Predicting Maternal Mortality Risk: Sub Urban and Rural India

MSc Research Project
Data Analytics

Niharika Subbarayappa
Student ID: X23102420

School of Computing
National College of Ireland

Supervisor: Professor. John Kelly

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Niharika Subbarayappa
Student ID:	X23102420
Programme:	Data Analytics
Year:	2023 - 2024
Module:	MSc Research Project
Supervisor:	Professor. John Kelly
Submission Due Date:	12/08/2024
Project Title:	Predicting Maternal Mortality Risk: Sub Urban and Rural India
Word Count:	5955
Page Count:	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Niharika Subbarayappa
Date:	12th August 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Predicting Maternal Mortality Risk: Sub Urban and Rural India

Niharika Subbarayappa
X23102420

Abstract

Maternal mortality being the critical issue in the global healthcare, especially in developing countries like India, where there is limited access to maternal health care and complications during pregnancy are often fatal. This study addresses these challenges of predicting maternal mortality risk in the rural regions of India by integrating machine learning (ML) and deep learning (DL) models by using data collected through IoT. The study compares the performances of three models: Random Forest, XG Boost and FNN. A hybrid model is built to combine the strengths of these individual models. The dataset includes features such as age, blood sugar levels, temperature of body and heart rates. Data was extensively pre-processed and hyper-parameter tuning was performed to optimize each model. The evaluation of these models is calculated based on accuracy, recall, precision and f1-score. The results showed that the XG Boost model achieved the highest accuracy Manik et al. (2020), slightly outperforming Random Forest model. However, the developed hybrid model which takes the combined outputs of all models has input, demonstrated the improved stability in the predictions. This research shows the potential of machine and deep learning models to predict maternal mortality risk and aiming in reducing the rate and improving maternal health outcomes in resource-limited settings.

1 Introduction

One major issue in the health care field is maternal mortality, which needs to be addressed about. Women considering the pillars of family and society but these maternal deaths during pregnancy causes great loss to baby, family, society and country. Pregnancy is considered as the most important stage in women's life as it causes major physiological and psychological changes. Pregnant women often die from many complications especially due to lack of information about maternal health care throughout and after pregnancy and these very often takes place in rural areas and low-middle income households in developing countries. In such developing countries Shrivastava et al. (2023), maternal mortality rates are more common among young teenage girls.

Maintaining good maternal health during this time is very important not just for the mother's health but also for better growth and development of fetus. Pregnant women often die from complex conditions such as advanced maternal age, blood disorders and irregular heartbeat and many more. According to many research and studies on this topic, most important factors to be considered during this time are age, heart rate, diastolic blood pressure, systolic blood pressure and total blood pressure and trying to reduce these conditions will help in decreasing maternal and newborn deaths. Around 295000 Manik et al. (2020), women lost their lives in 2017, due to the complications as mentioned above, as per the reports from World Health Organization. Majorly, 94 percent of deaths occurred in rural and sub urban lands. An estimated figure of about 86 percent of deaths occurred in Asia and sub-Saharan Africa in the year 2017 Bogale et al. (2022). Given the fact that these fast-approaching deadlines for reaching the development goals, communities are making lot of effort in supporting these low- and middle-income countries to renew their commitment towards achieving the goals and reducing the maternal deaths.

India, being the largest populous country of the world, always has its drawing attention of globe towards its health profile. Worldwide, the maternal mortality rate decreased from 385 per 100000 live births in 1990 and to 216 per 100000 live births in 2015. Similarly, in India, it ranged from 556 to 174 per 100000 live births in between 1990 to 2015 and by 2015 as a whole country contributed to 15 percent of global maternal deaths Ukrit et al. (2024). As a developing country, India should focus on maternal health not only for maintaining public health but also aligning itself with broader goals that are helping in developing the country, provided by the UN SDG's.

Having considered the importance, this research aims in tackling the maternal mortality issue from a holistic approach of integrating machine learning, a perspective focused on modelling the data of maternal mortality and evolving in research fields and gaining significant popularity in many areas including health care and education.

Research Question: To what extent can machine and deep learning models utilizing IOT-collected data effectively predict maternal mortality risk in rural regions of India?

The document is followed by different sections such as section 2 comprising the information on related works including detailed analysis and review of related papers. Section 3, consists of research methodology and design specification. Section 4 has information on ethical considerations followed. Section 5, implementation of the individual models and hybrid model. Section 6, evaluation of the models. Section 7, Conclusion and future work discussing on outcomes and possible limitations of the study and followed by acknowledgement and references.

2 Related Work

Practice of Predictive and Classification, Machine Learning and Deep learning models in prediction of maternal mortality risk.

Several research and studies have integrated machine learning and deep learning models in the medical field, showing their potential in predictive analytics. They have often provided remarkable results in healthcare, including predicting mortality risk as one of the major works. A medical cyber physical system for predicting maternal health in developing countries was developed using machine learning by Hossain et al. (2024), where research focuses on the potential of integrating MCPS with machine learning to predict outcomes with real time data collection along with cutting edge algorithms, having XGB classifier outperforming other algorithms with higher accuracy, providing real time data driven insights for mother and child care during the period of pregnancy.

There are some risks for pregnant woman during the early stage of pregnancy and these are classified based on the parameters such as age, number of births, alcohol and many other, considering all these a study Mutlu et al. (2023), aimed in predicting the maternal risk and proposed a machine learning model, with decision tree classifier showing great results comparatively. A novel attempt was made to evaluate the maternal dataset and this study Shastri et al. (2021), aimed in decreasing the deaths of women during pregnancy by undertaking prediction procedures via base and meta learners and using a way of more effective nested stacking technique. The system's reliability was evaluated by computing the accuracy and ROC area of algorithms. The model is best suited to detect high or low maternal mortality rates.

Combining number of machine learning models and classifiers in predicting the maternal mortality risk and rates. Several studies incorporated the idea of integrating multiple machine learning techniques, a deep learning-based risk level prediction Ukrit et al. (2024), where in one dimensional Convolutional neural network-based model functions better comparatively. A paper proposed Assaduzzaman et al. (2023), an early prediction of maternal health risk involving improved data pre processing. Out of models such as random forest, XGB, decision tree, Gradient boosting, random forest performed well with hyper parameter tuning. In continuation, similarly a study by Shifa et al. (2024), focused on equipping healthcare professionals with precise tools for early detection and intervention.

Out of nine algorithms incorporated, XG Boost gave better results, but models interpret ability and explain ability continues to be major problem and addressing this issue will give better results. Another paper Manik et al. (2020), focused on maternal mortality classification for health pro motive in Dairi city, Indonesia. Using 149 samples from survey in 2017 and integrating the machine learning classifiers, indicated that the decision tree and navies' are performing good in reducing the mortality rate. Maternal risk analysis can improve prenatal care, mother health and child health and optimize medical resources using machine learning algorithms such as LDA, KNN, QDA, decision tree and many more, a comparative assessment of several effective machine learning classification methods for detecting maternal health risk was proposed by Raihen and Akter (2024).

Applying split validation technique on 800 observations for training and 214 observations as test and a 10-fold cross validation model using SVM, outperformed in the terms of accuracy. Predicting mortality based on maternal health status and health insurance service using homogeneous ensemble machine learning methods was proposed by Bogale et al. (2022). Ethiopia, having the highest number of maternal deaths across Africa, having 68 deaths per 1000 pregnancies. This research aimed in mortality rate by using data collected from Ethiopia demographic health survey from 2011 to 2019, with gradient boosting algorithm performed with higher accuracy than other algorithms.

Research and studies using machines and deep learning and predictive analytics to anticipate the risk of maternal mortality, and these practices will often give good results for evaluating and reducing the probable outcomes. According to reports from the US and Europe, machine learning analysis can investigate and evaluate factors that are little different from the traditional logistic regression model developments and perform better and predict mortality rate. One such study was developed by Shastri et al. (2021), and usage of different of machine learning models, including, XG Boost, logistic regression algorithms, and other conventional methods, to predict the pregnancy mortality rate for five years. AUC values were used to compare predictive performance to explanatory factors.

In this vast growing, with many advancements in the health care fields, maternal mortality rate remains increasingly high in the countries that are still developing and the hybrid and AI models were developed to carry out research by Togunwa et al. (2023), to investigate and interpret health data. Combining the Artificial neural networks and Random Forest classifier algorithm was used and selected based on majority voting techniques and random forest model was performed better as stand-alone model and as an overall hybrid model, performance was improved. Considering that RF performed better than ANN, research conducted by Sheakh et al. (2023), integration of seven ML algorithms to evaluate the performance of model was developed, RF performed better as compared to remaining models but the study had the limitation as primary data and the quality of the data was yet to be improved and collected for developing robust models.

Using both the applications such as predictive analysis and explain-ability for predicting using ML models was proposed by Khan et al. (2020), and the analysis had been conducted in two stages, as one was to predict mortality rate, and another was to classify the problem in evaluating the high or low rate of mortality. The results showed that the predictive ability of the models was not better but the models exhibited the explain ability in a significantly stronger way. Having these considerations of better significance of explain ability another research was proposed by Rahman and Alam (2023), using explainable artificial intelligence-based ML and DL models that predicted mortality rate as high, mid or low and among 13 models considered, gradient boosting model had good results in terms of accuracy.

The usage of AI and ML has increased significantly, especially in the healthcare systems, where a study was conducted to improve the adverse effects of pregnancies by Shrivastava et al. (2023), in the urban regions of India, developing a robust model with cross-sectional study and aimed in predicting mortality rate but was limited to one geographic area with very less data. The ability of machine learning to transform and improving the health-related areas by reducing the work of physician and developing the cost-effective model was proposed by Pawar et al. (2022), which used different algorithms and developed multiple models to support wide range of medical care systems. The leading factors that are contributing to the increased number of mortality rate in women and the main requirement is to provide architecture to decrease the mortality rate and aimed in developing the predictive decision making and the risk assessment system for the sub urban and rural area was proposed by Goel and Ahuja (2023). Different classification models were used, and decision tree showed better results and further interpretation will show more enhanced model for higher-risk pregnancy.

3 Research Methodology and Design Specification

This research aims in predicting the maternal mortality risk rate in women to prevent deaths by providing proper precautions with early detection's and risk assessments followed the Crisp-DM methodology. Extracting meaningful information and data along with finding various patterns in data will be a major benefit for Crisp-DM methodology. Crisp-DM methodology can be used on unstructured and structured data or complex data structures, which makes it flexible to any domain Shastri et al. (2021). This section

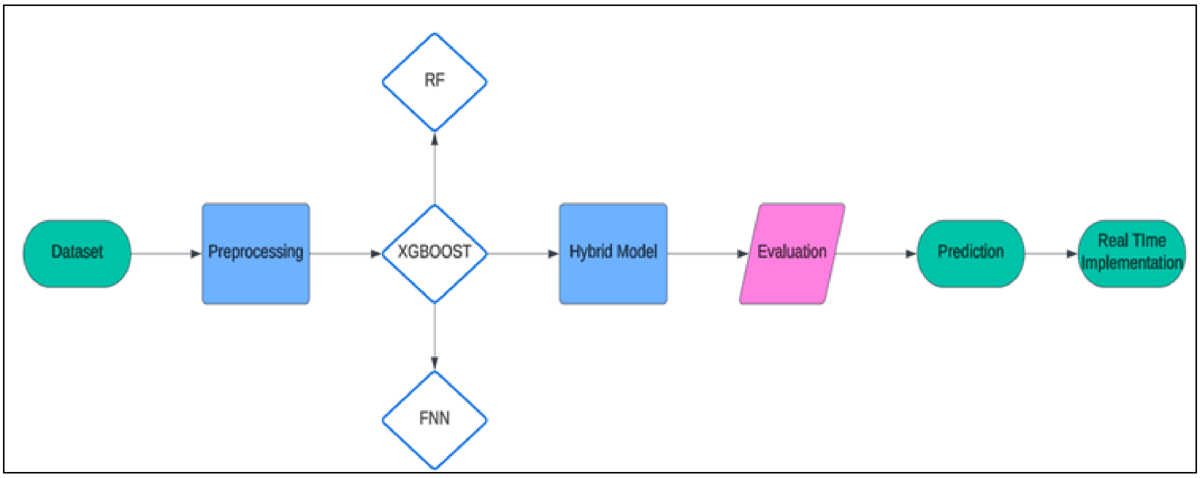

Figure 1: Design Flow

describes research methodology and design specification of the project. Tools used to perform these tasks are detailed in Configuration Manual document.

3.1 Methodology

Prediction of maternal health risk is important to prevent maternal morbidity and mortality. In this section, a clear process of assessing, predicting and comparing the performances of three different machine and deep learning models namely, Random Forest (RF), XG Boost (XGB) and Feed Forward Neural Networks (FNN) models will be trained in predicting maternal health risk levels. In the end, a hybrid model will be developed by combining the three approaches which will potentially enhance the model's accuracy. This following section will provide an in-depth explanation of the methodology and design specification covering data collection, data preparation and pre-processing, modelling and evaluation of results. The following figure 1 shows the visual representation of the methodology design.

3.2 Data Collection

The dataset considered for this research is available on Kaggle in an open dataset's repository. This data set maternal health risk has following seven features as followed, Age: age of the patient, Systolic Blood Pressure: systolic BP of the patient, Diastolic Blood Pressure: diastolic BP of the patient, Blood Sugar: blood sugar of the patient, Body Temperature: body temperature of the patient, Heart Rate: heart rate of patient, Risk Level: this is the target variable that indicates the risk level of the patient whether Low, Medium or High. All the six variables were numerical variables. The data file is saved as csv format, and this makes it easier for formatting and easy reading into python script.

3.2.1 Data Pre-Processing

Initially, everything starts by preparing the data that is converting data into a form where it can be easily trained to build a machine learning model. This leads to the approach of exploratory data analysis (EDA) to understand the characteristics and features of

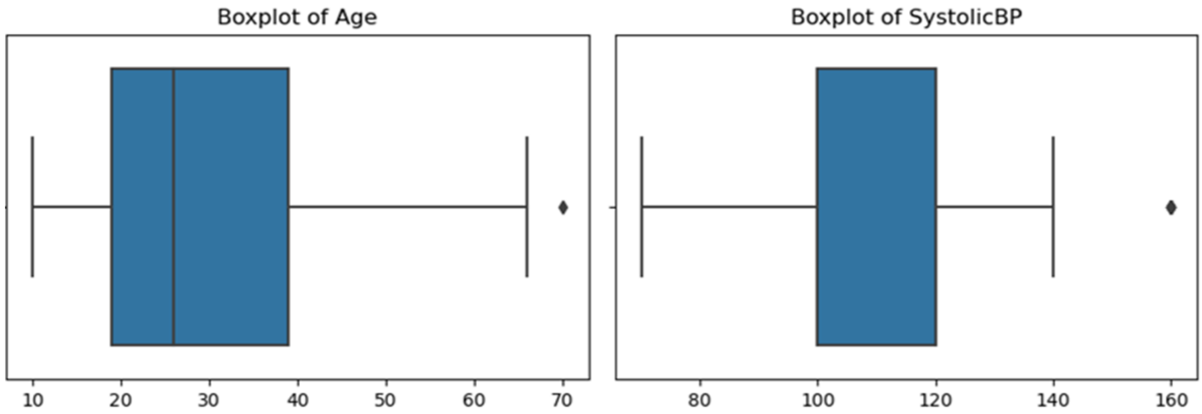

Figure 2: Box plot visualization

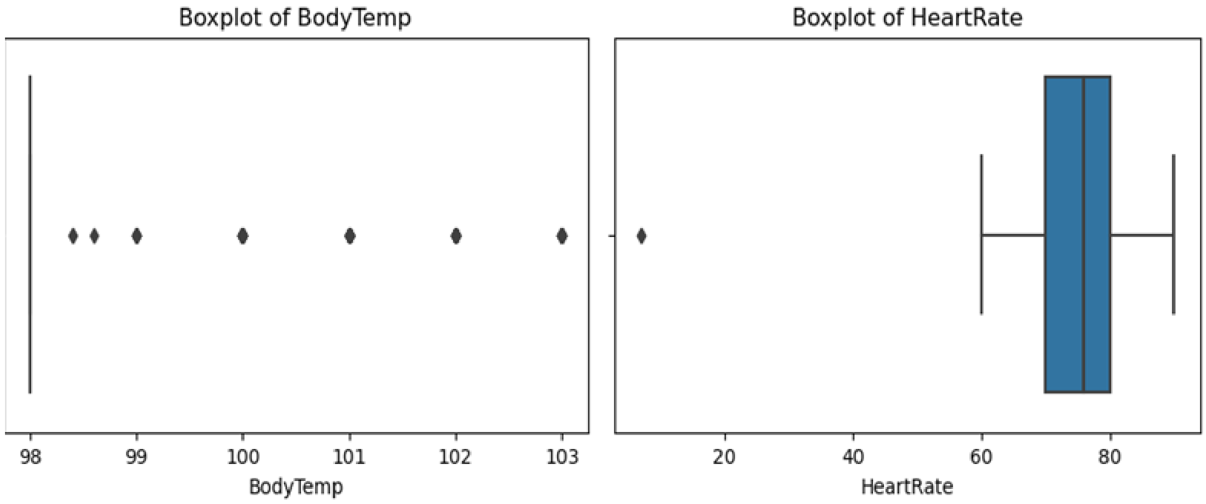

Figure 3: Box plot visualization

dataset. Identifying potential outliers which leads to understanding the range of values for each variable was possible through histogram visualizations where for each feature, graph will be produced to assess their distribution, which will show that some features were skewed while others distributed normally. The histogram visualizations are shown in figures 4 and 5.

Box plots will be developed as shown in the figure 2, to analyze the numerical variables that spread and will detect potential outliers. The correlation heat map will be developed to provide insights into the relations among the variables and feature inter-dependencies and also to identify any potential multi-col linearity, which might possibly influence the stability of the mode and its performance is shown in the following figure 6.

The stronger correlations are showed by values to 1 or -1 and close to 0 will show that there is little or no correlation. The visualization showed that systolic BP and Diastolic BP have a high positive correlation, showing they increase or decrease together. Body Temperature and heart rate show significantly less correlation, with mostly positive or negative values. The correlations and inter-dependencies are clearly shown in the figure 6.

The target variable risk factor being a categorical variable should be transformed into numerical variable for prediction purpose and this was done using Label Encoder. The

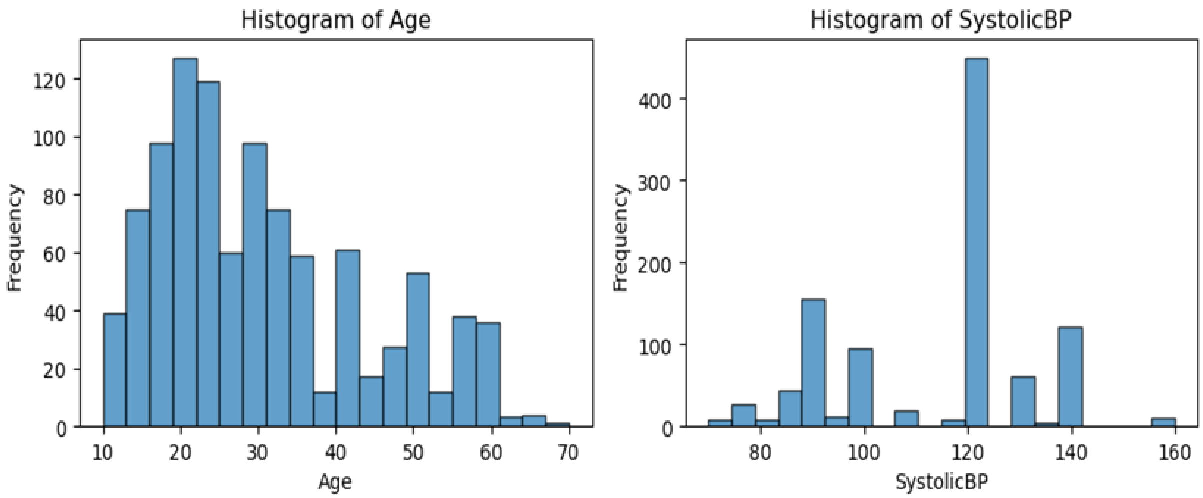

Figure 4: Histogram Plotting

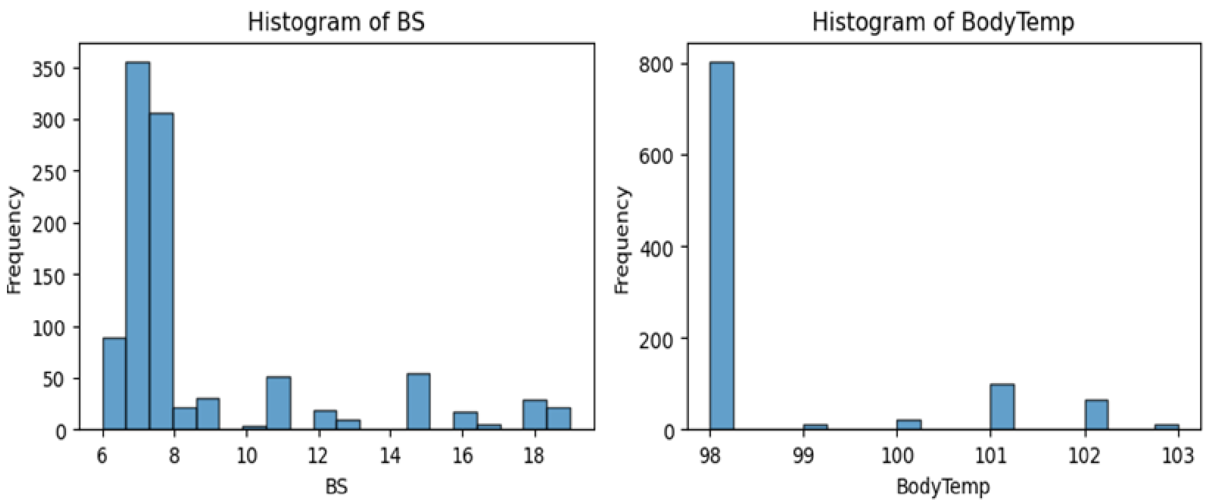

Figure 5: Histogram Plotting

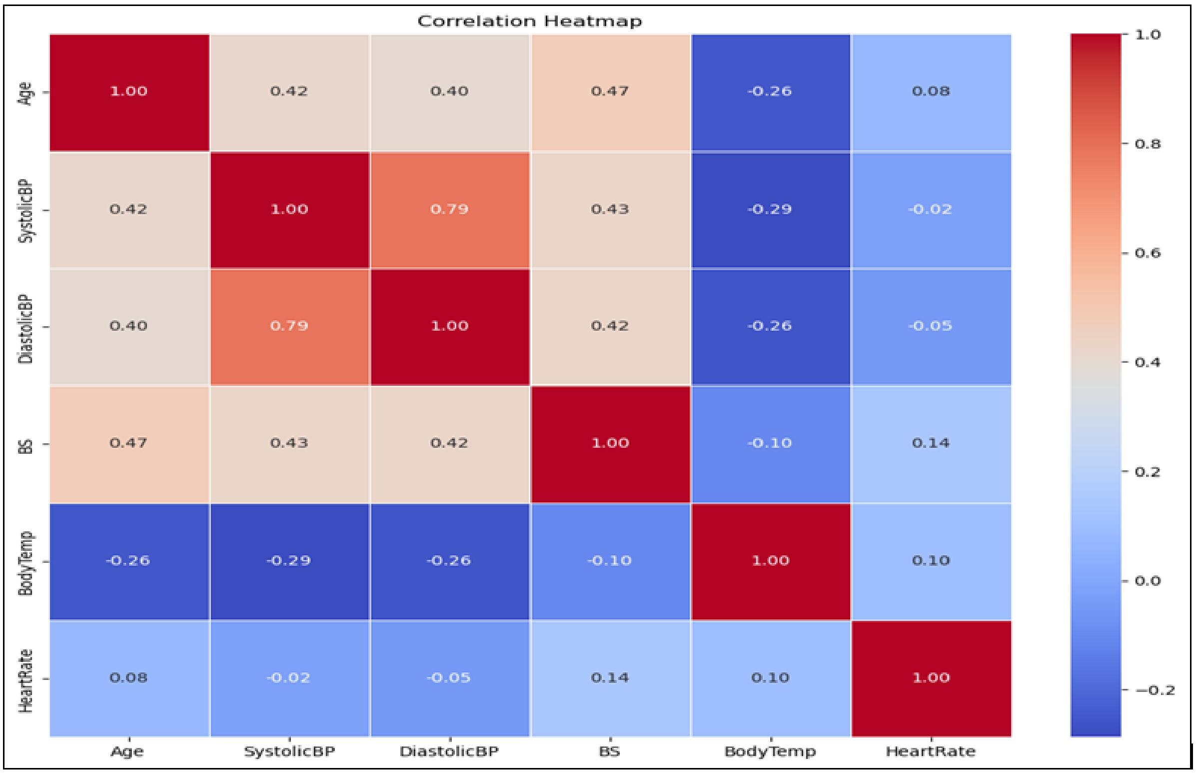

Figure 6: Correlation Heat Map

data set will be divided into training and testing of 70 percent and 30 percent respectively. Having the varying scales of input features, it is necessary to standardize them to bring them all to a common scale. This involves transforming the data such that it has a mean of zero and a standard deviation of one, which is particularly important for models sensitive to feature scaling, like neural networks.

Standardization of features will be performed using the standard scaler, which will make adjustments to values to have a mean of zero and the standard deviation of one. This step of standardization is particularly important for algorithms that are sensitive to input data such as neural networks.

3.3 Data Modelling – Design Specification

3.3.1 Random Forest

Random Forest is a model designed to capture the most complex patterns and decision boundaries in the maternal data set that is considered Togunwa et al. (2023). The algorithm will be using decision trees, where each tree will ask to be trained on the random subset of the variables and samples from the dataset. Random Forest is an ensemble learning method that operates by constructing multiple decision trees during training and outputting the mean prediction of the individual trees. For Random Forest, to optimize the model for its better performance hyper-parameter tuning is critical step. Grid Search CV will be used as a tuning parameter and this exhaustive search over a specific parameter grid will allow to identify the model's best parameters that contribute to the optimal performance.

3.3.2 XG BOOST

Another powerful method known or its efficiency and its high performance is XG Boost machine learning model. It works iteratively by fitting new models to the results of previous models, which will improve the model's accuracy Sheakh et al. (2023). Similarly, XG Boost requires a careful hyper-parameter tuning and parameters such as rate, maximum tree depth, sample ratio will be fine-tuned using Grid search CV, which will minimize the overall model's loss function that is in this study it will be multi-class logarithmic loss. XG Boost will prevent over-fitting and will play important role in enhancing the model's overall consistence performance.

3.3.3 Feed Forward Neural Network

Feed Forward neural network, a deep learning model consisting of multiple dense layers is chose along with machine learning models as a combination of neural networks with classifiers for prediction purpose. Tuning in FNN is difficult because of its complex nature of non-linearity in the model. In this following model development, Kera's Tuner will be used for hyper-parameter optimization.

3.3.4 Hybrid Model

Considering the importance of predicting the risk factor in reducing the maternal mortality, a hybrid model development will be done combining both machine learning classifiers and deep learning neural networks to improve the overall model's performance, accuracy and efficiency in risk assessment. The outputs of models are fed as meta features input to the hybrid model that will be developed and this hybrid model is designed to leverage the strengths of the base models that are integrated and learning from their combined outputs Sheakh et al. (2023). The final evaluation metrics, accuracy will be calculated by comparing the hybrid model's predictions against the testing dataset.

3.3.5 Evaluation Metrics and Interpretation of Results

After hyper parameter tuning, the best models for RF, XG Boost and FNN are selected and trained for model building and these models will be evaluated against test data set using various metrics such as the accuracy defined as the proportions of correct outputs out of all predictions, precision being the proportion of true values among all positive predictions given by the model, recall defined as the ratio of true positive outputs among all proper positive instances. F1 score means of all precision and recall together, giving out a single metric which balances both the metrics. For Interpretation of results, confusion matrices for all the individual models and hybrid model will be plotted. Along with them learning curves will be plotted to visualize the training and validation accuracy scores and in the end a bar plot to compare all the evaluation metrics will be plotted.

4 Ethical Considerations

This study involves data that is collected and analyzed sensitive data, which will raise several ethical considerations. First, the most important is data privacy and data security, especially to protect the identities of individuals and secure and prevent unauthorized access from anywhere and anyone. Having the use of machine learning models in healthcare introduces to the risk of datasets to avoid reinforcing the existing disparities in the healthcare access and outcomes. The possible impact of misclassified predictions on patient care should also be considered in emphasizing the need of these models that are to be used in decision-making tools rather than finding the replacements for professional medical decisions. One more important factor to be considered is having informed consent obtained directly from all individuals and this study should follow all relevant regulations and ethical guidelines, including those from different data protection authorities.

5 Implementation

This section of the study dwells deep into explanation of implementation of the proposed methodology, to assess, predict and compare the performances of different machine learning models namely, Random Forest, XG Boost, and Feed Forward Neural Network that was used to build a model to predict the maternal mortality risk levels. In addition to this, a combined hybrid model was developed to potential enhance the prediction accuracy. This section will in detail explain data pre processing, model training, model optimization, evaluation and construction of hybrid model. The following figure 7, picture's the flow and design of implementation.

The implementation starts by importing the necessary libraries, including Pandas, NumPy, and other machine learning and deep learning libraries such as Grid Search, Kera's for optimization purpose, and other libraries like mat plot for visualization purposes. The dataset used is in the CSV format is loaded using Pandas. It is very important mention correct file path so that correct data is been read without any errors. The dataset contains several features, which are: Age of the patient, systolic Blood pressure, diastolic blood pressure, Blood sugar levels, Body temperature of patient, heart rate and health risk level of the patients.

As a next step of this process, exploratory data analysis was done with help of histogram and box plot visualizations and the data was checked for spread and detection of potential outliers, data distribution, skewing of the data as shown in the figure. Along with this a correlation heat map was generated to analyze the inter-dependencies of the features and relationship patterns between the numerical variables as shown in the figure. Having the dataset loaded and ready, the initial step will be to check for all the variables and separate them from categorical values, in this case, the target variable 'Risk Level' which will be indicating the levels of maternal health risk is a categorical variable and this was converted to numerical value by using Label Encoder from Scikit-learn library. This is will assign a unique integer number to each category during the conversion process.

The next step was to split the datasets into training and testing sets to evaluate the performance of the models. This was done using train test split function from the Scikit-learn library by randomly dividing the dataset into two parts. The training set was 70 percent of the entire data and the testing was 30 percent of the entire dataset, where the training data set was used to train the model and test data to evaluate the model performance. This split make sure that the models will test on unseen data which will provide an estimate of their generalization capability. This will be followed by the process of standardization of data, especially when models are sensitive to the scale of the data. In this step, the scaling of features is done to make sure that each feature will equally contribute the learning process of the model and it helped in preventing the features

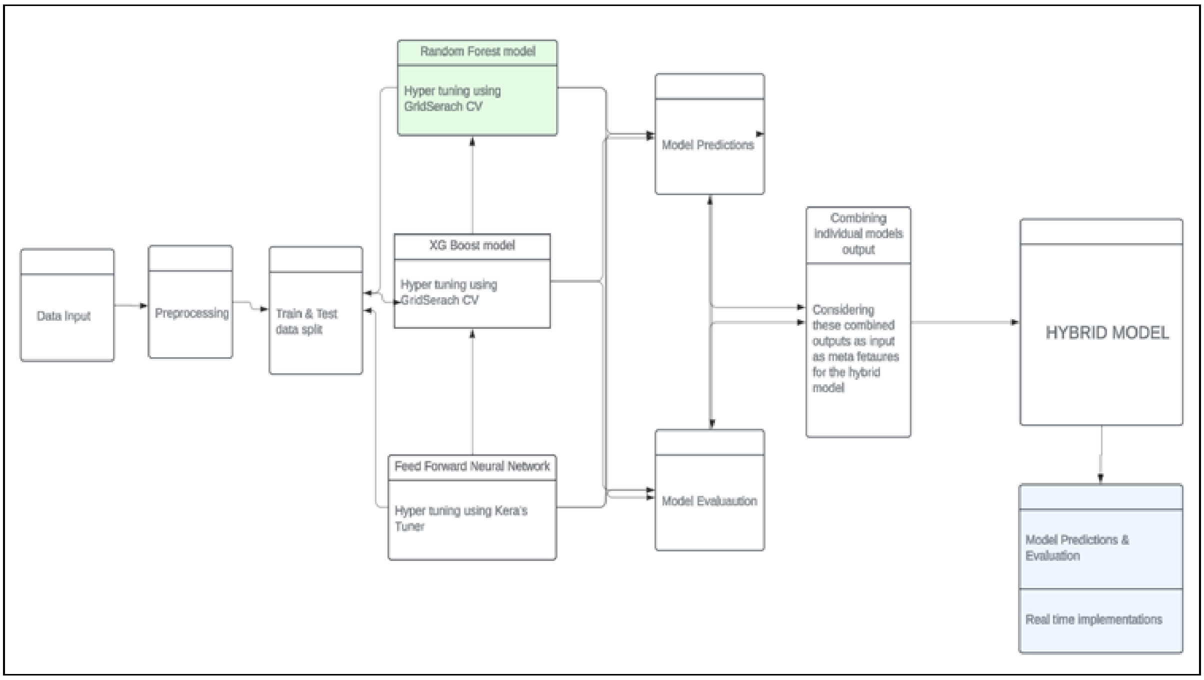

Figure 7: Implementation Framework

were dominating the model's performance and this was done using standard scaler from Scikit-learn library to standardize both training and testing data.

5.1 Random Forest Model

Random Forest is an ensemble method which will construct multiple decision trees during the training and gives outputs in mode for each class of individual trees Sheakh et al. (2023). In this scenario, to optimize the model's performance Grid Search CV was used to tune the performance of the model and it exhaustively searched for specified parameter values to find the best combination for improving the model's performance. The key hyper-parameters that were included in the random Forest model were, n estimators which were the number of trees in the forest, maximum depth of each tree, minimum sample split was the number of samples that were required to split an internal node and minimum sample leaf were the number of samples required to be a leaf node. The possible values were defined for the above-mentioned hyper parameters and Grid search CV was defined to evaluate each combination provided using 3-fold cross-validation technique and selected the best performing model based on the accuracy. The model is trained using the training data and the best set of hyper parameters was assigned to best rf and the tuned Random Forest accuracy was calculated.

5.2 XG Boost Model

Amongst the gradient boosting implementations, XG Boost is powerful and efficient. Similarly, like Random Forest, XG Boost also requires hyper-parameter optimization to achieve best performance Shastri et al. (2021). The key tuning parameters considered for this implementation were estimators, the total number of boosting rounds, maximum depth of the trees, to prevent over-fitting the step size of shrinkage is measured by learning rate, sub samples, the fraction of samples that were considered as overfilling for the individual base learners and the sample by tree is used as the fraction of features that will be used for fitting the individual base learners. Finally, Grid Search CV was used to perform the hyper-parameter tuning for XG Boost by defining the possible values and evaluated the combinations using cross-validation technique and the best estimator was assigned to best of XGB.

5.3 Feed Forward Neural Network Model

Similarly, For Feed Forward Neural Network, hyper parameter tuning was done using Kera's tuner, which involves selecting the optimal architecture and possible best learning parameters which includes, the total number of units specified in each layer, what are the number of layers given and how much is the learning rate considered. In this implementation, Kera's tuner was used because of its capability of evaluating different model architectures and hyper parameters on the training data and selects the best performing configuration based on validation accuracy, as the primary objective to optimize the overall model performance. Maximum trials are the maximum number of different hyper parameters combinations that was tried and executions per trial are the number of times each parameter combination was trained during optimization. The best set of parameters found is stored in best hyper parameters and these instances of parameters are used to sample different configurations.

The parameters which were optimized was, the number of layers, number of units per layer and the learning rate. The number of units in the first dense layer will vary between 32 to 256 in the steps of 32 with the activation function ReLu and soft max for the output layer which is suitable for multi-class classification with having 3 units. The categorical cross-entropy loss function was used, appropriate for multi-class classification tasks, along with the Adam optimizer, known for its efficiency and performance in deep learning applications with learning rat e ranging between 1e-3 or 1e-4.

It starts by converting the training labels to categorical values which is suitable for classification and initiates the search using the training data. The model was trained up to 50 epochs with having 30 percent validation split. The best set of hyper parameters was used in building the model and trains the model using training data for 50 epochs and a batch size of 10. Predictions for the test data set was also produced. The argmax function was used to convert the output form Soft Max layer to class labels by considering the maximum values of each prediction.

The accuracy was of the model was calculated on test set by comparing predicted labels with the true labels. Finally, function to evaluate the model's performance was done based on accuracy, precision, recall and F1 score, which will provide a comprehensive view of model. Each model undergoes hyperparameter tuning using Grid Search CV or Kera's Tuner to find the best hyperparameters for optimal performance. The tuned models are then used to predict on the test set, and their accuracies are evaluated and reported. This approach ensures that each model is optimized before combining them in the hybrid model.

5.4 Hybrid Model

5.4.1 Hybrid Model Construction

The hybrid model will aim in combining the strengths of all the three individual models Random Forest, XB Boost and Feed Forward Neural network to improve the overall performance. The approach used here involves stacking which is the outputs of individual models are combined and fed into a meta model. This meta model in this study is another neural network which learns to make final prediction based on these provided combined outputs.

5.4.2 Hybrid Model Design

The hybrid model's architecture includes input layer considering the input features that matches the combined outputs of all individual models built, a hidden layers having two dense layers with 64 and 32 units respectively using ReLu activation function and these layers was combined with meat features generated by individual layers and the final output dense layer with 3 units and a Soft Max activation function suitable for a three-class classification problem. The model will then be compiled with Adam optimizer, categorical cross-entropy loss and accuracy evaluation.

5.4.3 Training the Hybrid Model

The hybrid model was trained on a new dataset created by stacking the outputs of the individual models considered as meta features for each training instance. This meta-dataset allows the hybrid model to learn the relationship between the predictions of the individual models and the true class labels. The predicted class probabilities for training and testing datasets for RF, XGB and FNN defined. Combining these predictions into a single array for training and testing datasets and this array becomes the input into hybrid model which makes it effective way of learning from ensemble predictions. The training labels were converted to categorical format and initializes the hybrid model with an input dimension matching the number of combined meta-features with the batch size of 10 and 50 epochs. The hybrid model than generated the predictions for the test data and finally accuracy was computed for the hybrid model.

6 Evaluation

6.1 Evaluation Metrics

The evaluation metrics are calculated using accuracy, precision, recall and f1-score for all three individual models and also for the hybrid model. The following figure 8 shows the calculated results of all models. Individually RF and XGB performed almost similar when compared with FNN, it performed low and when these outputs were combined and built a hybrid model, which outperformed and improved the overall accuracy and stability of the model.

6.2 Confusion Matrix for Models

The final evaluation and comparison were calculated based on accuracy, precision, recall and f1 score and comparison was done using generating confusion matrix, learning curves

MODEL	ACCURACY	PRECISION	RECALL	F1-SCORE
RF	0.800000	0.801026	0.800000	0.799268
XGB	0.803279	0.804707	0.803279	0.802981
FNN	0.737705	0.744310	0.737705	0.733749
HYBRID	0.816393	0.818208	0.816393	0.816274

Figure 8: Results

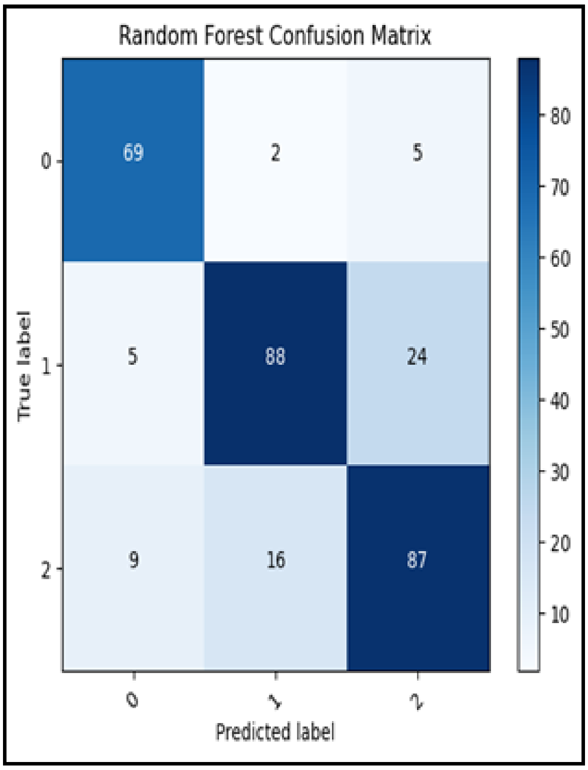

Figure 9: RF Model Confusion Matrix

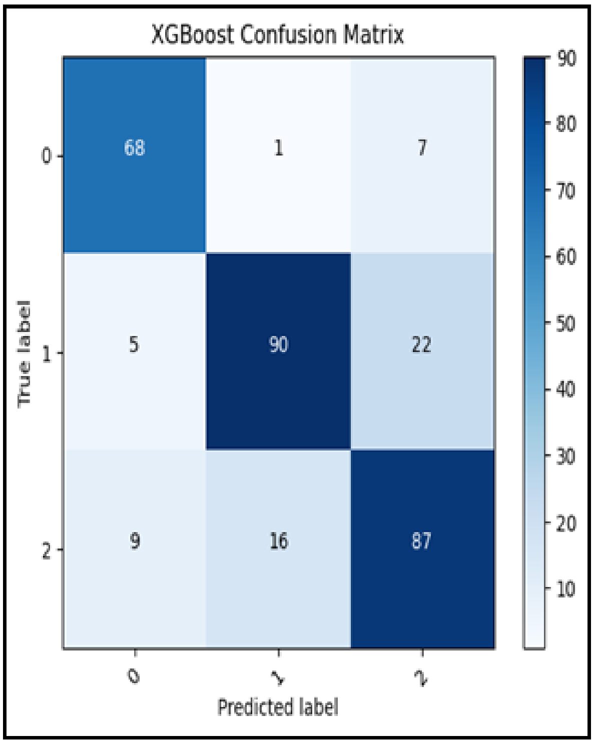

Figure 10: XG Boost Confusion Matrix

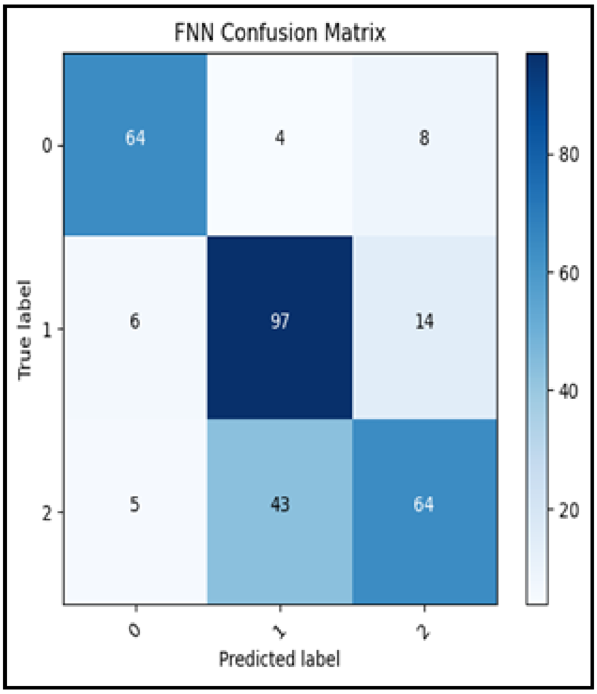

Figure 11: FNN Confusion Matrix

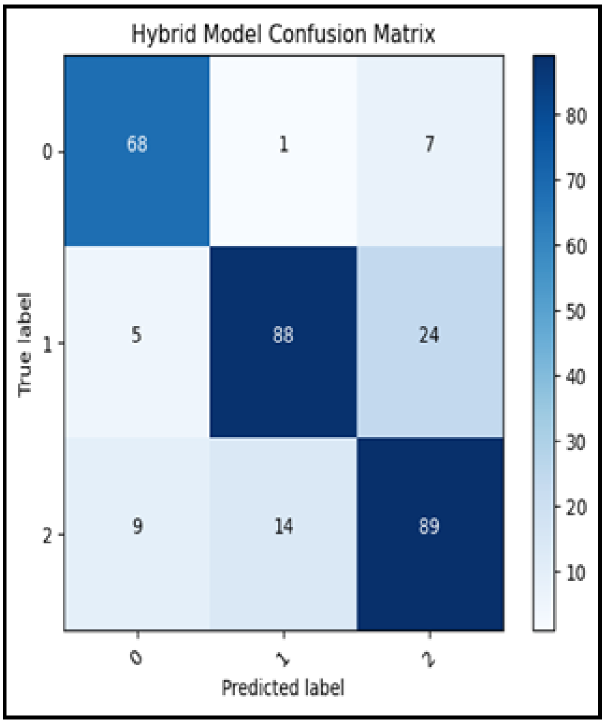

Figure 12: Hyvrid Model Confusion Matrix

and bar plot for comparing evaluating metrics in the end. Initially, the accuracy was calculated for all the models that was developed after hyper-parameter optimizations. Finally, all other metrics were calculated at the end of model building and was printed and following table will show the results.

In the Random Forest, the confusion matrix showed in the figure 9 explains that 69 high-risk instances were predicted correctly and miss-classification of low risk was 1 and miss-classification of mid risk was 6 instances. For the low risk, the correctly classified low risk 88 instances, 3 misclassified as high and 17 as mid risk misclassified. Finally for mid risk, there 87 instances correctly classified followed by 3 and 9 misclassified instances for high and low risk respectively. Similarly, for XG Boost model, figure 10 shows, the high risk had 68 instances classified properly with having 1 and 5 wrong classifications of low and mid risk each separately. Same way for low risk, there were 90 correct classifications with having 4 and 10 miss-classifications of high and mid risk respectively and for mid risk there were 87 correctly classified instances with 3 and 9 wrong classifications for high and low risks. This was similar for Feed Forward model as shown in the figure 11, with having 64, 97 and 64 correct classification instances for high, low and mid risks respectively and followed same for hybrid model classified, 68, 88 and 89 instances correctly for high, low and mid risks separately, is shown in the following figure 12 showing the plotting of all correct predictions as diagonal values and misclassified instances as off-diagonal values.

6.3 Learning Curves and Bar Plot

6.3.1 Learning Curves plotting for Random Forest Model

The red line represents the training score and has its growth high close to 1.0 indicating that the model fits very well with training data. The green line shows the cross-validation

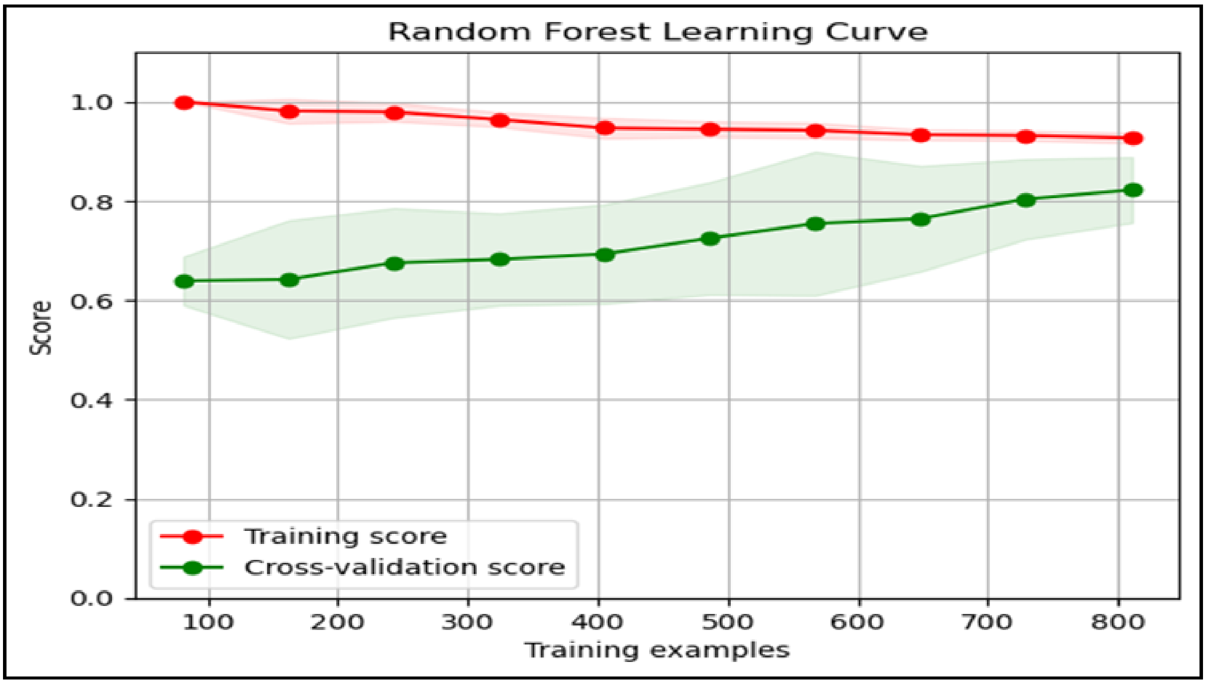

Figure 13: Learning Curve for Random Forest Model

score starts with lower score but gradually the curve increases as when the number of training samples increases. There might be chances of over-fitting because initially the training score was more compared to cross-validation score but when the training size was increased the score for cross-validation curve increased gradually, improving the model's performance and is plotted in the figure 13.

6.3.2 Learning Curves plotting for XG Boost Model

Similar to the Random Forest, the training score for XG Boost remains high, indicating good fitting to the training data. The cross-validation score starts lower but shows a consistent improvement as the number of training examples increases. Like the Random Forest, there is an initial gap between the training and cross-validation scores, suggesting over-fitting. The improvement in cross-validation score is seen as when training data size is increased as shown in the figure 14.

6.3.3 Learning Curves plotting for FNN Model

The training accuracy and validation accuracy is shown for 50 epochs. The accuracy of training improves consistently and stabilizing over the time. The validation accuracy fluctuates in the beginning but improves as the training size is increased as plotted in the figure 15.

6.3.4 Learning Curves plotting for Hybrid Model

The training accuracy is stable with minor fluctuations while validation shows more variance however, the validation accuracy is relatively close to training accuracy as visualized in the figure 16.

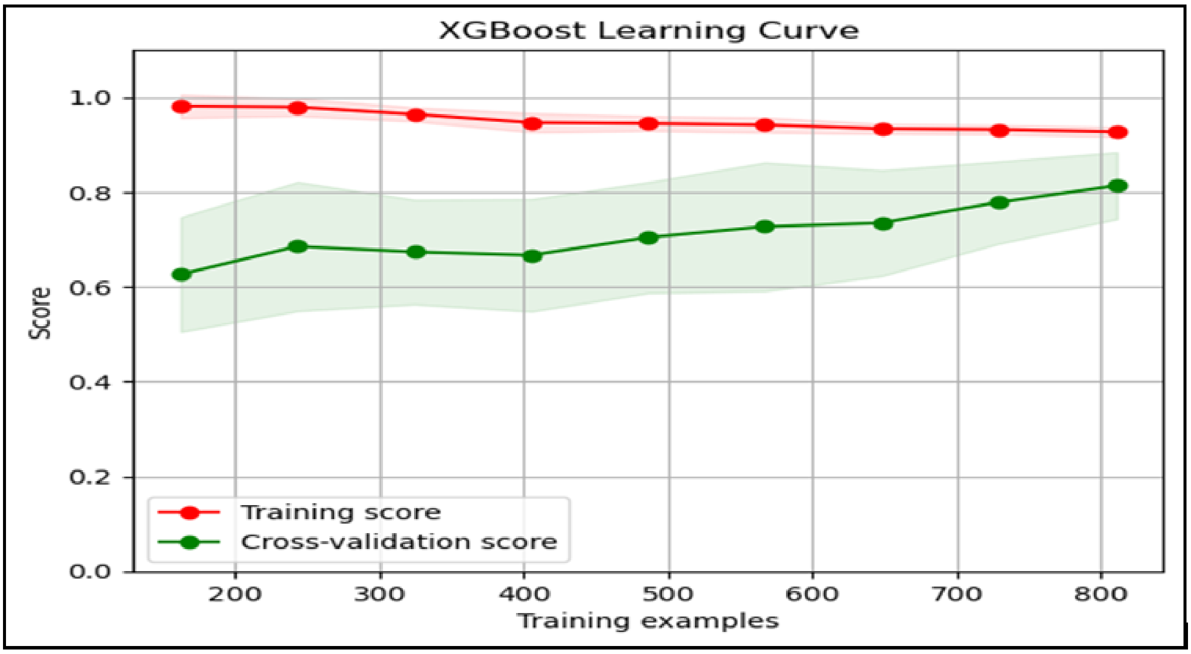

Figure 14: Learning Curve for XG Boost Model

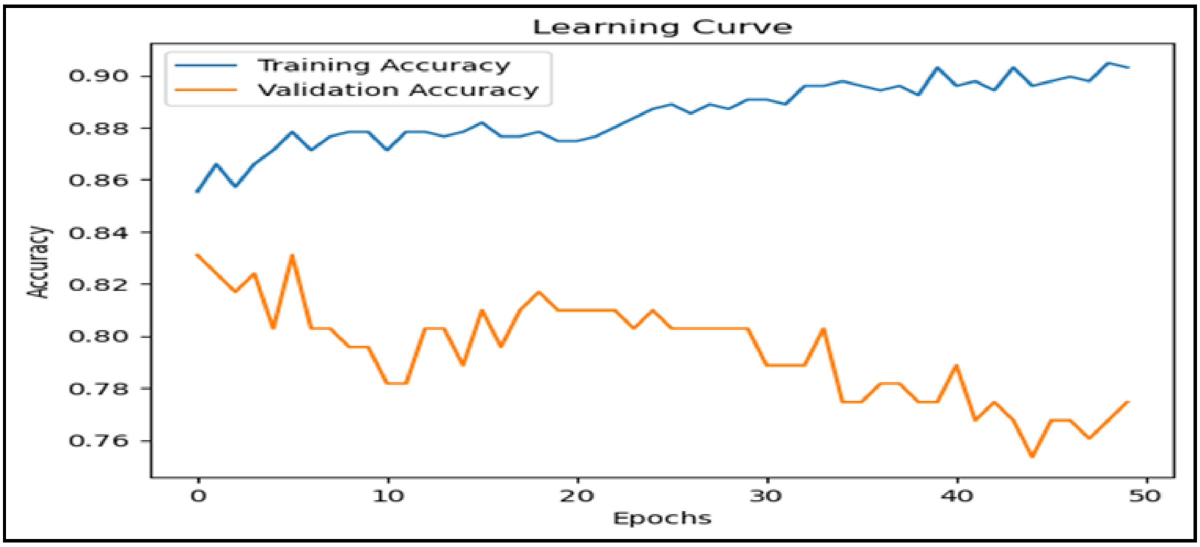

Figure 15: Learning Curve for FNN Model

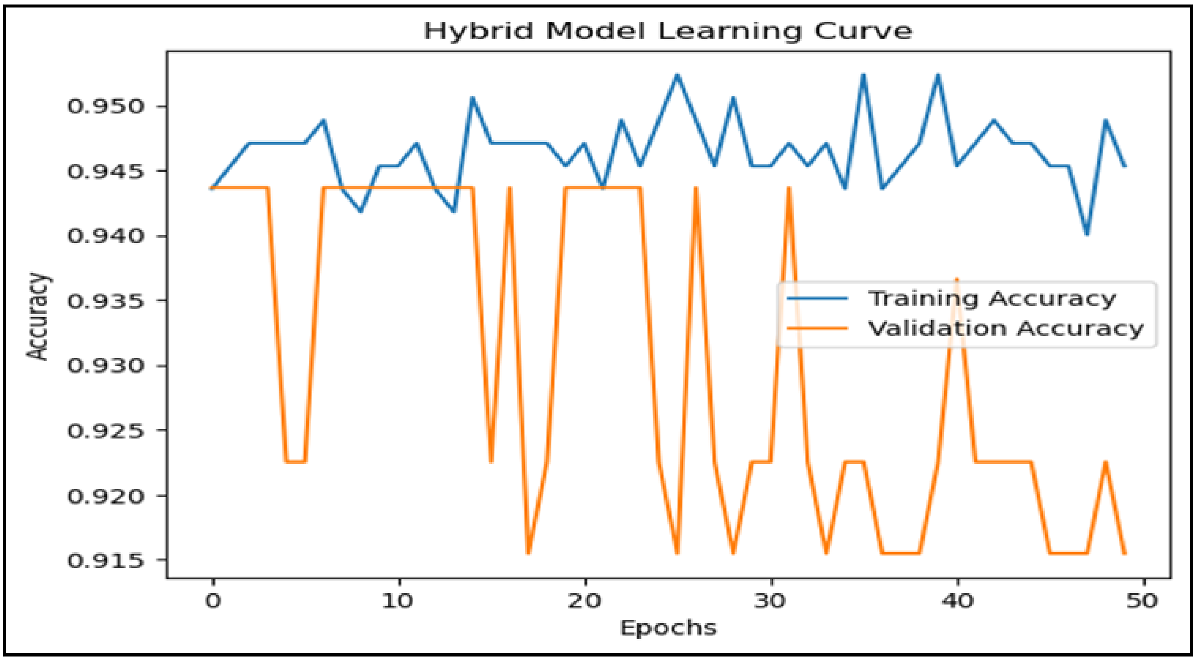

Figure 16: Learning Curve for Hybrid Model

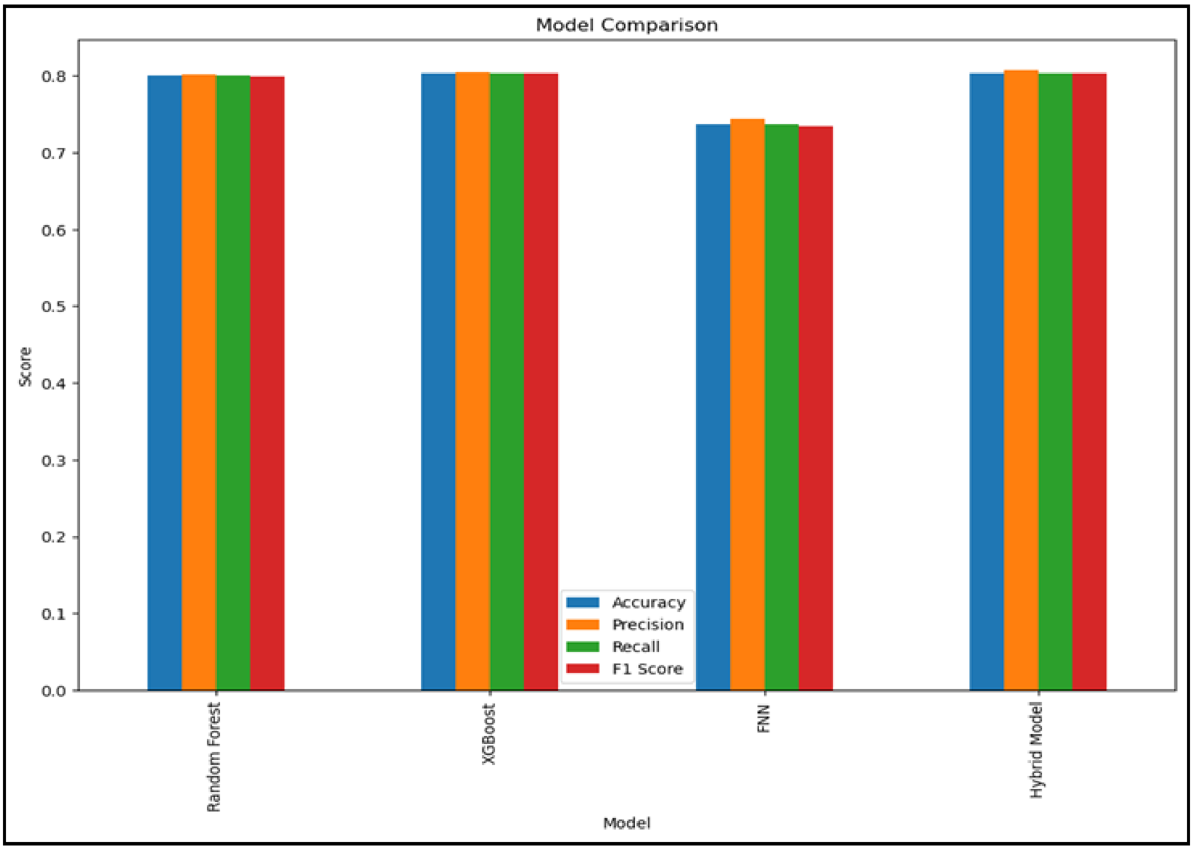

Figure 17: Bar Plot Visualization

6.3.5 Final Comparison with Bar Plot Visualization

The bar chart compares four metrics (Accuracy, Precision, Recall, F1 Score) across four models (Random Forest, XG Boost, FNN, Hybrid Model). Random Forest and XG Boost: Scores almost similar across all metrics. FNN (Feed Forward Neural Network), which is lower compared to the other models. Hybrid Model, outperformed all other models comparatively, resulting in balanced performance and is shown clearly in the figure 17.

7 Conclusion and Future Work

This study shows the strengths of integrating machine learning and deep learning models to predict maternal mortality risk, especially in rural and suburban regions of India, having always limited access to quality healthcare. The XG Boost model, along with Random Forest model, provided the best predictive performance as individual models, hybrid model that was developed offered a balanced approach, by improving the overall prediction stability and performed well. These results and findings suggest that with further refinement such models can be instrumental in aiding with early detection and intervention and ultimately focusing on reducing the maternal mortality rates. However, study also highlights the importance of addressing the possible ethical considerations as the data is sensitive and also prevent algorithmic bias.

As apart of future works and other researches should focus on addressing the limitations in this study, particularly, the fact that the data set used is limited to one demographic region and collecting more quality, quantity and diverse datasets is important Shrivastava et al. (2023). Expanding the data set and including larger regions will enhance model's performance and also reduce the risk of bias in terms of model building and predictions and in terms of healthcare, it will help them understand on what basis or need to provide healthcare at proper time, understand all the patterns of different regions and come up with better ways to prevent deaths during pregnancy.

One more area to focus as part of future works is integrating IoT devices for real-time data collection and monitoring, which will significantly improve the accuracy and timeliness of predictions. Including user friendly interfaces will play a major role in practical implementation. Finally, an interdisciplinary collaboration between the data experts and healthcare professionals will be essential to make sure that these technological advancements will translate into better improvements in maternal health outcome, which includes the ongoing evolution and refinement of models as new data and insights become available.

8 Acknowledgement

The author is grateful to the Professor. John Kelly, for guidance and supervision throughout this research project design and implementation, and author is thankful to her family and friends in providing moral support through the process.

References

Assaduzzaman, M., Al Mamun, A. and Hasan, M. Z. (2023). Early prediction of maternal health risk factors using machine learning techniques, *2023 international conference for advancement in technology (ICONAT)*, IEEE, pp. 1–6.

Bogale, D. S., Abuhay, T. M. and Dejene, B. E. (2022). Predicting perinatal mortality based on maternal health status and health insurance service using homogeneous ensemble machine learning methods, *BMC Medical Informatics and Decision Making* **22**(1): 341.

Goel, P. and Ahuja, S. (2023). Different machine learning prediction models using pregnancy risk prediction systems, *2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT)*, IEEE, pp. 1–6.

Hossain, M. M., Kashem, M. A., Nayan, N. M. and Chowdhury, M. A. (2024). A medical cyber-physical system for predicting maternal health in developing countries using machine learning, *Healthcare Analytics* **5**: 100285.

Khan, Y. A., Abbas, S. Z. and Truong, B.-C. (2020). Machine learning-based mortality rate prediction using optimized hyper-parameter, *Computer Methods and Programs in Biomedicine* **197**: 105704.

Manik, H., Siregar, M. F. G., Rochadi, R. K., Sudaryati, E., Yustina, I. and Triyoga, R. S. (2020). Maternal mortality classification for health promotive in dairi using machine learning approach, *IOP conference series: materials science and engineering*, Vol. 851, IOP Publishing, p. 012055.

Mutlu, H. B., Durmaz, F., Yücel, N., Cengil, E. and Yıldırım, M. (2023). Prediction of maternal health risk with traditional machine learning methods, *Naturengs* **4**(1): 16–23.

Pawar, L., Malhotra, J., Sharma, A., Arora, D. and Vaidya, D. (2022). A robust machine learning predictive model for maternal health risk, *2022 3rd international conference on electronics and sustainable communication systems (ICESC)*, IEEE, pp. 882–888.

Rahman, A. and Alam, M. G. R. (2023). Explainable ai based maternal health risk prediction using machine learning and deep learning, *2023 IEEE World AI IoT Congress (AIIoT)*, IEEE, pp. 0013–0018.

Raihen, M. N. and Akter, S. (2024). Comparative assessment of several effective machine learning classification methods for maternal health risk, *Computational Journal of Mathematical and Statistical Sciences* **3**(1): 161–176.

Shastri, S., Kour, P., Kumar, S., Singh, K., Sharma, A. and Mansotra, V. (2021). A nested stacking ensemble model for predicting districts with high and low maternal mortality ratio (mmr) in india, *International Journal of Information Technology* **13**: 433–446.

Sheakh, M. A., Tahosin, M. S., Hasan, M. M., Islam, T., Islam, O. and Rana, M. M. (2023). Child and maternal mortality risk factor analysis using machine learning approaches, *2023 11th International Symposium on Digital Forensics and Security (ISDFS)*, IEEE, pp. 1–6.

Shifa, H. A., Mojumdar, M. U., Rahman, M. M., Chakraborty, N. R. and Gupta, V. (2024). Machine learning models for maternal health risk prediction based on clinical data, *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)*, IEEE, pp. 1312–1318.

Shrivastava, R., Singhal, M., Gupta, M., Joshi, A. et al. (2023). Development of an artificial intelligence–guided citizen-centric predictive model for the uptake of maternal health services among pregnant women living in urban slum settings in india: protocol for a cross-sectional study with a mixed methods design, *JMIR Research Protocols* **12**(1): e35452.

Togunwa, T. O., Babatunde, A. O. and Abdullah, K.-u.-R. (2023). Deep hybrid model for maternal health risk classification in pregnancy: synergy of ann and random forest, *Frontiers in Artificial Intelligence* **6**: 1213436.

Ukrit, M. F., Jeyavathana, R. B., Rani, A. L. and Chandana, V. (2024). Maternal health risk prediction with machine learning methods, *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*, IEEE, pp. 1–9.