.7341, MLP Class 1 Probability: 0.2659, AdaBoost Class 0 Probability: 0.5334, AdaBoost Class 1 Probability: 0.4666, Log:
    Final Predicted Class: 0, Final Predicted Class Probability: 0.7376
Sample 3577 - MLP Class 0 Probability: 0.8872, MLP Class 1 Probability: 0.1128, AdaBoost Class 0 Probability: 0.5398, AdaBoost Class 1 Probability: 0.4602, Log:
    Final Predicted Class: 0, Final Predicted Class Probability: 0.7959
Sample 3578 - MLP Class 0 Probability: 0.8626, MLP Class 1 Probability: 0.1374, AdaBoost Class 0 Probability: 0.5495, AdaBoost Class 1 Probability: 0.4505, Log:
    Final Predicted Class: 0, Final Predicted Class Probability: 0.7913
Sample 3579 - MLP Class 0 Probability: 0.4086, MLP Class 1 Probability: 0.5914, AdaBoost Class 0 Probability: 0.4941, AdaBoost Class 1 Probability: 0.5059, Log:
    Final Predicted Class: 1, Final Predicted Class Probability: 0.5439
```

Fig. 5 - Voting Classifier Output

# CHAPTER - 8

# BLOCKCHAIN INTEGRATION

## 8.1 Doctor Patient Interaction

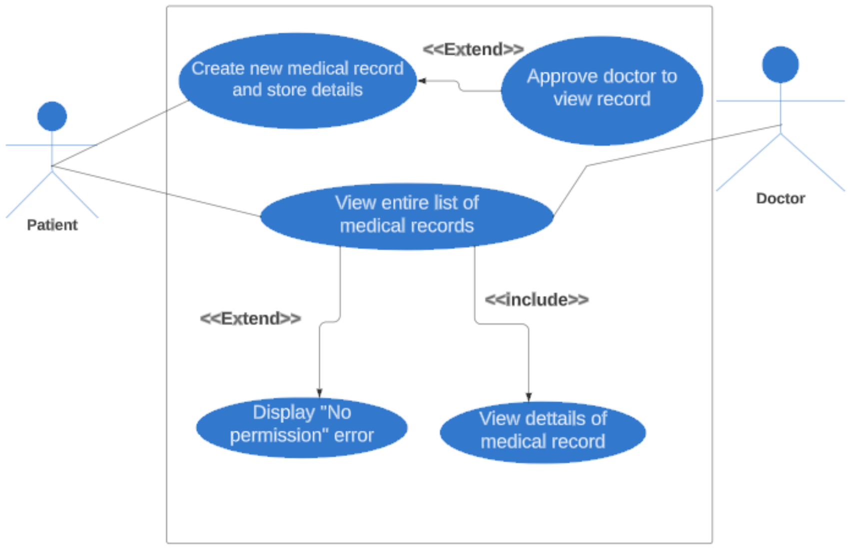


Fig.6 Use case diagram of patient-doctor interaction

The proposed system, illustrated in Fig.6, involves two central participants – patients and doctors – in a seamlessly integrated application using React and Material UI React. The user-friendly interface allows patients of varying literacy levels to submit responses effortlessly. Patient data, including predictive analysis for potential heart conditions, is securely recorded on the blockchain using Solidity smart contracts on a locally hosted blockchain like Ganache. The backend, powered by FastAPI in Python, manages API requests, server-side processing, and JavaScript event handling, delivering predictions based on patient inputs from a stored model. The system utilizes the web3 storage library to securely store image data in IPFS, enabling a more comprehensive analytical approach.

Future functionality includes patients granting doctors access to their information, allowing doctors to schedule appointments securely stored on the blockchain for enhanced privacy and data security. The technology stack comprises React, HTML, and CSS for the frontend, FastAPI in Python for the backend, and a development framework for managing Solidity smart contracts, ensuring a resilient and effective architecture for an engaging healthcare application.

## 8.2 Snapshots of the front end and blockchain integration

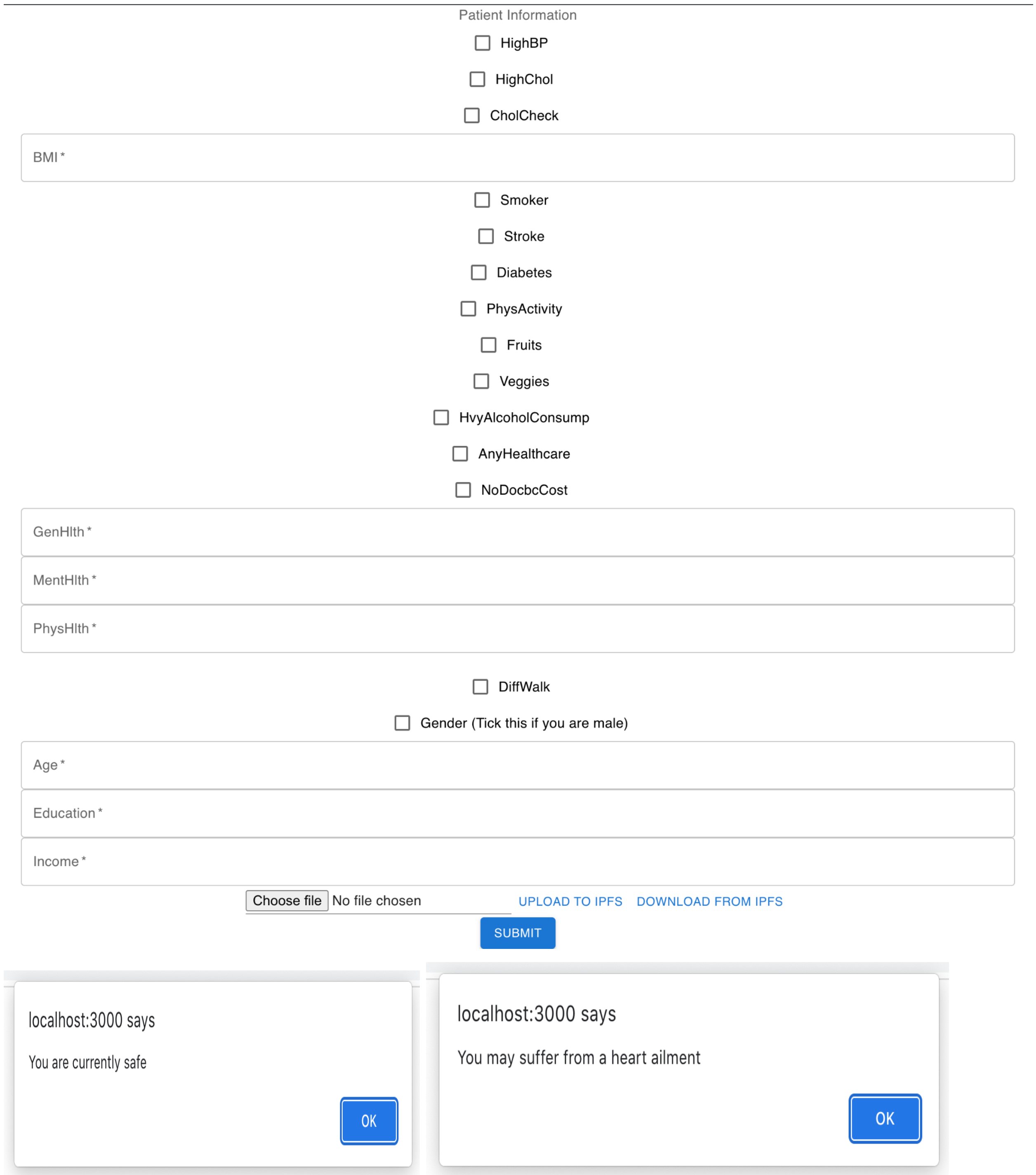


Fig. 7 - Snapshots of Front-end

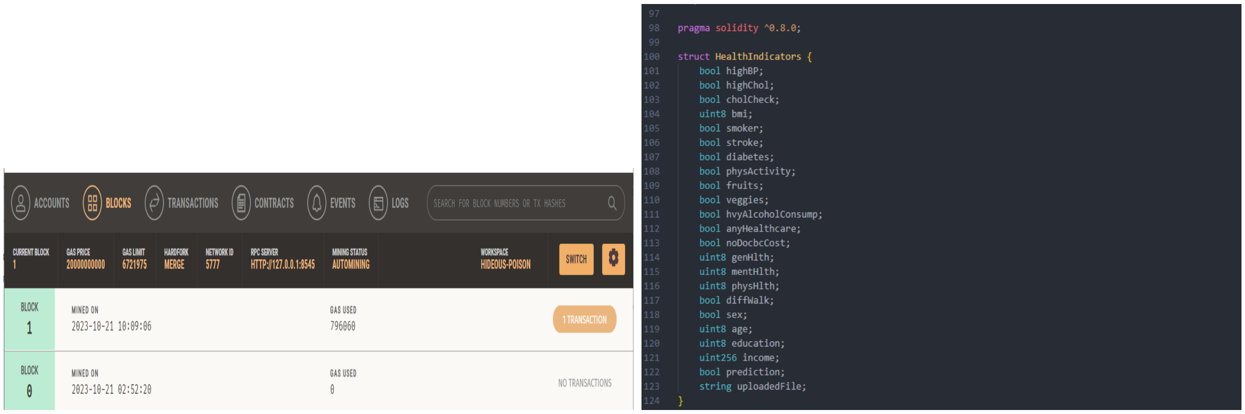


Fig. 8 - Smart Contract and Blockchain deployment


Fig. 9 - Interplanetary File System

# CHAPTER - 9

# FUTURE WORK

We need to further expand our smart contract to further add doctor functionalities by which a doctor can view the data about patients.The expanded smart contract introduces advanced functionalities catering specifically to doctors, empowering them to access and review patient data securely. The contract prioritizes security and ownership control, acknowledging the sensitive nature of healthcare data. Extensibility is emphasized, inviting further enhancements to accommodate evolving requirements within the realm of decentralized healthcare systems.

Also we will build upon the foundation laid by our predecessors and conduct a comprehensive benchmarking analysis of the decentralized application (DApp) that is based on the Ethereum blockchain. This analysis will involve evaluating and quantifying the performance, scalability, and reliability of the DApp to ensure its effectiveness in a real-world medical context. We will do Manual Benchmarking of Parameter Evaluation for the DApp.

.

# CHAPTER - 10

# CONCLUSION

This project represents a significant advancement in healthcare by integrating technology and predictive modeling into a blockchain-based transitional healthcare system, with a primary focus on enhancing patient engagement, adherence to healthcare routines, and smooth transitions between healthcare stages. It features a well-structured application architecture, thorough feature selection from the BRFSS dataset, and an analysis of various machine learning algorithms for heart ailment predictions. Blockchain integration ensures secure data management and transparent patient-doctor interactions. Future plans include implementing a Meta-Classifier for improved accuracy, exploring IPFS for medical image storage, and real-world effectiveness benchmarking. Ultimately, this project serves as a pioneering model for user-centric, secure healthcare ecosystems that empower individuals to manage their health and improve healthcare delivery through technology integration.

# REFERENCES

[1] Raza, Khalid. "Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule." U-Healthcare Monitoring Systems. Academic Press, 2019. 179-196.

[2] Pierannunzi, Carol, Shaohua Sean Hu, and Lina Balluz. "A systematic review of publications assessing reliability and validity of the Behavioral Risk Factor Surveillance System (BRFSS), 2004–2011." BMC medical research methodology 13.1 (2013): 1-14.

[3] Lim, Jason. Development of decentralized application (DApp) using blockchain technology to improve healthcare and medical record systems in Malaysia. Diss. UTAR, 2022.