IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

MATERNAL RISK LEVEL PREDICTION USING ENSEMBLE MODEL

[1] Nirmala, [2] Rekha S Kambli

[1]Department of Electronics and Communications
Government polytechnic Nagamangala, Karnataka, INDIA
[2]Department of Computer Science and Engineering
Government polytechnic K R pet, Karnataka, INDIA

Abstract: In the context of Bangladesh, this work has created a system for accurately monitoring and forecasting a pregnant woman's risk level. Pregnant women's health information and risk factors will be examined by this method to determine the risk intensity level. By 2030, the United Nations wants to lower mother and infant deaths and improve maternal health, but the rate is not declining as quickly as it should. This study evaluated the risk level based on risk factors in pregnancy using the relevant analytical tools and machine learning algorithms. Data on maternal health was acquired for this study from the UCI machine learning library. Risk has been examined using categorization and classification techniques based on risk level. The Random Forest Algorithm provides the highest accuracy of 96% for training data, when it comes to classification and prediction of the risk level, according to a comparison of certain groups of machine learning algorithms.

***Index Terms* - Maternal health risk, Prediction, Random Forest, kNN, Neural network, Naïve Bayes**

I. INTRODUCTION

Many pregnant women die from pregnancy-related illnesses because there is a paucity of information regarding maternal health care during pregnancy and after birth. It primarily affects emerging countries' low middle classes and rural areas [1]. Every stage of pregnancy must be constantly monitored to guarantee a healthy delivery of the child and normal development of the foetus. Large-scale traffic jams, unfavourable weather, pollution, etc. are significant problems for the hospital's ongoing examinations. For any early warning indications, determining the underlying causes of the majority of maternal health-related problems is essential[2]. The results of this research have ensured that pregnant women obtain alternative care, particularly those who reside in remote areas. IoT-based smart health monitoring solutions may be advantageous to them. The ability of the Internet of Things (IoT) to establish connectivity between equipment and humans is extraordinary, which is essential in the healthcare industry to provide around-the-clock monitoring [3].

Maternal health should be considered as a physical entity for the purposes of interfacing with modern technology, such as Cyber-Physical Systems (CPS). CPS, one of the fundamental technologies of Industry 4.0, has ramifications for communication, processing, and control of the multiple states of a physical item. As a result, the main objectives of Industry 4.0 are to use cutting-edge technologies like IOT, Big Data, Block-Chain, Hadoop, etc. to minimise or do away with the current level of complexity in real-world applications [4]. Throughout pregnancy, there are numerous risk factors to keep an eye on. The variables are age, BMI (body mass index), blood oxygen (BO), blood pressure (BP), body temperature, [4] and physical activity. maternal ECG, vaginal discharge in the first trimester, sickness in the first trimester, contractions in the third trimester, foetal heart rate, abnormal foetal protein, electrical uterine activity, and so on. The cutoff thresholds or units of measurement for these variables must be considered [5]. Prompt diagnosis and adequate treatment are the two main priorities for maintaining optimum health throughout pregnancy. Preventing avoidable mother and infant fatalities, especially in rural areas, requires early prenatal care and diagnosis[6].

II. LITERATURE REVIEW

The applications of both data mining and statistical methods are compared for the prediction of diseases using different data set [7-9]. Risk analysis, risk prediction, and implementing devices/tools to diagnose the disease have a common trend nowadays [10, 11]. A heart attack risk prediction study in [12] created a smartphone-based risk prediction tool that revealed diagnostic and prognostic appeal through the analysis of risk factors in recent medical research.

According to WHO and UNICEF, many pregnant women die from preventable diseases. Machine learning techniques are focused on dealing with this kind of crisis recommended by new researchers [5]. Machine learning algorithms are capable of classifying and predicting unknown level high risk situations. Decision trees outperform regression models in terms of accuracy and prediction among all other medical algorithms with low mean absolute error[13, 14].

In this essay, several risk factors for maternal mortality have been thoroughly examined and demonstrated. Risk factors are those things that make getting sick more likely. The classification and forecasting of pregnancy risk intensity is a complex issue. The risk parameter gathered and displayed here is very important. Few methods have been developed to categorise, examine, and forecast the amount of risk for maternal and neonatal health.

III. MATERIALS AND METHODS

Data set used for the study: In order to monitor a pregnant woman's and her foetus' health, this research suggests a model of maternal healthcare. The dataset, which contains around 1014 instances with 7 attributes—including the class attribute—is drawn from the UCI machine learning repository. Low risk, mid risk, and high risk are the three categories of risks that have been taken into consideration. There are 1014 records total, of which 406 were classed as low-risk levels, 336 as mid-risk levels, and 272 as high-risk levels[5] . Age, body temperature, blood sugar, heart rate, and systolic and diastolic blood pressure are the independent characteristics considered. Figure 1 is the violin curve which illustrates the attributes' detailed statistics, and violin plots display each variable's density and summary statistics. Figure 2 and 3 shows how attributes are distributed according to risk levels.

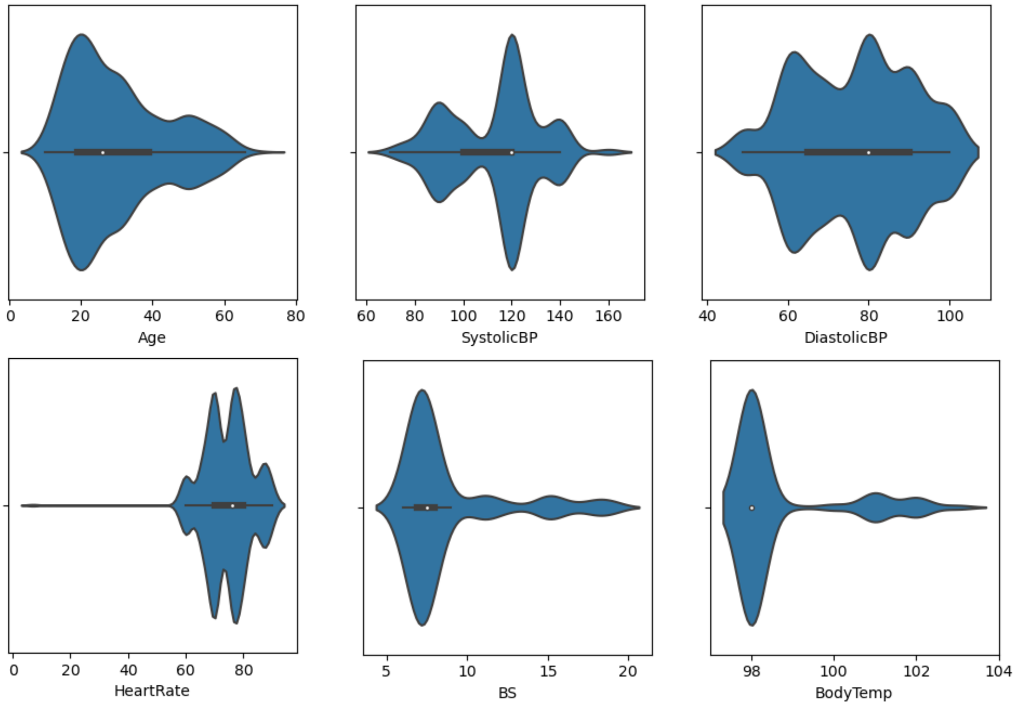

Figure1.Voiline curve of the attributes

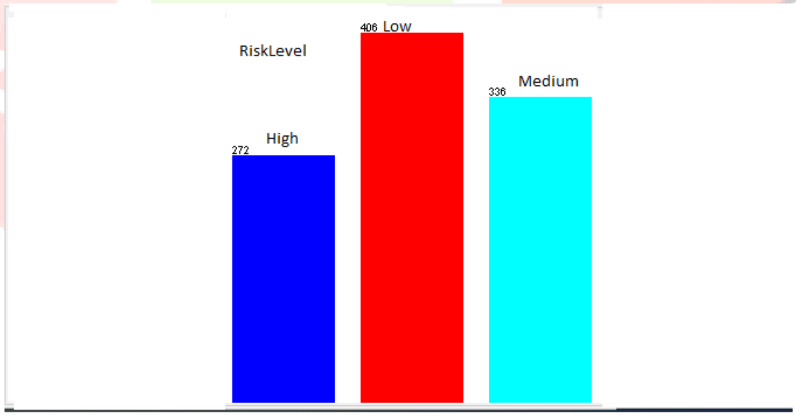

Figure 2. Distribution of different risk levels

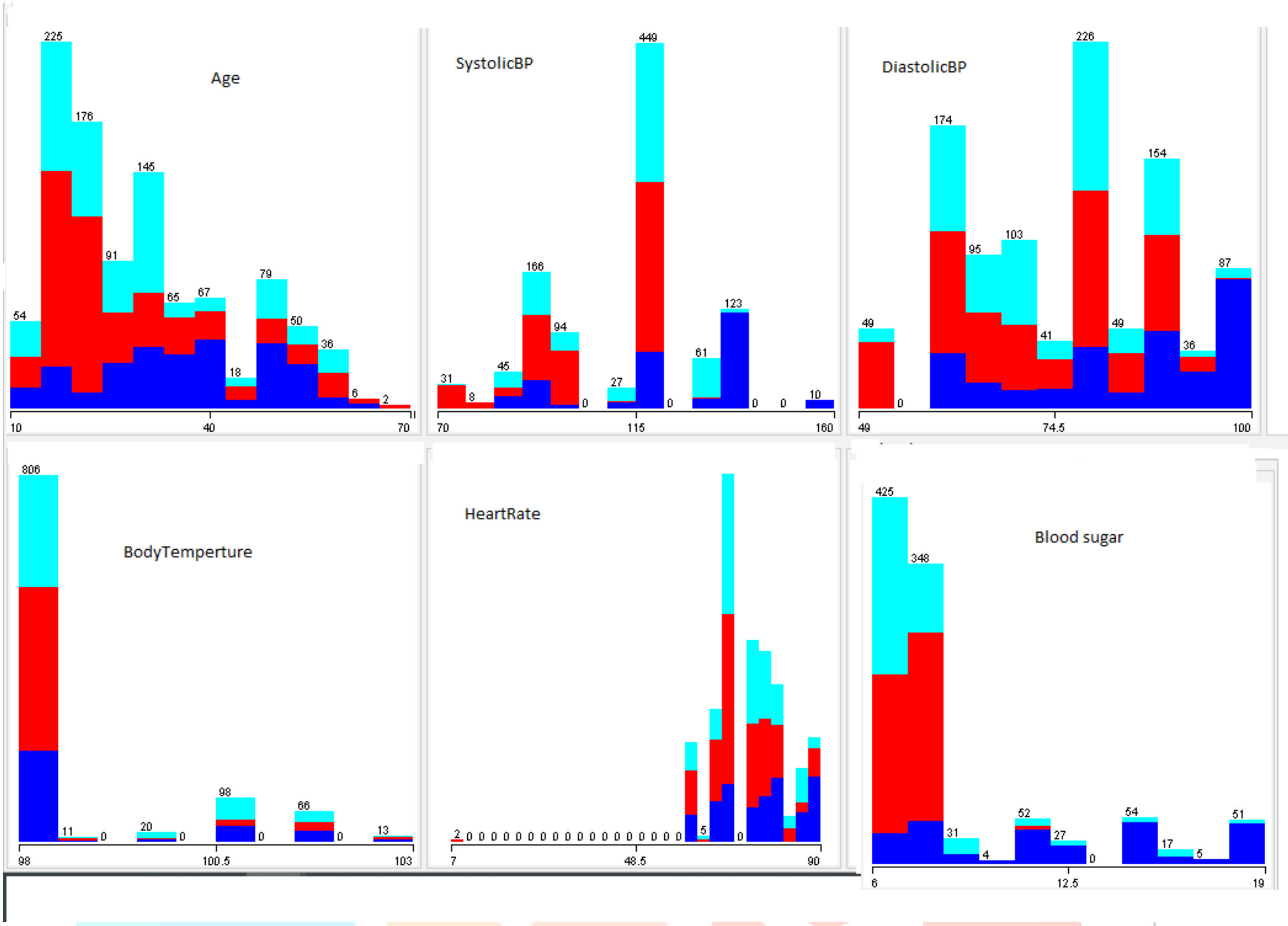

Figure 3: The distribution in attributes values in different risk levels

Methodology: The data set is used to implement the four different classification models, kNN, Random forest, Naive Bayes, and neural network model. Also, the kNN, Naive Bayes, and neural network model were considered when developing the stacking ensemble model. In the section that follows, the classification models are explained in detail.

K Nearest Neighbor: The 'feature similarity' component of the KNN method is used to forecast the values of any additional data points. In other words, the value given to the new point depends on how much it resembles the points in the training set. By averaging the observations in the same neighbourhood, KNN regression is a non-parametric technique that intuitively approximates the relationship between independent variables and the continuous result.

Random Forest: An approach for supervised machine learning called Random Forest (RF) builds a forest and randomises it. A forest, or collection of Decision Trees, is trained using the bagging approach. To produce a reliable and accurate classification, Random Forest builds many decision trees and combines them. The Random Forest algorithm's main benefit is that it may be used for both classification and regression analysis. [15].

Neural Network: Neural networks (NN), is a learning model whose operation is impacted by the operation of a biological neuron. The neural network is made up of nodes, which process the data provided to them as input and transmit the results to other nodes. The activation or node value is referred to as each node's output. Weights attached to the nodes can be changed to aid network learning. These weights show how much an input may or may not influence an outcome.

Naïve Bayes: Naive Bayes is a straightforward method for building classifiers. These models assign class labels to problem cases, which are represented as vectors of feature values, and the class labels are chosen from a finite set. For training such classifiers, there isn't just one technique, but rather a family of algorithms built on the premise that, given the class variable, the value of one feature is independent of the value of every other feature.

Stacking: A machine learning ensemble algorithm is called stacking or stacked generalisation. To determine how to combine the predictions from two or more underlying machine learning algorithms, it employs a meta-learning algorithm. The advantage of stacking is that it can use a variety of effective models to accomplish classification or regression tasks and produce predictions that perform better than any one model in the ensemble.

IV. RESULTS AND DISCUSSION

The prediction model is applied for different set attributes. The data set in divided into 70% training and 30% testing. The first Several set attributes are applied to the prediction model. The data set is split into 30% for testing and 70% for training. The initial model used is kNN, where 15 neighbours are considered. Following that, a random forest model is produced using around 10 trees, and a neural network model is constructed using 100 neuron hidden layers and ReLu activation function. 50 estimators are used to create the AdaBoost model, with a tree serving as the base estimator. The same classifiers are used to predict maternal risk using training and test data. Tables 1 and 2 for the training data and testing data, respectively, summarise the results of the classification models in terms of accuracy, F1 measure, precision, and recall.. The classifiers are compared using calibration curve as shown in figure 3, Comparisons of calibration curves, often called reliability diagrams, show how effectively a binary classifier's probabilistic predictions are calibrated. For binned predictions, it plots the actual frequency of the positive label against its expected likelihood.

Table 1: Performance of the prediction model for training data

Model	Accuracy	F1	Precision	Recall
kNN	0.74	0.72	0.74	0.74
Stack	0.78	0.74	0.77	0.78
Random Forest	0.89	0.89	0.89	0.89
Neural Network	0.77	0.74	0.76	0.77
Naive Bayes	0.75	0.71	0.73	0.75

Table 2: Performance of the prediction model with testing data

Model	Accuracy	F1	Precision	Recall
kNN	0.53	0.51	0.65	0.53
Stack	0.56	0.52	0.71	0.56
Random Forest	0.83	0.83	0.85	0.83
Neural Network	0.57	0.55	0.67	0.57
Naive Bayes	0.54	0.51	0.65	0.54

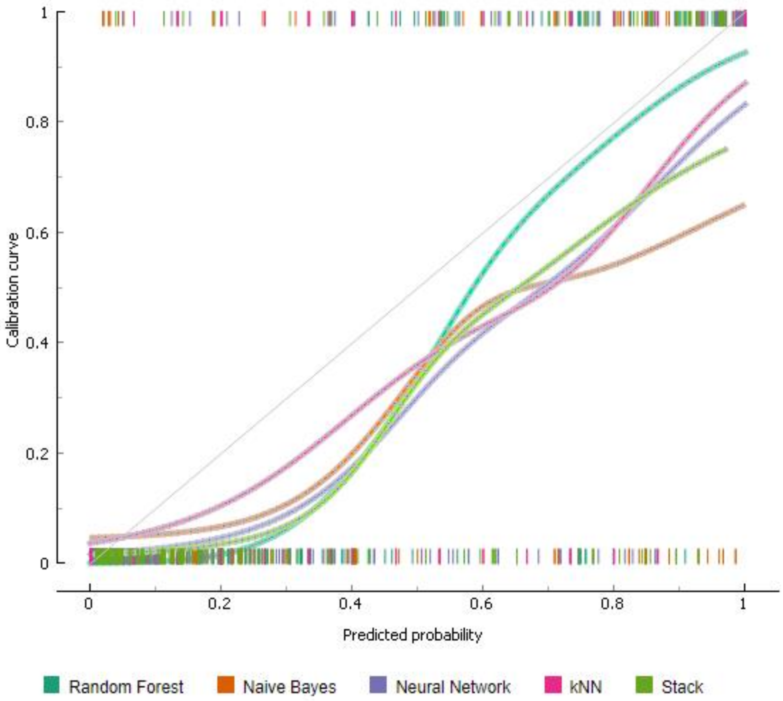

Figure 3 Calibration curve

The performances of the prediction models are compared in terms pf accuracy, Area under Curve (AUC), F1 measure, precision and recall. The highest accuracy of 89% and 83% is obtained by random forest prediction model for training and testing data respectively.

V. CONCLUSION AND FUTURE WORK

To identify the significant risk component for evaluating, categorising, and predicting risk intensity, a different strategy has been taken. The dataset's maximum accuracy is provided by the ensemble random forest model.

Future research on maternal and foetal health, as well as immunology, can be made possible by the addition of more attributes to the analysis. The accuracy can be further increased by using a more accurate hybrid model.

.

REFERENCES

1. Ahmed, M., et al., Review and analysis of risk factor of maternal health in remote area using the Internet of Things (IoT). InECCE2019, 2020: p. 357-365.
2. Say, L., et al., Global causes of maternal death: a WHO systematic analysis. The Lancet global health, 2014. 2(6): p. e323-e333.
3. Talpur, M.S.H., et al., Internet of things as intimating for pregnant women's healthcare: an impending privacy issues. TELKOMNIKA Indonesian Journal of Electrical Engineering, 2014. 12(6): p. 4337-4344.
4. Mustafa, R., et al., A comprehensive review of hypertension in pregnancy. Journal of pregnancy, 2012. 2012.
5. Ahmed, M. and M.A. Kashem. IoT based risk level prediction model for maternal health care in the context of Bangladesh. in 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI). 2020. IEEE.
6. Santhi, V., et al., IOT based wearable health monitoring system for pregnant ladies using cc3200. International Journal of Advanced Research Methodology in Engineering & Technology, 2017. 1(3): p. 56-59.
7. Ahmed, K. and T. Jesmin, Comparative analysis of data mining classification algorithms in type-2 diabetes prediction data using weka approach. International Journal of Science and Engineering, 2014. 7(2): p. 155-160.
8. Ahmed, K., T. Jesmin, and M.Z. Rahman, Early prevention and detection of skin cancer risk using data mining. International Journal of Computer Applications, 2013. 62(4).
9. Ahmed, K., et al., Prediction of breast cancer risk level with risk factors in perspective to bangladeshi women using data Mining. International Journal of Computer Applications, 2013. 82(4).
10. Thomas, G., et al., A novel method for interrogating receiver operating characteristic curves for assessing prognostic tests. Diagnostic and prognostic research, 2017. 1(1): p. 1-9.
11. van Doorn, S., et al., The effects of misclassification in routine healthcare databases on the accuracy of prognostic prediction models: a case study of the CHA2DS2-VASc score in atrial fibrillation. Diagnostic and prognostic research, 2017. 1(1): p. 1-9.
12. Raihan, M., et al. Smartphone based ischemic heart disease (heart attack) risk prediction using clinical data and data mining approaches, a prototype design. in 2016 19th International Conference on Computer and Information Technology (ICCIT). 2016. IEEE.
13. Mehta, R., N. Bhatt, and A. Ganatra, A survey on data mining technologies for decision support system of maternal care domain. International Journal of Computers and Applications, 2016. 138(10): p. 20-4.
14. Abegaz, K.H. and E.M. Habtewold, Trend and barriers of antenatal care utilization from 2000 to 2016 Ethiopian DHS: a data mining approach. Scientific African, 2019. 3: p. e00063.
15. Hassan, M.M., et al. Diabetes Prediction in Healthcare at Early Stage Using Machine Learning Approach. in 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT). 2021.