



## Machine learning-based maternal health risk prediction model for IoMT framework

Subhash Mondal, Amitava Nag, Anup Kumar Barman and Mithun Karmakar\*

Department of Computer Science and Engineering, Central Institute of Technology Kokrajhar,  
Kokrajhar, Assam, India

E-mail/Orcid Id:



SM,  [ph22cse1001@cit.ac.in](mailto:ph22cse1001@cit.ac.in),  <https://orcid.org/0000-0002-4203-8467>; AN,  [amitava.nag@cit.ac.in](mailto:amitava.nag@cit.ac.in),  <https://orcid.org/0000-0003-4408-7307>; AKB,  [ak.barman@cit.ac.in](mailto:ak.barman@cit.ac.in),  <https://orcid.org/0000-0002-5697-0431>; MK,  [m.karmakar@cit.ac.in](mailto:m.karmakar@cit.ac.in),  <https://orcid.org/0000-0003-0683-88000000-0003-4408-7307>

### Article History:

Received: 15<sup>th</sup> Apr., 2023

Accepted: 28<sup>th</sup> Jul., 2023

Published: 30<sup>th</sup> Aug., 2023

### Keywords:

Maternal Health Risk,  
Internet of Medical  
Things (IoMT),  
Prediction Model,  
Exploratory Data  
Analysis (EDA),  
Android-based  
Application, Random  
Forest Classifier

**Abstract:** The Internet of Things (IoT) is vital as it offers extensive applicability in various fields, including healthcare. In the context of the risk level during pregnancy, to monitor and predict abnormalities, IoT devices provide a means to collect real-time health data, enabling continuous monitoring and analysis in the Internet of Medical Things (IoMT) environments. By integrating IoT devices into the system, crucial signs such as Heart Rate (HR), Systolic and Diastolic Blood Pressure (BP), Fetal Movements (FM), and Temperature (T) can be tracked remotely and non-invasively. This allows for the timely detection of abnormalities or potential risk factors during pregnancy, empowering healthcare professionals to intervene proactively and provide personalized care. This research focuses on developing a system for observing and predicting the maternal risk level in the IoT environment, mainly in remote areas. The goal is to improve maternal health and reduce maternal and child mortality rates, a significant decline according to United Nations targets for 2030. The research utilizes analytical tools and Machine Learning (ML) algorithms to analyze health data and risk factors associated with pregnancy. The acquired dataset contains various risk factors categorized and classified based on intensity. After comparing different ML models' experimental results, Exploratory Data Analysis (EDA) approaches to determine the most effective risk factors. The fine-tuned Random Forest Classifier (RF) achieves the highest accuracy of 93.14%. An Android-based application has also been developed to deploy the prediction model to determine risk levels based on the different parameters.

### Introduction

Maternal Health Risk (MHR) refers to potential health problems arising during pregnancy, childbirth, and postpartum. According to WHO, there are around 280,000 fatalities of women due to pregnancy complications, which means a woman dies approximately every two minutes (WHO, 2023). The various factors increase the mortality rate of maternal women and childbirth, including the shortage of doctors and nurses and the localization, time, and distance (Redondi et al., 2013). According to WHO's report in 2020, around 800 women die daily due to poor resources and care

(Castillejo et al., 2013). Despite recent technological advances, the rate of maternal death is decreasing, making it difficult to ensure both the mother's and child's safety during pregnancy. Pregnancy-related risks can be reduced in this scenario by anticipating complications and taking precautions.

Some studies have been conducted in recent years to predict certain risks that can occur during pregnancy and to predict the birth method best suited to mothers' pregnancy characteristics. For example, Pereira et al. (2015) used different supervised ML algorithms to predict the best



delivery method among vaginal, cesarean, forceps, and vacuum delivery. In another study, Chen et al. (2011) used a Neural Network (NN) and Decision Tree (DT) algorithm to predict the factors associated with preterm birth. Similarly, Rawashdeh et al. (2020) used Random Forest (RF), DT, K-Nearest Neighbors (KNN), and NN to predict the risk of premature birth. For different data types, different Machine Learning techniques are used, with varying results and performance.

This research study focuses on deploying the ML classifiers prediction model that determined maternal time frame health risk. Initial, five ML classifiers, namely RF, DT, KNN, Logistic Regression, and Support Vector Machine, were deployed after performing some data preprocessing techniques on the acquired dataset consisting of 1014 instances and six related factors that contribute to determining the "Risk Level" as target outcomes in multiclass classification in the First Stage. In the second stage of the prediction model, an immense data analysis approach was performed on the entire feature levels by considering the Exploratory Data Analysis (EDA) techniques in multifold to decide the more contributing features that predict the outcome level. The best-performing RF model is deployed on the processed dataset after eliminating the noncontributing feature using EDA. Under the best configurable test condition, the processed RF model performed well with an improved accuracy of 91.18%. The hyper-parameter tuning approach was applied using the Grid Search CV to derive the best estimator values corresponding to each parameter. The best hyper-parameterized RF model was employed to tune the experimental results under the same test condition and achieved the highest accuracy of 93.14%.

### Motivation

Predicting MHR aims to improve the overall health of pregnant women and their babies. MHR can occur during pregnancy, childbirth, and the postpartum period. However, it is most prevalent during pregnancy when women are at a higher risk of developing health issues, which can lead to miscarriage and death in certain circumstances (Hussain et al., 2014). By identifying and assessing the potential health risks early on, healthcare professionals can take measures to prevent, manage, or treat these conditions.

Predicting health risks can also help healthcare systems to allocate resources more effectively.

Healthcare providers can prioritize their care by identifying women at higher risk. It can also empower pregnant women with information about their risk factors and allow them to make informed decisions about their health.

Overall, the motivation behind predicting maternal health risks and implementing it as an Android application is to enhance pregnant women's health and well-being, reduce complications, and improve outcomes for both mother and babies.

The study aims to develop an Android application that integrates with IoT devices, such as wearable sensors and remote monitoring systems, to predict and mitigate maternal health risks that arise during pregnancy. The article mainly focuses on the following:

- To introduce an IoT-based framework that is capable of monitoring maternal health. The medical sensors/devices collected data samples (blood pressure, body temperature, heart rate, etc.) that are directly fed into machine learning models for the risk prediction of maternal health.
- To create and deploy the ML model on an Android-based application to generate an emergency alert and medical reports to the user, their relatives, and doctors.
- To perform feature selection via the Exploratory Data Analysis (EDA) approach to decide the important and relevant factors contributing to maternal health risk prediction.

### Related works

This section demonstrates a few related kinds of literature conducted before using approaches like Neural Networks (NN), ML classifiers, and the ensemble technique to combine the different architectures for predicting maternal health risk factors. Some of the studies focus on monitoring systems during pregnancy time.

Ali Raza et al. (2022) proposed an ensemble method, BiLTCN that combined the NN-based BiLSTM, Temporal Convolutional Network, and Decision Tree as a classifier using the clinical dataset of 1218 instances collected by the IoT-enabled system. The proposed system observed results after balancing using SMOTE with an average accuracy of 88%. Also, they applied feature selection techniques and used SVM along with BiLTCN, claiming 98% accuracy on the reduced feature model.

Ahmed et al. (2020) executed research by using the ML models and concluded that the Logistic Model Tree (LMT) classifier performs better in analyzing the factors related to maternal health. The IoT-enabled system data were collected and deployed on the LMT model, producing 90% accuracy.

The mortality prediction rate was developed using the ML models, and the two-class SVM model produced a more accurate accuracy of 86.7% compared to other models (Rani and Kumar, 2021). Also, Akbulut et al. (2018) developed the fetal health monitoring system using the Decision Forest Model with an accuracy of 89.5% under test conditions compared to other ML models. Sarhaddi et al. (2021) proposed an IoT-based Maternal health monitoring system for long-term uses that monitor pregnant women the entire time.

Assaduzzaman et al. (2023) focused on ML model to develop risk factors for maternal health using a dataset that preprocessed and applied feature engineering techniques to develop a prediction model using RF and other ML classifiers; among them, RF achieved an accuracy of 90% which was a most top model. Pereira et al. (2020) addressed the health monitoring system of maternal risk factors using six ML models and applied the feature elimination technique RFE to the feature set. The RF classifier with RFE achieved the highest mean accuracy of 93.24%. Pawar et al. (2022) deployed eight ML models using the k-fold cross-validation technique to classify maternal risk into three classes. Among the models, RF provided the best results, with a mean accuracy of 70.21%.

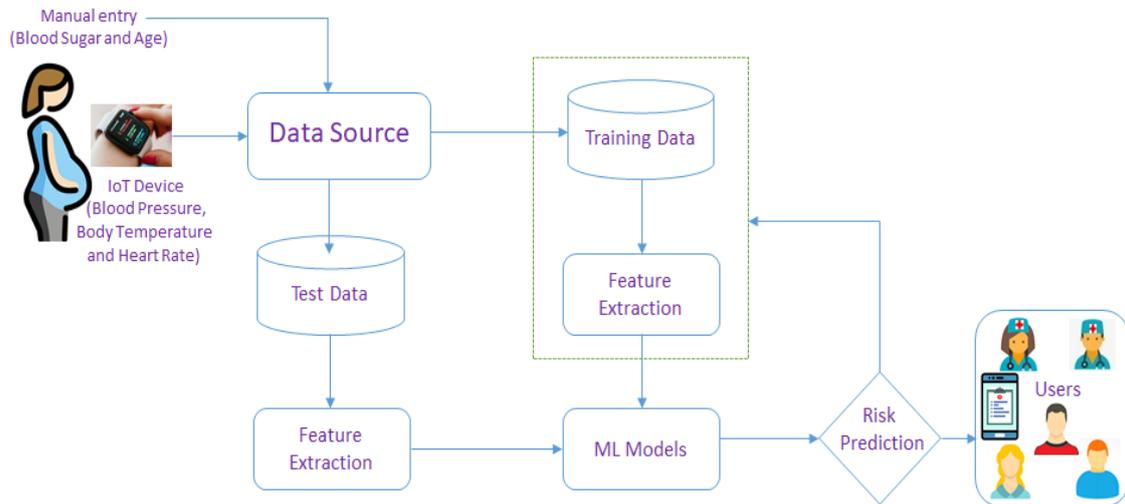
Maternal health risk prediction aims to develop and implement models and systems that can effectively predict the risk associated with maternal health outcomes during pregnancy. It involves research, data collection, model development, result validation, and implementation to improve maternal health care and reduce mortality rates. The concepts used in this study are ML, IoT, and Software Development (Android application).

ML techniques have an important role in maternal health risk prediction. It has been widely used in predicting the mode of childbirth and assessing the potential maternal risk during pregnancy. These techniques allow us to develop prediction models to analyze data and identify patterns, correlations and predictive factors that give rise to adverse maternal health outcomes. Machine

Learning can be utilized through Data Analysis and Feature Selection, Model Development, Training and Validation, and Predictive analysis. The classification task of predicting a specific disease, malware, or conditions using ML techniques enables one to reduce the dimension of the features using feature selection techniques or applying the data analysis approaches and combining the different model's predictions using ensemble techniques (Islam et al., 2023).

The upcoming challenges in the medical field are the development of modern IoT devices and the environment provided by the technology enhancement and the uses of IoT applications. With the recent development of the new Medical 4.0 in the healthcare sector, everything is now connected through IoT nodes, even hospital beds, to patients' physical and biological characteristics. The application of Medical 4.0 in healthcare sectors is discussed by Haleem et al. (2022) and provides the details to decrease the cost of healthcare expenses in underdeveloped or developed countries. Patient data is digitalized, and the transformation of doctor-centric treatment at a hospital or clinic is replaced by IoT technology to patient-centric approaches. Medical 4.0 is embedded with industry 4.0 at the manufacturing level with high safety, security and privacy and is more effective (Oliveira et al., 2021; Al-Jaroodi et al., 2020). The IoT has a significant role in maternal health risk prediction. It can provide real-time monitoring, data collection, and connectivity between devices. In this research study, three types of IoT devices (Heart rate, blood pressure, and body temperature measuring) will be used; these devices will provide real-time data for risk assessment. Many IoT-based software applications are developed to increase the satisfaction level of patients through smooth communication among the hospitals and are always connected through IoT-enabled applications regardless of the physical locations (Pang et al., 2018; Gupta et al., 2020; Celdrán et al., 2018; Jaleel et al., 2020).

From the above analysis, we note that there is a lack of work on automatic health risk prediction and monitoring of a woman during their maternal. Therefore, the proposed work is important because it integrates IoT and ML to automatically diagnose abnormalities of a woman during their maternal smart at early stage.



**Figure 1. IoT-based automated smart maternal health risks monitoring system: based on machine learning**

## Materials and methods

The proposed system is an android-based maternal health risk prediction system in an IoT environment, designed to analyze data from IoT devices and predict the health risk level of a pregnant woman during pregnancy. Its primary objective is to improve maternal health risk outcomes by identifying high-risk cases early on. The system architecture of the proposed model based on ML classifiers is depicted in Fig. 1. The detailed step-by-step explanation of the system workflow is discussed below in phases.

**Data Collection:** The system would gather relevant data about pregnant women, including age, blood pressure (from IoT device), blood sugar, body temperature (from IoT device) and heart rate (from IoT device).

**Data Preprocessing:** The system would preprocess raw data to make it suitable for further analysis and modelling.

**Exploratory Data Analysis (EDA):** EDA is an approach for analyzing and visualizing data to gain insights, understand the underlying patterns, and identify relationships between variables. It helps in understanding the structure of the data, detecting outliers, and assessing variables.

**Feature Selection/Feature Engineering:** It is the process of choosing a subset from a large set of available features in a dataset.

**Machine Learning Models:** The system would then utilize machine learning algorithms to analyze the collected data and identify patterns and correlations between risk factors and potential health risks.

**Risk Level Assessment:** Based on the analysis, the system would assign a risk level to each pregnant woman, indicating the likelihood and severity of potential health risks. This scoring system can help prioritize high-risk cases for further and immediate medical attention.

**Early Warning:** The system can generate alerts and notifications for healthcare professionals and registered family members when a patient's risk level crosses a certain threshold.

**Android Application Deployment:** Deploying an Android app that utilizes machine learning models involves several steps. First, the machine learning model must be trained and optimized for mobile deployment. Then, the model is integrated into the Android app, ensuring compatibility and efficient resource usage. Finally, the app and the embedded machine learning model are packaged and benefit from the intelligent functionalities.

## Implementation details

As per the proposed system architecture, using the first approach after collecting the raw dataset from the open source, we performed some data preprocessing techniques to transform the raw dataset into a processed dataset to perform the ML model deployment for deciding any risk of abnormalities during pregnancy time. The two-stage prediction model based on the ML technique in the IoT environment is illustrated in stage 1 for initial model prediction, and in the second stage, a unique approach, EDA was applied for feature selection for the final model deployment. The detailed architecture is depicted in Figure 2.

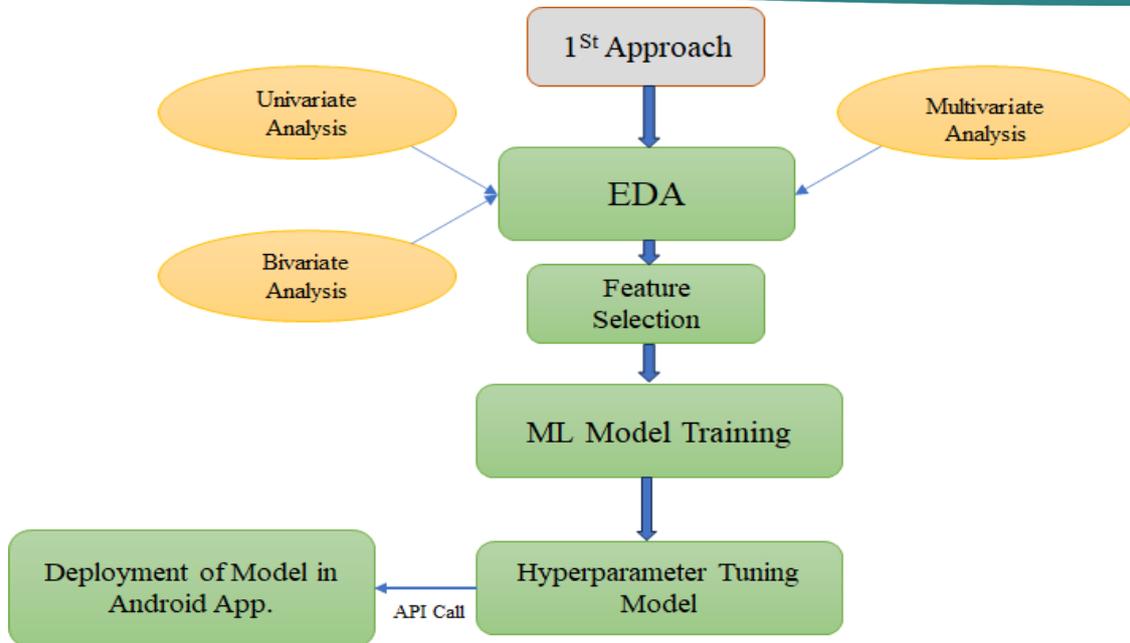


Figure 2. Two-stage prediction model workflow diagram

Table 1. Dataset feature description with null values

#	Column/Feature	#NullValues	Dtype
0	Age	0	int64
1	SystolicBP	0	
2	DiastolicBP	0	
3	BS	0	float64
4	BodyTemp	0	
5	HeartRate	0	int64
6	RiskLevel	0	object

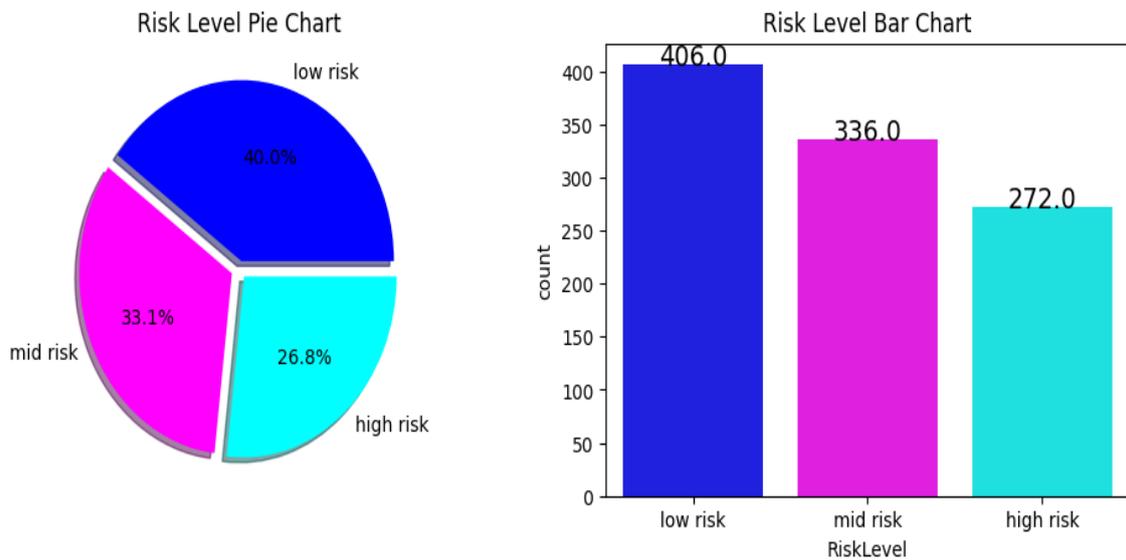


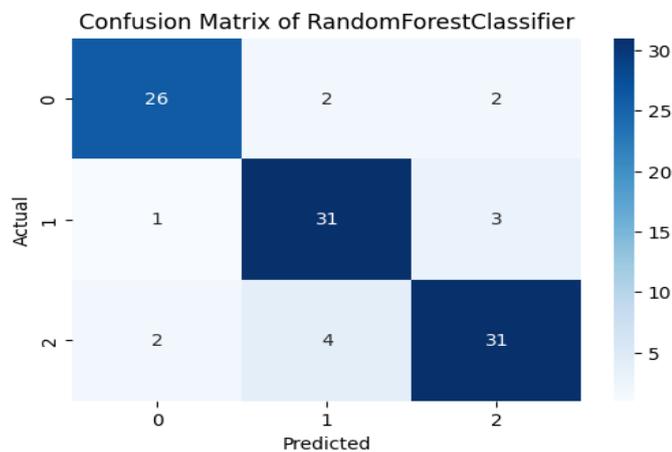
Figure 3. Target outcome data distribution

**Table 2. Experimental results in 1<sup>st</sup> stage model prediction**

Model	Acc %)	Pre	Re	KMA (%)	Fs
RF	86.275	0.864	0.863	83.556	0.863
DT	86.275	0.862	0.863	81.471	0.862
KNN	72.549	0.729	0.725	68.312	0.722
SVM	68.628	0.684	0.686	67.979	0.672
LR	65.686	0.655	0.657	63.700	0.643

### Data preprocessing

Data preprocessing includes cleaning data, removing impossible or replacing null values, and checking categorical features. The entire dataset does not have any null values, and to convert the categorical column, the Label Encoding technique was used to numerical ones for the "RiskLevel" column. To standardize each feature value with a specific range between "0" and "1", the normalization technique was applied to the entire raw dataset using the MinMax scaler to scale down the cell values.

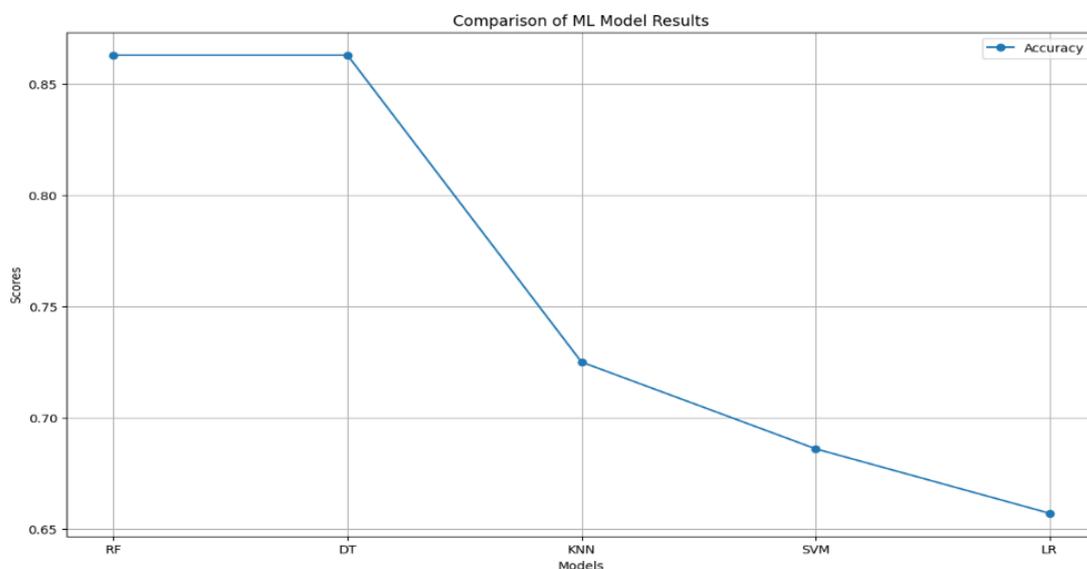
**Figure 4. The CM of the unbalanced RF model**

During the data analysis phase, we checked the data distribution of the target column; the target label was multivalued and categorical in nature, and the class distribution was not equal instances. The pictorial presentation of class distribution is depicted in Figure 3.

Our First Approach to deploying the ML-based model used all the features as independent variables of model input and the target level by considering the actual values of the risk level. For the model creation, we split the dataset of total instances into the ratio of 0.90:0.10 used for the model training and the rest for the model validation.

### Model training & validation

After preprocessing, we moved towards the next stage of model training. The five ML multiclass classifiers, namely Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), k-Nearest Neighbor (KNN), and Logistic Regression (LR), were deployed under the best configurable Python environment using the training dataset. The classification report was derived by considering the performance metrics Accuracy (Acc), Precision (Pre), Recall (Re), and F1 Score (Fs) to evaluate the model performance using the test dataset. The cross-validation (CV) technique was also applied to the entire dataset to handle the low-resource dataset

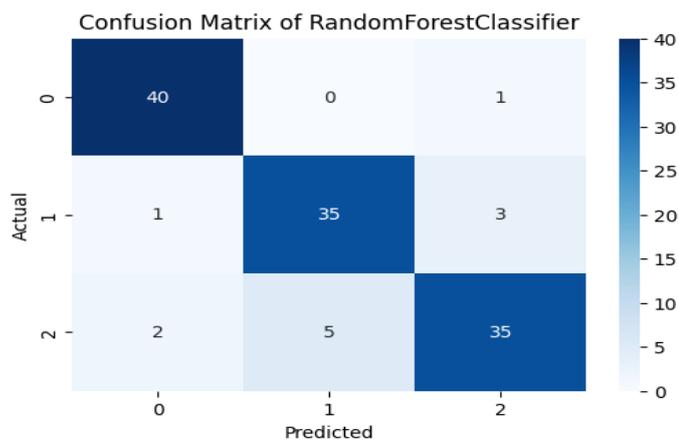
**Figure 5. Accuracy comparison of the deployed ML models**

**Table 3. The balanced employed models’ experimental findings**

Model	Acc (%)	Pre	Re	KMA (%)	Fs
RF	90.164	0.901	0.902	88.140	0.901
DT	88.525	0.885	0.885	86.867	0.883
KNN	77.049	0.769	0.770	73.088	0.768
SVM	67.213	0.663	0.672	68.696	0.655
LR	65.574	0.659	0.656	58.932	0.657

instances situation, in this case, to overcome the model overfitting and underfitting problems. The five-fold CV results and all other metrics outcomes of all the deployed models are represented in Table 2. The Confusion Matrix (CM) of the best-performing RF model is depicted in Figure 4. The experimental results of the deployed models in terms of Acc are depicted in Figure 5.

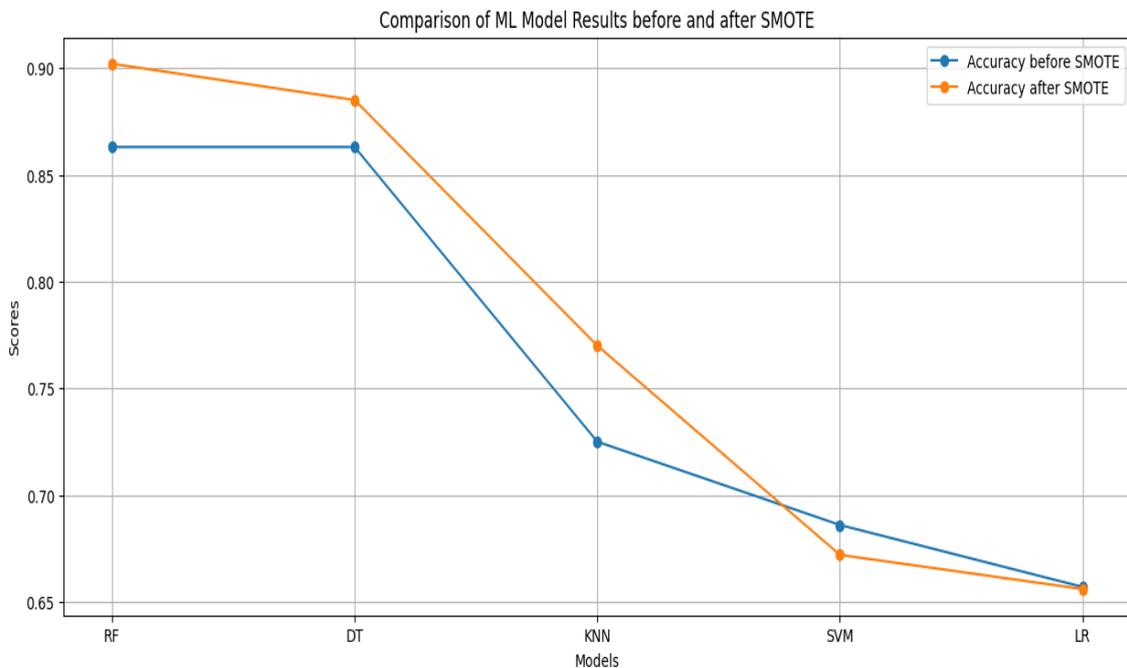
Technique, an algorithm used to address the class imbalance in supervised learning problems. It is designed to oversample the minority class by creating synthetic examples. Both under-sampling and over-sampling have their disadvantages: data loss for under-sampling and overfitting for oversampling. SMOTE has no disadvantages since it creates synthetic examples to balance the data. The results could have been more accurate, but in the case of multiclass, it was impressive. The model's outcome is tabulated in Table 3.



**Exploratory Data Analysis (EDA)**

In the second stage of model training and testing, before that, we performed the exploratory data analysis phases among the features. The three types of EDA approaches were executed by taking the features as a factor and applying Univariate, Bivariate, and Multivariate analysis on the six features corresponding to the target variable “RiskLevel.”

**Figure 6. The CM of the balanced RF model**

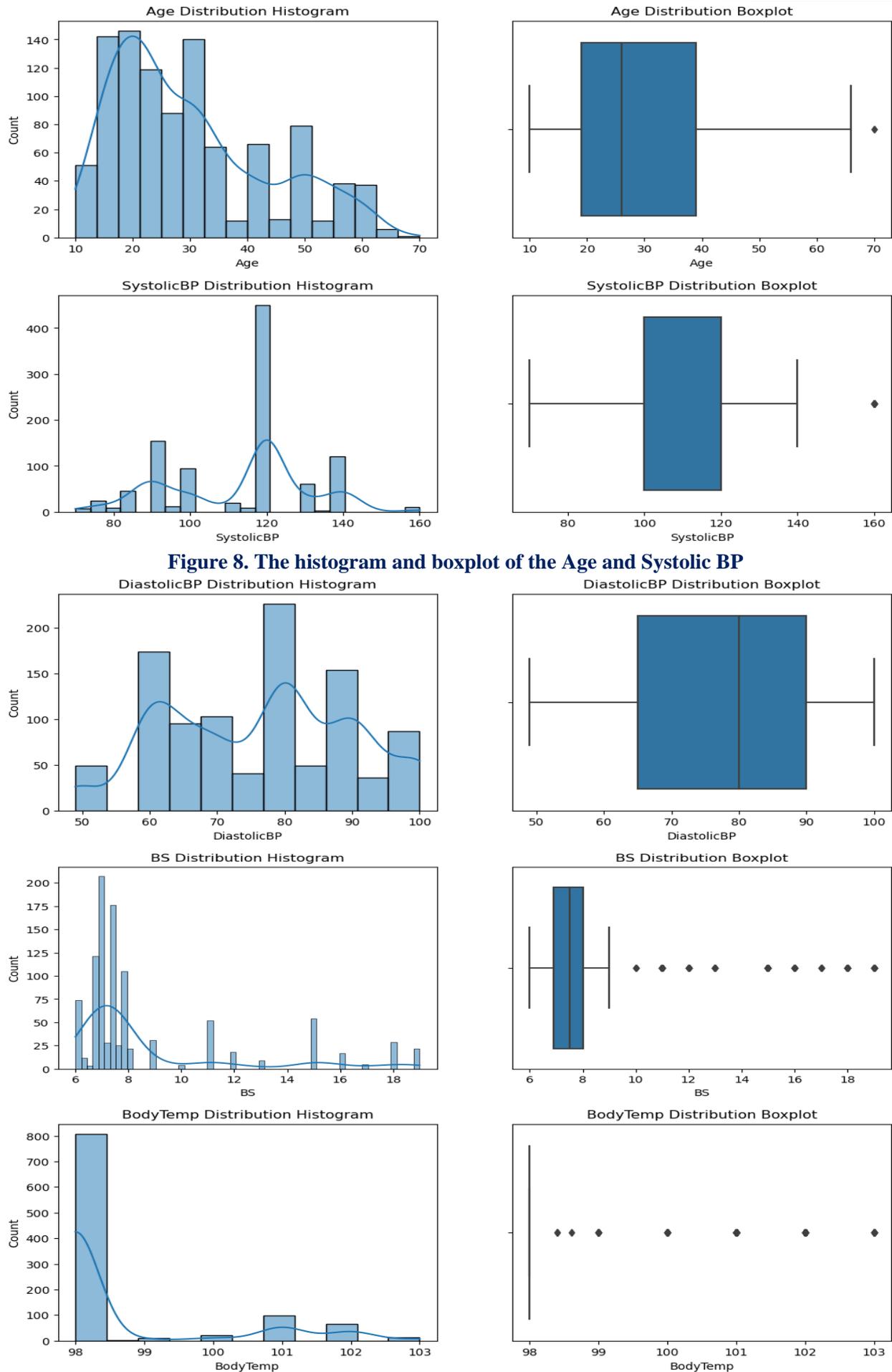


**Figure 7. In 1<sup>st</sup> stage, accuracy comparison of all models**

**Univariate analysis**

The results obtained after implementation were normal, so we used SMOTE, Synthetic Minority Over-Sampling

Univariate analysis separately explores the distribution of each variable in a data set. It looks at the



**Figure 8. The histogram and boxplot of the Age and Systolic BP**

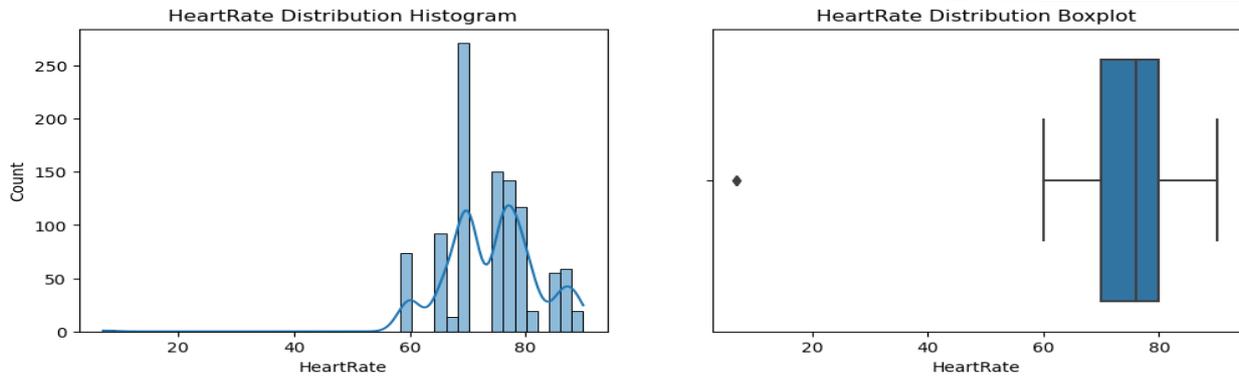


Figure 9. The histogram and boxplot of the Diastolic BP, BS, Body Temp and Heart Rate

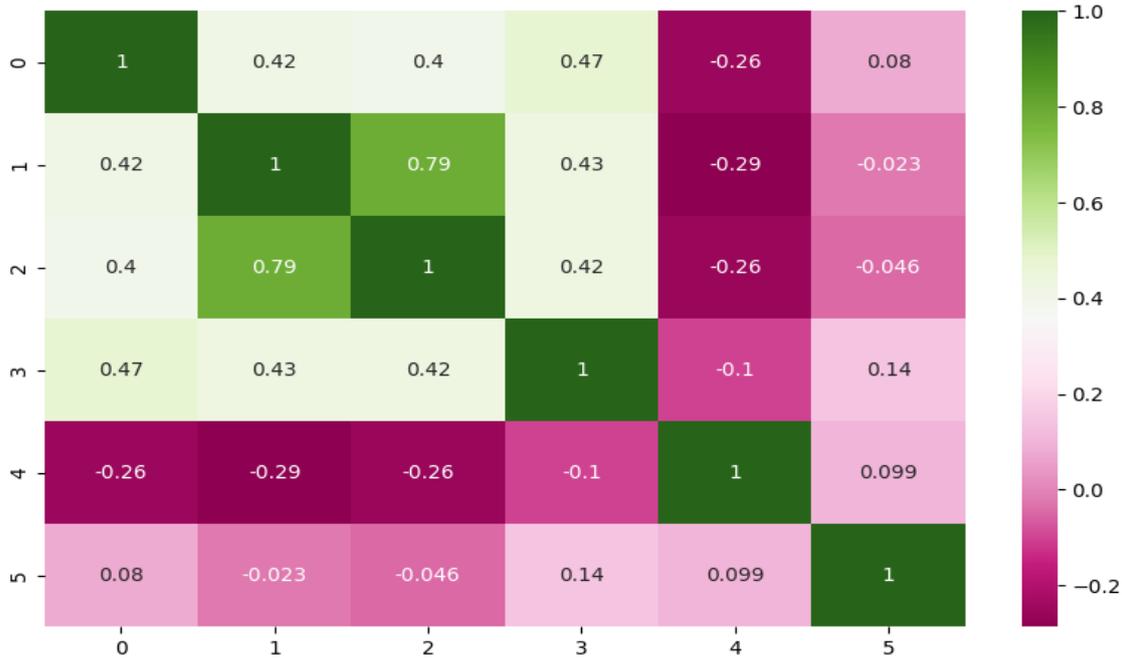


Figure 10. The Correlation Heatmap of all features

range of values and the central tendency of the values. Univariate data analysis does not look at relationships between variables (like bivariate and multivariate analysis); rather, it summarizes each variable independently. Methods to perform univariate analysis will depend on whether the variable is categorical or numerical. For the numerical variable, we would explore the shape of the distribution (distribution can either be symmetric or skewed) using histogram and density plots. We would use bar plots to visualize categorical variables' absolute and proportional frequency distribution.

The different univariate analyses were performed using the histograms and the boxplots of all the features depicted in Figures 8 and 9.

**Observation:** Almost all variables have outliers that cause skewed distribution. We will ignore that outlier for now because that value seems natural in this case, except for "Heart Rate." That variable has an outlier that is too far from the other values.

**Bivariate analysis**

Bivariate analysis helps study the relationship between two variables. It helps to find out if there is an association between the variables, and if yes, then what is the strength of the association? One variable here is dependent, while the other is independent. We used correlation coefficients to find out how high is the relationship between two variables. We also use scattered plots to show the patterns that can be formed using the two variables. The correlation among the features and with the target column, the heatmap was derived to check the inertia values among the features are depicted in Figure 10.

**Observation:** "Systolic BP" and "Diastolic BP" are highly correlated. As we can see from the graph, they have a positive correlation with a correlation coefficient value of 0.79. This means that SystolicBP and DiastolicBP variable contains highly similar information, with very little or no variance in information. This is known as a problem called multicollinearity, which undermines the statistical significance of an independent

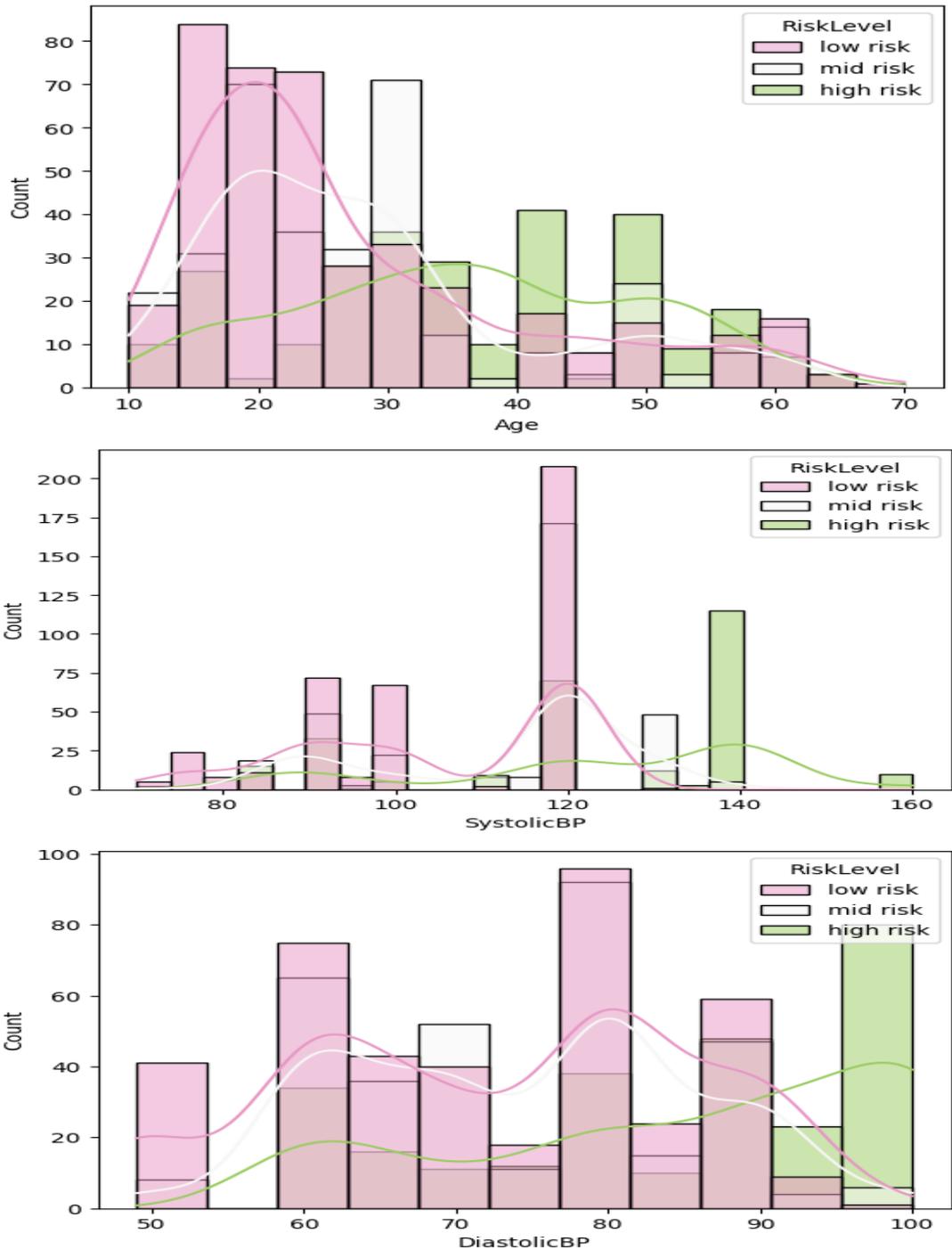
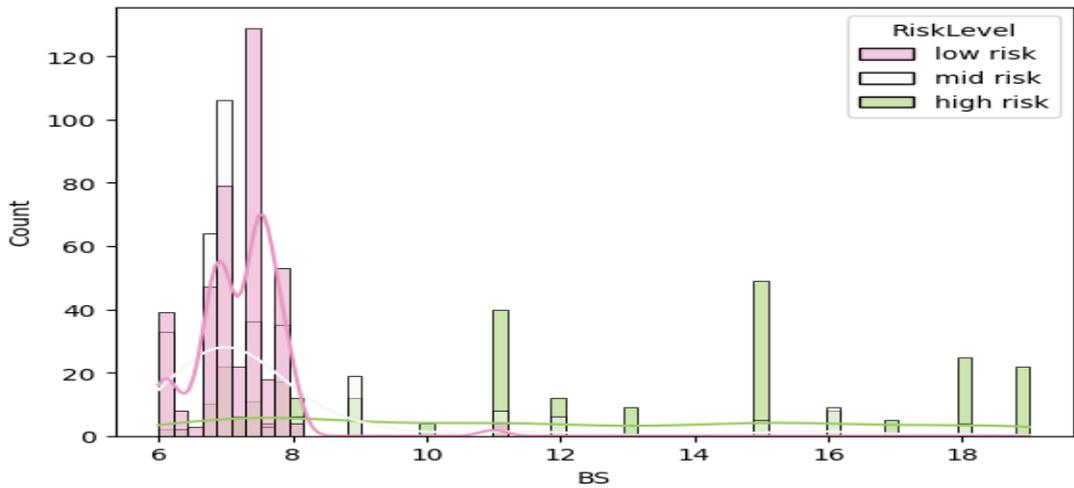
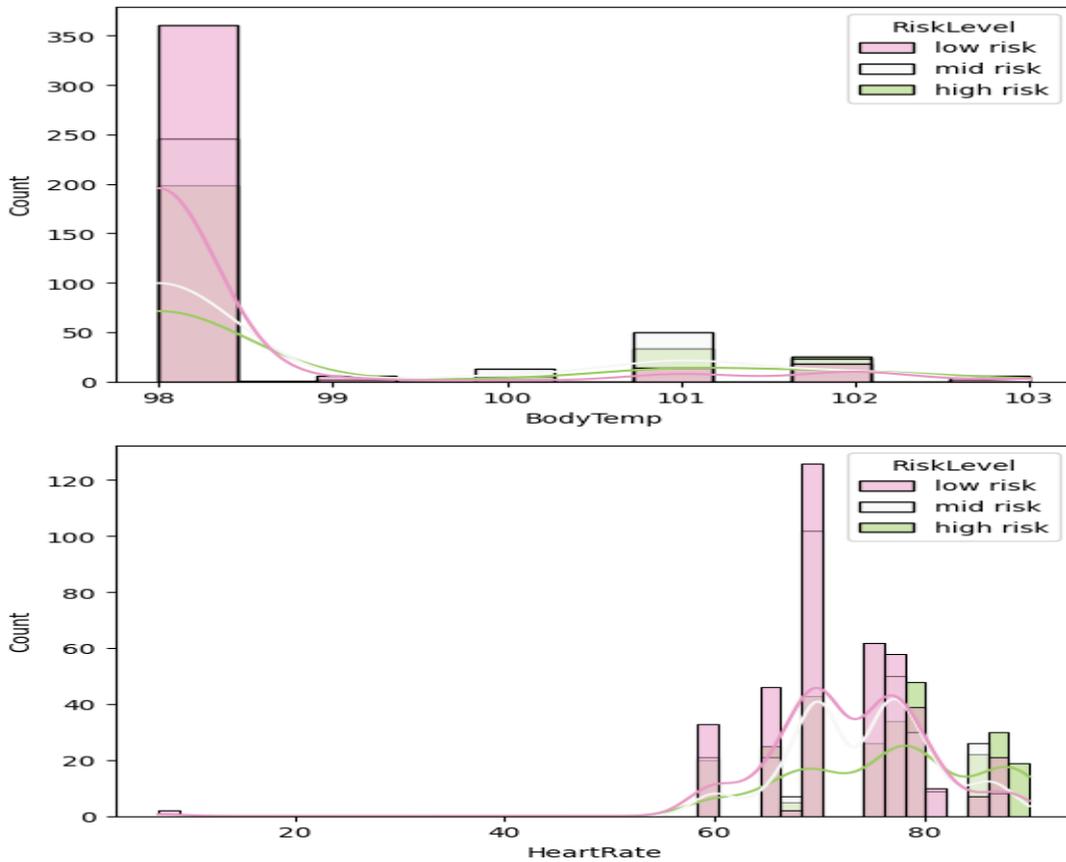


Figure 11. Bivariate histogram diagram of features concerning target outcome





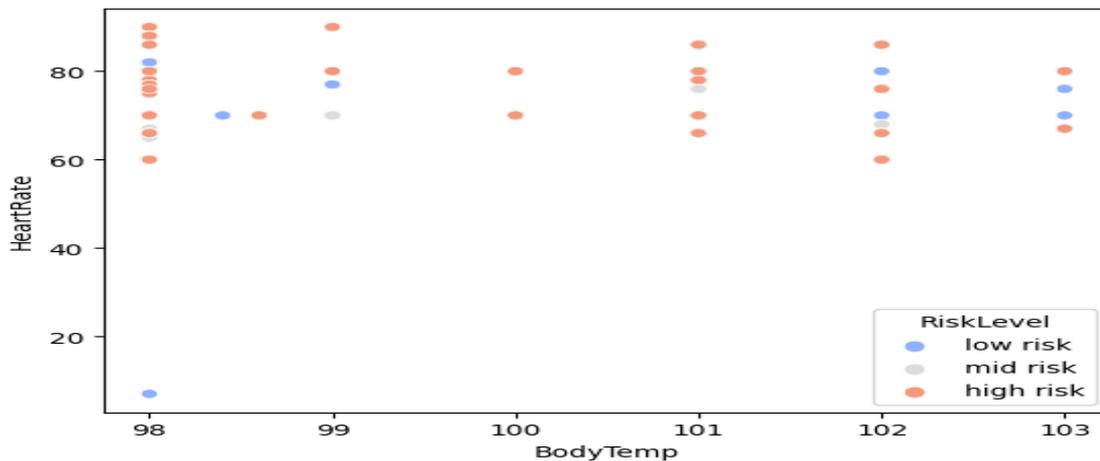
**Figure 12. Bivariate histogram diagram of features concerning target outcome**

Health risks seem to be getting higher along with the number of heart rates.

**Multivariate analysis**

Multivariate analysis involves analyzing multiple variables (more than two) to identify any possible association and find the relationship among them. More

variable. We can remove one of them because we do not want a **redundant** variable while making or training our model. However, we will dig deeper to decide whether we need to remove this variable and which variable we



**Figure 13. Multivariate histogram of Body Temp and Heart Rate concerning Risk Level**

should remove.

We used the histogram with hue mapping to visualize the predictor variables' data distribution based on the target variable and patronized in Figures 11 and 12 sequentially.

**Observation:** As mentioned before, the "Heart Rate" variable has an outlier with an unnatural value of 6 bpm.

specifically, we tried associating more than one predictor variable with the response variable.

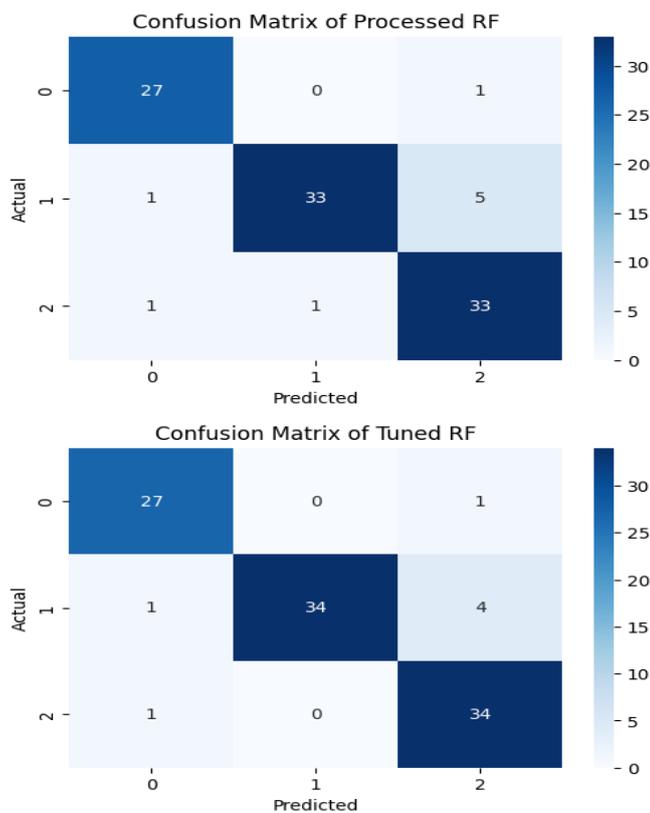
In this case, we analyzed the impact of two different predictor variables simultaneously on the "RiskLevel" variable. We used a scatter plot since all the predictor variables have numerical values and then grouped them using Risk Level values with different colours. We

analyzed the risk level by considering two variables at a time. We observed that in the previous two stages, “Heart Rate” and “Body Temperature” were highly correlated with the response. In this case, only one scatter plot is provided for the conclusion in Figure 13.

**Observation:** Pregnant women with higher body temperature seem to have a higher health risk, regardless of their heart rate; also noted, according to the previous analysis, pregnant women in this observation mostly have a 98 F body temperature. The HeartRate variable could be more helpful in this case.

**Table 4. Proposed prediction model experimental results**

Model	Acc (%)	Pre	Re	KMA (%)	Fs
Processed-RF	91.176	0.917	0.911	90.897	0.912
Tuned-RF	93.137	0.937	0.932	93.111	0.932



**Figure 14. The CM of the processed and tuned RF model**

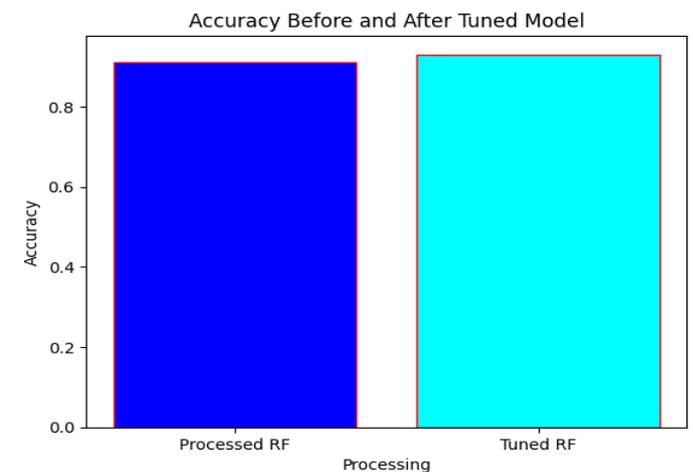
**Discussion**

In this dataset, several variables have outliers, but most of those values still make sense in real life. The only variable that has an outlier with an unreasonable value is "Heart Rate." In this variable, two observations have a heart rate value of 7 bpm (beats per minute). The average resting heart rate for adults ranges from 60 to 100 beats per minute, and the lowest recorded resting heart rate in human history was 25 bpm. Therefore, we will drop these two records with a heart rate value of

“Seven” because that value does not make sense and is most likely an input error.

We will not store processed data in the original variable; instead, we will store it in the new variable to compare it with the original data. Then, after conducting several analyses of the predictor variables, we conclude that the "Heart Rate" variable is less helpful in determining the health risks of pregnant women. So, it is safe to remove that variable. If we delete that variable, one might wonder why we drop records with outliers on the HeartRate variable. The answer is that it has an input

error, so the records may need to be legit. The label is also incorrect, misleading the training process and making the model less accurate.

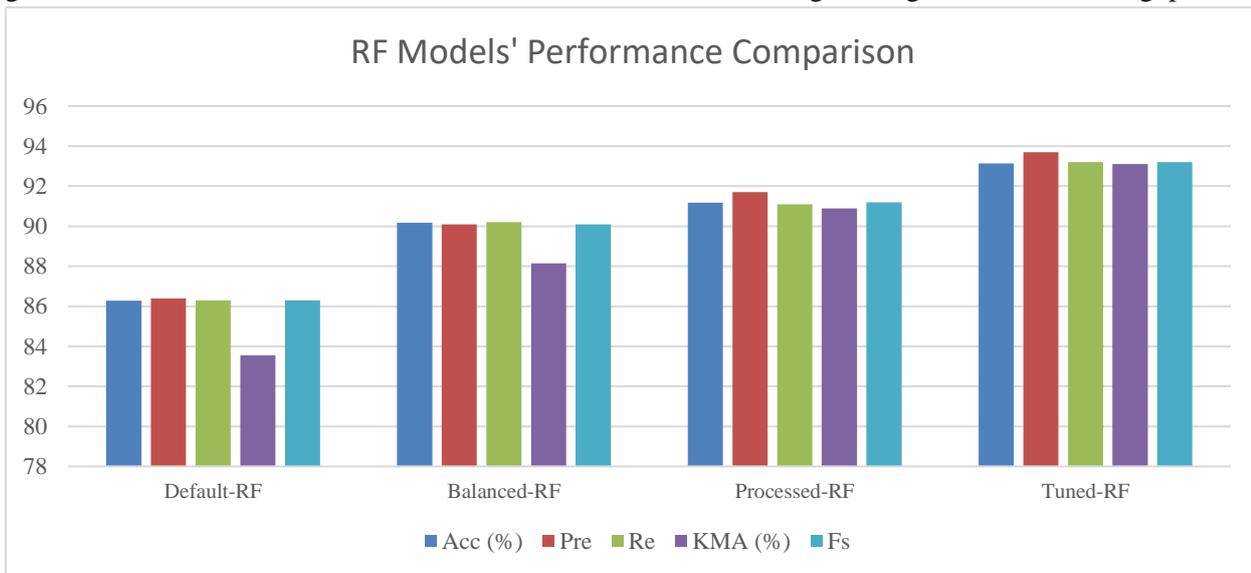


**Figure 15. The accuracy comparison of the processed and tuned RF model**

This research study concludes with an analysis of the acquired dataset after performing EDA technique; we can wind up that BS level is the most important variable in determining the health level of pregnant women. Pregnant women with high blood glucose levels tend to have high health risks. Over 75% of pregnant women with a BS of 8 or more have a high health risk. BS also has a relatively strong positive correlation to Age, Systolic BP, and Diastolic BP, so pregnant women with high Age, Systolic BP, and Diastolic BP must be vigilant. Age is also an important variable, where the health risks of pregnant women seem to start to increase starting from the age of 25 years. For Systolic BP and Diastolic BP, these two variables have a strong relationship, as evidenced by the correlation coefficient value of 0.79. About Body Temp, this variable does not give much information because more than 79% of the total value is 98°F. However, this variable shows that pregnant women

with a body temperature above 98.4°F tend to have a greater health risk. The last one is Heart Rate, the least relevant variable in determining the health level of pregnant women.

initial dataset. This resulted in an accuracy of 90.16% throughout the testing phase, within the optimal customizable setting. During the subsequent phase, feature engineering and data cleaning procedures were



**Figure 16. The performance comparison of the different RF models**

### Experimental results

Based on the Second Approach, the results obtained after doing EDA and then training and testing the best-performing ML model, RF Classifier and then again fine-tuning the model using Grid Search CV along k-fold CV are shown below.

After applying for EDA and eliminating the feature “Heart Rate,” the prediction model is trained using the 90% instances, and the model is validated over the 10% data instances. The accuracy is observed significantly under the same test condition. We fine-tuned the RF model using the grid hyperparameter values and performed the Grid Search CV for better prediction outcomes. The processed data and hyper-tuned RF model results are summarized in Table 4. The CM of the processed and Tuned RF model and their accuracy comparison are depicted in Figures 14 to 15, respectively. Finally, the performance improvement of the RF prediction model is significantly noticeable and represented in Figure 16.

### Conclusion & future scope

This work culminates by constructing a stage prediction model. In the initial phase of constructing the classification model, five machine-learning classifiers were employed. Among these classifiers, the Random Forest (RF) classifier demonstrated an accuracy of 86.28% when applied to the obtained dataset. Subsequently, we implemented the balanced Synthetic Minority Over-sampling Technique (SMOTE) on the

executed, involving the removal of data outliers and the deletion of extraneous variables. As a result, the accuracy of the model exhibited an improvement, reaching a value of 91.18%. The results indicate that the suggested model exhibits superior generalization capabilities when applied to the processed dataset. In addition, hyperparameter tuning was conducted to determine the optimal values for the hyperparameter estimator in the Random Forest method. By utilizing the optimal hyperparameter determined by the Grid Search CV tuning technique, the model achieves an enhanced accuracy rate of 93.14%. The use of cross-validation, employing a five-fold data-splitting methodology throughout the entirety of the dataset, resulted in a noteworthy mean accuracy of 93.11%. This outcome suggests the presence of a stable prediction model that is not prone to overfitting.

This research study could be the scope of the real-time alerts and interventions system that can be enhanced to provide real-time alerts and interventions based on risk prediction to enable timely notifications to healthcare professionals, allowing personalized care. The Android app's user experience and interface can be improved to ensure its effectiveness and widespread adoption. A feedback mechanism can be created by getting input from healthcare professionals and pregnant women and can be incorporated to enhance the usability and accessibility of the system. Since the study involves collecting sensitive health data and ensuring robust data privacy and security measures, which is of utmost importance, a strong encryption technique can be developed, and compliance

with privacy regulations should be ensured to protect the confidentiality and integrity of the collected data.

### Conflicts of interest

There are no known conflicts of interest for the authors in the publication of this work.

### References

- Ahmed, M. (2022). Maternal Health Risk Data. From Kaggle: <https://www.kaggle.com/datasets/csafrif2/maternal-health-risk-data>
- Ahmed, M., Abul Kashem, M., Rahman, M., & Khatun, S. (2020). Review and Analysis of Risk Factor of Maternal Health in Remote Area Using the Internet of Things (IoT). *In ECCE2019*, pp. 357–365. [https://doi.org/10.1007/978-981-15-2317-5\\_30](https://doi.org/10.1007/978-981-15-2317-5_30)
- Assaduzzaman, M., Al Mamun, A., & Hasan, M. (2023). Early Prediction of Maternal Health Risk Factors Using Machine Learning Techniques. *2023 International Conference for Advancement in Technology (ICONAT)*. Goa, India. <https://doi.org/10.1109/ICONAT57137.2023.10080700>
- Celdrán, A., Gil Pérez, M., García Clemente, F., & Martínez Pérez, G. (2018). Sustainable securing of Medical Cyber-Physical Systems for the healthcare of the future. *Sustainable Computing: Informatics and Systems*, *19*, 138-146. <https://doi.org/10.1016/j.suscom.2018.02.010>
- Pereira, S., Costa Filho, R., Ramos, R., & Oliveira, M. (2020). Improving Maternal Risk Analysis in Public Health Systems. *5th International Conference on Smart and Sustainable Technologies (SpliTech)*. Split, Croatia. <https://doi.org/10.23919/SpliTech49282.2020.9243769>
- Raza, A., Rehman Siddiqui, H., Munir, K., & Almutairi, M. (2022). Ensemble learning-based feature engineering to analyze maternal health during pregnancy and health risk prediction. *PLoS ONE*, *17*(11), e0276525. <https://doi.org/10.1371/journal.pone.0276525>
- Akbulut, A., Ertugrul, E., & Topcu, V. (2018). Fetal health status prediction based on maternal clinical history using machine learning techniques. *Computer Methods and Programs in Biomedicine*, *163*, 87-100. <https://doi.org/10.1016/j.cmpb.2018.06.010>
- Al-Jaroodi, J., Mohamed, N., & Abukhousa, E. (2020). Health 4.0: On the Way to Realizing the Healthcare of the Future. *IEEE Access*, *8*, 211189 - 211210. <https://doi.org/10.1109/ACCESS.2020.3038858>
- Castillejo, P., Martinez, J. F., Rodriguez-Molina, J., & Cuerva, A. (2013). Integration of wearable devices in a wireless sensor network for an E-health application. *IEEE Wireless Communications*, *20*(4), 38 - 49. <https://doi.org/10.1109/MWC.2013.6590049>
- Chen, H.-Y., Chuang, C. H., Yang, Y. J., & Wu, T. P. (2011). Exploring the risk factors of preterm birth using data mining. *Expert Systems with Applications*, *38*(5), 5384-5387. <https://doi.org/10.1016/j.eswa.2010.10.017>
- Gupta, R., Shukla, A., Mehta, P., & Bhattacharya, P. (2020). VAHAK: A Blockchain-based Outdoor Delivery Scheme using UAV for Healthcare 4.0 Services. *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. Toronto, ON, Canada. <https://doi.org/10.1109/INFOCOMWKSHPS50562.2020.9162738>
- Haleem, A., Javaid, M., Singh, R., & Suman, R. (2022). Medical 4.0 technologies for healthcare: Features, capabilities, and applications. *Internet of Things and Cyber-Physical Systems*, *2*, 12-30. <https://doi.org/10.1016/j.iotcps.2022.04.001>
- Hussain, T. M., Shaikh, M., Ali, B. R., & Talpur, H. (2014). Internet of Things as Intimating for Pregnant Women's Healthcare: An Impending Privacy Issues. *The Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, *12*(6), 4337-4344. <https://doi.org/10.11591/telkomnika.v12i6.4227>
- Islam, R., Sayed, M., Saha, S., & Jamal Hossain, M. (2023). Android malware classification using optimum feature selection and ensemble machine learning. *Internet of Things and Cyber-Physical Systems*, *3*, 100-111. <https://doi.org/10.1016/j.iotcps.2023.03.001>
- Jaleel, A., Mahmood, T., Awais Hassan, M., & Bano, G. (2020). Towards Medical Data Interoperability Through Collaboration of Healthcare Devices. *IEEE Access*, *8*, 132302 - 132319. <https://doi.org/10.1109/ACCESS.2020.3009783>
- Oliveira, M.Y., Pesqueira, A., Sousa, M., & Dal Mas, F. (2021). The Potential of Big Data Research in

- HealthCare for Medical Doctors' Learning. *Journal of Medical Systems*, 45(13). <https://doi.org/10.1007/s10916-020-01691-7>
- Pang, Z., Yang, G., Khedri, R., & Zhang, Y.T. (2018). Introduction to the Special Section: Convergence of Automation Technology, Biomedical Engineering, and Health Informatics Toward the Healthcare 4.0. *IEEE Reviews in Biomedical Engineering*, 11, 249 - 259. <https://doi.org/10.1109/RBME.2018.2848518>
- Pawar, L., Malhotra, J., Sharma, A., & Arora, D. (2022). A Robust Machine Learning Predictive Model for Maternal Health Risk. *3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)*. Coimbatore, India. <https://doi.org/10.1109/ICESC54411.2022.9885515>
- Pereira, S., Portela, F., & Santos, M. (2015). Predicting Type of Delivery by Identification of Obstetric Risk Factors through Data Mining☆. *Procedia Computer Science*, 64, 601-609. doi:10.1016/j.procs.2015.08.573
- Rani, S., & Kumar, M. (2021). Prediction of the mortality rate and framework for remote monitoring of pregnant women based on IoT. *Multimedia Tools and Applications*, 80, 24555–24571. <https://doi.org/10.1007/s11042-021-10823-1>
- Rawashdeh, H., Awawdeh, S., Shannag, F., & Henawi, E. (2020). Intelligent system based on data mining techniques for prediction of preterm birth for women with cervical cerclage. *Computational Biology and Chemistry*, 85, 107233. <https://doi.org/10.1016/j.compbiolchem.2020.107233>
- Redondi, A., Chirico, M., Borsani, L., & Cesana, M. (2013). An integrated system based on wireless sensor networks for patient monitoring, localization and tracking. *Ad Hoc Networks*, 11(1), 39-53. <https://doi.org/10.1016/j.adhoc.2012.04.006>
- Sarhaddi, F., Azimi, I., Labbaf, S., & Niela-Vilén, H. (2021). Long-Term IoT-Based Maternal Monitoring: System Design and Evaluation. *Sensors*, 21(17), 2281. <https://doi.org/10.3390/s21072281>
- WHO.(2023). Maternal mortality. Retrieved June 01, 2023 from <https://www.who.int/news-room/fact-sheets/detail/maternal-mortality>

### How to cite this Article:

Subhash Mondal, Amitava Nag, Anup Kumar Barman and Mithun Karmakar (2023). Machine Learning-based maternal health risk prediction model for IoMT framework. *International Journal of Experimental Research and Review*, 32,145-159.

DOI : <https://doi.org/10.52756/ijerr.2023.v32.012>



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.