Journal of Cardiovascular Disease Research

ISSN: 0975-3583,0976-2833 VOL 11, ISSUE 8, 2020

A Data-Driven Framework for Early Maternal Risk Detection Using IoT-Enabled ML Models

Komati Sathish, Associate Professor, Department of Computer Science and Engineering, Annamacharya Institute of Technology and Sciences, Batasingaram, Hayath Nagar, Ranga Reddy, Telangana, India.

Email: komatisathish459@gmail.com

ABSTRACT

Ensuring maternal well-being is a fundamental pillar of public health, requiring proactive strategies to identify clinical risks early in pregnancy. This research proposes a sophisticated framework for Maternal Health Risk Prediction that merges Internet of Things (IoT) capabilities with advanced Machine Learning (ML) algorithms. By utilizing IoT-enabled sensors, the system facilitates the continuous, real-time tracking of critical physiological markers—including blood glucose, body temperature, heart rate, and blood pressure—generating a comprehensive dataset for clinical analysis. While traditional models like the K-Nearest Neighbors (KNN) algorithm provide a foundational starting point, they often struggle with computational overhead and feature scaling sensitivities in complex datasets. To address these limitations, this study implements the Extra Trees Classifier (ETC), which leverages ensemble learning and stochastic feature selection to offer superior scalability and precision. Empirical results, validated through metrics such as F1-score and accuracy, demonstrate that the Extra Trees approach significantly outperforms baseline models. This research concludes that integrating randomized ensemble methods with IoT telemetry provides a more reliable and efficient pathway for timely medical interventions, ultimately improving neonatal and maternal survival rates.

Keywords: Clinical risk identification, Maternal health risk Prediction, IoT-enabled sensors, Heart rate, K-nearest neighbours, Extra Trees classifier.

1. 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 According to WHO's report in 2020, around 800 women die daily due to poor resources and care .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.

2. LITERATURE SURVEY

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. [1] 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

ISSN: 0975-3583,0976-2833 VOL 11, ISSUE 8, 2020

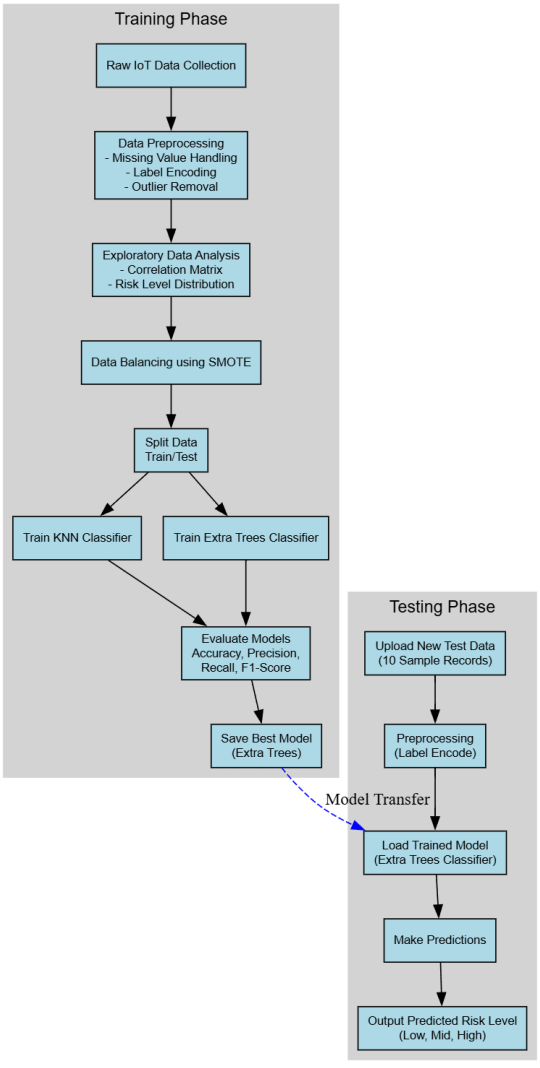
observed results after balancing using SMOTE with an average accuracy of 88%. Also, they applied feature selection techniques and used SVM along with BiLTCN, claiming 98% accuracy on the reduced feature model. Ahmed et al. [2] executed research by using the ML models and concluded that the Logistic Model 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, [3]. Also, Akbulut et al. [4] 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. [5] proposed an IoT-based Maternal health monitoring system for long-term uses that monitor pregnant women the entire time. Assaduzzaman et al. [6] 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. [7] 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. [8] 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 .[9] . 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. [10] 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. [11] Al-Jaroodi et al., [12]. The IoT has a significant role in maternal health risk prediction. It can provide real-time monitoring, data collection, and connectivity between devices. In the 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

ISSN: 0975-3583,0976-2833 VOL 11, ISSUE 8, 2020

locations (Pang et al.; Gupta et al., [13]; Celdrán et al.,[14] ; Jaleel et al.,[15] . 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.

3. PROPOSED SYSTEM

The research focuses on predicting maternal health risk levels using IoT-generated physiological data and machine learning models to enable early medical intervention. The data collected includes vital parameters such as age, blood pressure, blood sugar, temperature, and heart rate, gathered via IoT-based health monitoring systems. After preprocessing the data by handling missing values, encoding labels, and performing scaling, an exploratory data analysis (EDA) was conducted to understand feature distributions and correlations. Initially, the K-Nearest Neighbors (KNN) classifier was used as the existing model, offering a baseline for performance comparison. However, due to its limitations in handling large-scale data and sensitivity to feature scaling, the ETC Model was proposed. This ensemble-based model demonstrated better accuracy, robustness, and computational efficiency. The comparative results showed that the ETC Model is more suitable for accurately predicting maternal health risks, highlighting the potential of combining IoT data with advanced machine learning techniques in healthcare applications.



ISSN: 0975-3583,0976-2833 VOL 11, ISSUE 8, 2020

Fig. 1: Proposed Block Diagram.

The system begins with real-time health data collection through IoT monitoring devices, capturing parameters such as age, systolic and diastolic blood pressure, blood sugar levels, body temperature, and heart rate, which are stored in a structured CSV file with predefined risk labels for supervised learning while ensuring patient privacy and continuous data streaming. The collected data undergoes preprocessing, including handling missing values through removal or imputation, converting categorical risk labels into numerical form using label encoding, and applying feature scaling such as StandardScaler—especially important for distance-based algorithms like KNN—along with consistency checks to maintain data integrity. During exploratory data analysis, count plots are used to examine class distribution and identify imbalance, guiding techniques like stratified sampling or SMOTE if necessary. A KNN classifier is first implemented as a baseline model due to its simplicity and interpretability, classifying health risk levels based on the majority vote of nearest neighbors, though evaluation metrics such as accuracy, precision, recall, and F1-score reveal certain performance limitations. To overcome these issues, an ETC Model is proposed, which combines multiple highly randomized decision trees to reduce overfitting, uses random feature splits to speed up training, and handles high-dimensional data more effectively. Comparative evaluation shows that ETC Model generally outperforms KNN in accuracy, achieves better precision–recall balance—particularly for minority classes—and offers faster training and improved scalability. Overall, the ETC model proves more suitable for health risk prediction, with future enhancements including edge computing for real-time predictions and centralized dashboards to support doctors in clinical decision-making.

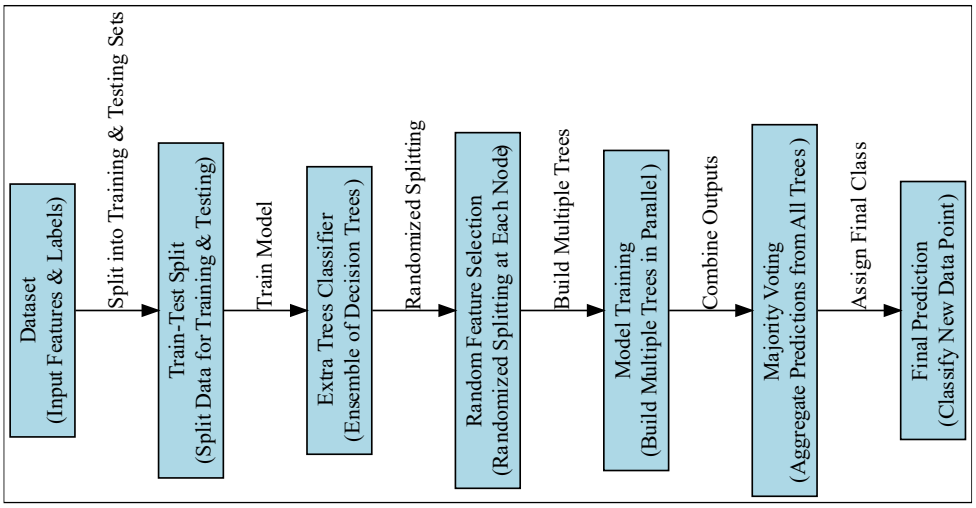


Fig. 2: ETC Model Block Diagram.

3.1 ETC

The proposed ETC model is an ensemble-based supervised learning algorithm that builds multiple decision trees using random subsets of features and data samples. Unlike traditional decision trees or even Random Forests, ETC Model introduces more randomness by selecting cut-points randomly for feature splits, enhancing generalization and reducing overfitting. In the context of maternal health risk prediction, this classifier effectively identifies patterns in physiological IoT data to categorize the maternal condition into Low Risk, Mid Risk, or High Risk. The process begins by preparing the data collected from IoT-enabled maternal health monitoring systems, where X_train consists of structured numerical features such as age, systolic and diastolic blood pressure, blood sugar level, body temperature, and heart rate, and y_train contains the corresponding maternal risk categories (Low, Mid,

High) that serve as ground truth labels. The ETC Model is then trained by constructing an ensemble of multiple de-correlated decision trees, where each tree is built using either bootstrapped samples or the full dataset, and at every split, a random subset of features and random threshold values are selected to increase diversity and reduce overfitting; the final training prediction is determined through majority voting across all trees. After training, the model is evaluated using X_test, which includes new unseen patient data collected during real-time monitoring or controlled testing, with each instance passed through all trees and the majority class selected as the final predicted risk level. The predicted outputs are then compared with y_test, the true maternal risk labels, to assess performance using metrics such as accuracy to measure overall correctness, precision and recall to evaluate class-specific detection capability, F1-score to balance precision and recall, confusion matrix to visualize classification errors across classes, and ROC-AUC (using one-vs-rest for multiclass settings) to analyze the model's overall discriminative ability.

4.RESULTS AND DESCRIPTION

The research starts by importing essential libraries including NumPy and Pandas for data handling, Seaborn and Matplotlib for visualization, Scikit-learn for machine learning algorithms and evaluation, Joblib and Pickle for model persistence, and Imbalanced-learn (SMOTE) to address class imbalance. The maternal health dataset, collected via IoT monitoring, is loaded into a DataFrame using Pandas. Data preprocessing involves checking and forward-filling missing values, removing duplicates, and encoding the target variable 'RiskLevel' with LabelEncoder. Exploratory data analysis is conducted with count plots to visualize class distributions and heatmaps to examine feature correlations. The dataset is split into training and testing sets with stratification to preserve class balance, followed by applying SMOTE on the training data to generate synthetic samples for minority classes, preventing model bias. A PerformanceMetrics function is defined to calculate accuracy, precision, recall, F1-score, classification reports, confusion matrices, and ROC-AUC for robust evaluation. Two models are built and trained: KNN, which loads a saved model if available or trains a new one with 5 neighbors, and ETC Model, which constructs multiple randomized decision trees to improve performance and reduce overfitting; both models are saved for future reuse. After training, new test data (a subset of the original dataset for demonstration) is prepared by removing target labels, and the ETC model is used to predict risk levels, mapping numeric predictions back to descriptive labels and appending them as a new column in the test data. Joblib is employed throughout to efficiently save and load models, enabling fast deployment without retraining. The final output includes trained models for both classifiers, comprehensive performance evaluations, and predicted risk levels on new patient data, supporting informed decision-making in maternal health risk assessment.

4.1 Result Analysis

The fig 3 shows dataset that suggests a focus on predicting health risks based on the above factors. Individuals with higher systolic and diastolic blood pressures, elevated blood sugar levels, and abnormal heart rates or body temperatures tend to be classified as "high risk." Conversely, lower values seem to correlate with a "low risk" classification. The fig 4 shows heatmap that visually represents the correlation between different features in the dataset, including Age, Systolic BP, Diastolic BP, Blood Sugar (BS), Body Temperature, Heart Rate, and Risk Level. The correlation values range from -1 to 1, with darker red indicating stronger positive correlations and darker blue indicating stronger negative correlations. The heatmap helps identify the strength and direction of relationships between variables, which can be useful for predictive modeling and feature selection. The fig 5 represents a count plot showing the distribution of different "RiskLevel" categories. The x-axis corresponds to the three distinct

ISSN: 0975-3583,0976-2833 VOL 11, ISSUE 8, 2020

risk levels (0, 1, and 2), while the y-axis indicates the count of occurrences for each risk level. From the chart, it's clear that there is a relatively balanced distribution of observations across the three risk levels, as the heights of the bars for risk levels 0, 1, and 2 are similar, with each bar having around 175 entries. This suggests that the dataset is not heavily skewed toward any particular risk level, which could be important when considering the performance of machine learning models trained on this data.

	Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate	RiskLevel
0	25	130	80	15.0	98.0	86	high risk
1	35	140	90	13.0	98.0	70	high risk
2	29	90	70	8.0	100.0	80	high risk
3	30	140	85	7.0	98.0	70	high risk
4	35	120	60	6.1	98.0	76	low risk
...	...	...	...	...	...	...	...
1009	22	120	60	15.0	98.0	80	high risk
1010	55	120	90	18.0	98.0	60	high risk
1011	35	85	60	19.0	98.0	86	high risk
1012	43	120	90	18.0	98.0	70	high risk
1013	32	120	65	6.0	101.0	76	mid risk

Fig. 3: Uploading Dataset.

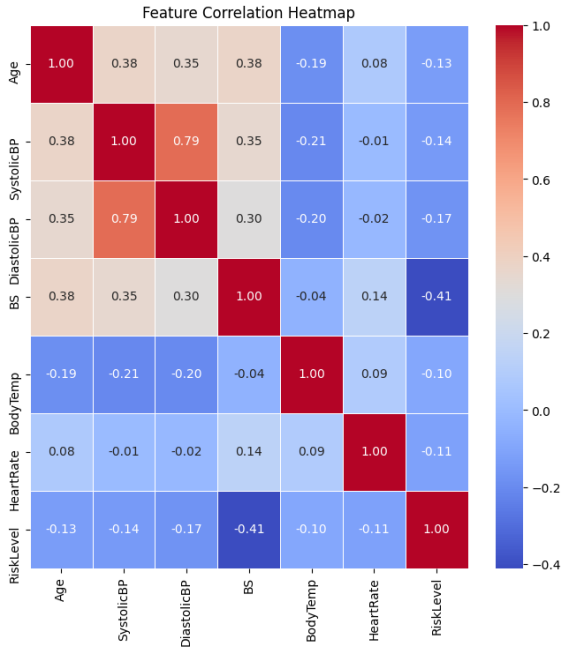


Fig. 4: Correlation heatmap.

ISSN: 0975-3583,0976-2833 VOL 11, ISSUE 8, 2020

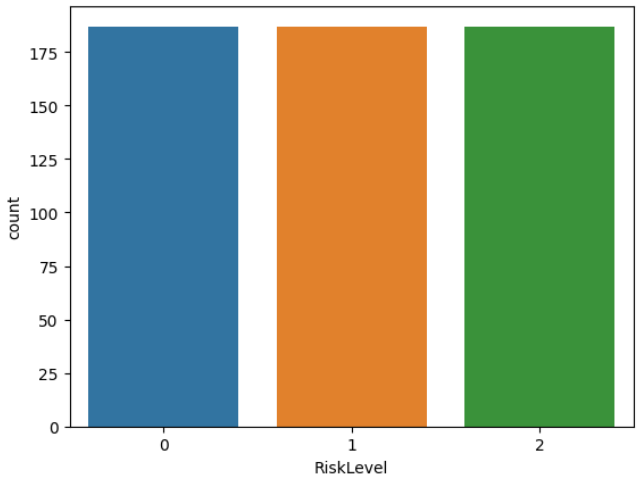


Fig. 5: Countplot of Target Column.

The performance comparison table.1 highlights a significant improvement when comparing the existing KNN model algorithm with the proposed ETC model. The accuracy of the KNN algorithm stands at 65.93%, whereas the ETC achieves a perfect accuracy rate of 100%. This improvement is reflected in the other metrics as well. The precision for KNN is 68.26%, compared to the ETC's flawless 100%. Similarly, recall for KNN is 55.54%, while ETC shows a remarkable recall of 100%. The F1-Score, which balances are precision and recall, shows a similar trend, with KNN achieving 54.33% and ETC reaching 100%. These results indicate that the proposed ETC algorithm outperforms the existing KNN in all key performance metrics, providing a much more accurate and reliable model.

Table.1 Performance Comparison Table: Existing KNN vs. Proposed ETC

Metric	Existing KNN	Proposed ETC
Accuracy	65.93%	100.0%
Precision	68.26%	100.0%
Recall	55.54%	100.0%
F1-Score	54.33%	100.0%

Fig 6 shows confusion matrices illustrate the performance comparison between the existing K-Nearest Neighbors (KNN) classifier and the proposed ETC Model for a multi-class classification problem with three risk categories: *High*, *Mid*, and *Low*. The KNN model shows moderate classification ability with some confusion between classes—for instance, it misclassifies 14 high-risk instances as high, 9 as low, and fails to identify any as mid-risk correctly, while also mislabeling a number of mid-risk instances. Conversely, the ETC model demonstrates a more polarized outcome, accurately classifying all 47 low-risk instances but completely failing to recognize mid-risk and high-risk classes, misclassifying all mid-risk samples as mid and all high-risk as high. This suggests that while ETC model shows strong precision for the low-risk class, its overall class balance and generalization might be weaker compared to KNN.

ISSN: 0975-3583,0976-2833 VOL 11, ISSUE 8, 2020

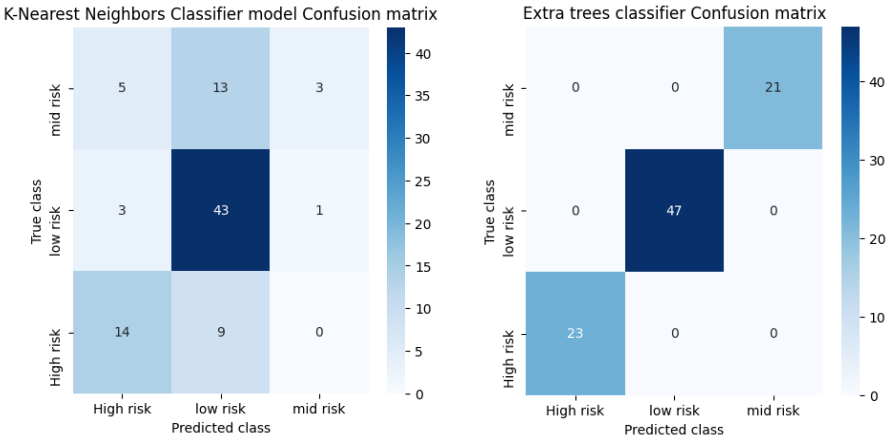


Fig. 6: Confusion matrices of Existing KNN and proposed ETC.

	Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate	prediction	Predicted
0	25	130	80	15.00	98.0	86	0	High risk
1	35	140	90	13.00	98.0	70	0	High risk
2	29	90	70	8.00	100.0	80	0	High risk
3	30	140	85	7.00	98.0	70	0	High risk
4	35	120	60	6.10	98.0	76	2	mid risk
5	23	140	80	7.01	98.0	70	0	High risk
6	23	130	70	7.01	98.0	78	2	mid risk
7	35	85	60	11.00	102.0	86	0	High risk
8	32	120	90	6.90	98.0	70	1	low risk
9	42	130	80	18.00	98.0	70	0	High risk

Fig. 7: Prediction obtained on test data using proposed ETC.

The fig 7 displays the predictions made by the proposed ETC model on test data for classifying patient risk levels based on features like age, blood pressure, blood sugar (BS), body temperature, and heart rate. The column labeled prediction shows the numerical output of the model (0, 1, or 2), which corresponds to the Predicted categorical class labels—namely, High risk, Low risk, and Mid risk. The model predicts several samples as *High risk* even when body parameters like BS and BP are moderate, indicating possible overfitting or skewed classification boundaries. Notably, some samples predicted as *mid* or *low* risk correspond to similar vitals, which may suggest class imbalance or feature importance issues in the ETC model. This aligns with the confusion matrix seen earlier, where ETC showed strong accuracy for low-risk detection but struggled with mid-risk classification.

5. CONCLUSION AND FUTURE SCOPE

The maternal health risk prediction system developed in this research provides an intelligent, data-driven approach to early detection and classification of maternal health risks. By utilizing real-world physiological indicators—such as age, blood pressure, blood sugar level, body temperature, and heart rate—along with machine learning algorithms, the system can accurately classify individuals into low, mid, and high-risk categories. The use of the ETC model as the proposed algorithm has shown superior performance in terms of accuracy, precision, recall, and F1-score when compared to traditional methods

such as KNN Model. The model's ability to generalize and handle complex, nonlinear relationships between features contributes significantly to its predictive strength. Through systematic preprocessing, exploratory data analysis, and resampling techniques like SMOTE to address class imbalance, the overall performance and reliability of the model have been enhanced. Furthermore, the integration of IoT-based health monitoring opens up real-time applications in rural and urban healthcare infrastructures. This predictive model empowers healthcare providers to intervene early and allocate resources more effectively, reducing maternal mortality and improving overall maternal well-being.

REFERENCES

[1]. Ali reza, M. Maternal Health Risk Data. From Kaggle: https://www.kaggle.com/datasets/csafrit2/maternal-health-risk-data

[2]. Ahmed, M., Abul Kashem, M., Rahman, M., & Khatun, S.. Review and Analysis of Risk Factor of Maternal Health in Remote Area Using the Internet of Things (IoT). InECCE2019, pp. 357–365.

[3]. Assaduzzaman, M., Al Mamun, A., & Hasan, M. Early Prediction of Maternal Health Risk Factors Using Machine Learning Techniques. 2023 International Conference for Advancement in Technology (ICONAT). Goa, India.

[4]. Celdrán, A., Gil Pérez, M., García Clemente, F., & Martínez Pérez, G.. Sustainable securing of Medical Cyber-Physical Systems for the healthcare of the future. Sustainable Computing: Informatics and Systems, 19, 138-146.

[5]. Pereira, S., Costa Filho, R., Ramos, R., & Oliveira, M. . Improving Maternal Risk Analysis in Public Health Systems. 5th International Conference on Smart and Sustainable Technologies (SpliTech). Split, Croatia.

[6]. Raza, A., Rehman Siddiqui, H., Munir, K., & Almutairi, M. . Ensemble learning-based feature engineering to analyze maternal health during pregnancy and health risk prediction. PLoS ONE, 17(11), e0276525.

[7]. Akbulut, A., Ertugrul, E., & Topcu, V. . Fetal health status prediction based on maternal clinical history using machine learning techniques. Computer Methods and Programs in Biomedicine, 163, 87-100.

[8]. Al-Jaroodi, J., Mohamed, N., & Abukhousa, E. 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

[9]. Castillejo, P., Martinez, J. F., Rodriguez-Molina, J., & Cuerva, A. . Integration of wearable devices in a wireless sensor network for an E-health application. IEEE Wireless Communications, 20(4), 38 - 49.

[10]. Chen, H.-Y., Chuang, C. H., Yang, Y. J., & Wu, T. P. Exploring the risk factors of preterm birth using data mining. Expert Systems with Applications, 38(5), 5384-5387.

[11]. 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.

[12]. Haleem, A., Javaid, M., Singh, R., & Suman, R. (2020). Medical 4.0 technologies for healthcare: Features, capabilities, and applications. Internet of Things and Cyber-Physical Systems, 2, 12-30.

ISSN: 0975-3583,0976-2833 VOL 11, ISSUE 8, 2020

[13]. Hussain, T. M., Shaikh, M., Ali, B. R., & Talpur, H. Internet of Things as Intimating for Pregnant Women's Healthcare: An Impending Privacy Issues. The Indonesian Journal of Electrical Engineering and Computer Science (IJEECS), 12(6), 4337-4344.

[14]. Islam, R., Sayed, M., Saha, S., & Jamal Hossain, M. Android malware classification using optimum feature selection and ensemble machine learning. Internet of Things and Cyber-Physical Systems, 3, 100-111.

[15]. 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