MODEL FOR PREDICTING RISK LEVELS IN MATERNAL HEALTHCARE

Rekha S Kambli[1*], Nirmala[2]

[1]Department of Computer science and Engineering, Government polytechnic K R pet, INDIA
[2]Department of electronics and communications, Government polytechnic K R pet, INDIA

ABSTRACT

In the context of Bangladesh, a system has been built in this study for efficiently tracking and forecasting a pregnant woman's risk level. To determine the risk intensity level, this system will analyze the health information and risk factors of pregnant women. By 2030, the United Nations wants to improve maternal health and reduce mother and infant mortality; however, the rate is not decreasing as quickly as it should be. This study aimed to determine the risk level based on risk factors in pregnancy using appropriate analytical tools and machine learning algorithms. Data on maternal health were retrieved from the UCI machine learning repository for this study. According to the level of risk, categorization and classification methods have been utilized for the examination of risk variables. The AdaBoost Algorithm provides the highest AUC of 96%, when it comes to classification and prediction of the risk level, according to a comparison of certain groups of machine learning algorithms.

.

Keyword : - *Maternal health, Prediction, Random forest, kNN, Neural network, AdaBoost*

---

1. INTRODUCTION

There is a lack of information regarding maternal health care during pregnancy and after delivery, many pregnant women pass away from pregnancy-related illnesses. The poor middle class and rural parts of developing nations are particularly affected by it [1]. To ensure the safe delivery of the baby and the proper development of the foetus, every moment throughout pregnancy needs to be closely watched. Massive traffic congestion, unfavourable weather, pollution, etc. are major issues for the hospital's ongoing checkups. Finding the underlying causes of the majority of maternal health-related issues is crucial for any early warning signs [2].

This research has made sure that pregnant women receive alternative treatment, especially those who live in rural locations. They may benefit from IoT-based smart health monitoring systems. The Internet of Things (IoT) is incredibly capable of establishing communication between machines and people, which is crucial in the health care sector to provide round-the-clock monitoring [3].

For the purpose of interacting with contemporary technology, such as Cyber-Physical Systems, maternal health should be viewed as a physical entity (CPS). Communication, computation, and control of a physical entity's various states are implications of CPS, one of Industry 4.0's key technologies. As a result, the primary goals of Industry 4.0 are the utilization of cutting-edge technologies like IOT, Big Data, Block-Chain, Hadoop, etc. to reduce or eliminate current real-world complexity.

There are many risk factors to monitor throughout pregnancy. Age, Body Mass Index (BMI), Blood Oxygen (BO), Blood Pressure (BP), Body temperature, [4] and Physical Activity are the variables. maternal ECG, vaginal discharge during the first trimester, nausea during the first trimester, contractions during the third trimester, foetal heart rate (FHR), abnormal foetal protein (AFP), electrical uterine activity (EUA), foetal movement activity, and so forth. We must take into account the cutoff points or units of measurement for these variables [5].

The key focus for preserving excellent health throughout pregnancy is prompt diagnosis and appropriate treatment. Early prenatal care and diagnosis are crucial for preventing unnecessary maternal and newborn deaths, especially in rural settings [6].

2. 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 diagnosis 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]. Unknown level high risk prediction and classification can be done by machine learning algorithms. Among all other medical algorithms with low mean absolute error, decision trees have the greatest advantage over regression models in terms of accuracy and prediction [13, 14].

In this paper detailed analysis of a few risk factors of maternal mortality has been done and shown. Risk factors mean those reasons, which increase the chance of getting a disease. The classification and prediction of risk intensity during pregnancy is a multilayered problem. The risk parameter collected and shown here is highly significant. Few approaches have been designed to classify, analyze, and predict the risk level of maternal and neonatal health.

3. PROPOSED METHODOLOGY

This paper proposes a maternal healthcare model for observing a pregnant woman and fetal health. Data is collected from UCI machine learning repository, the dataset contains about 1014 instances with 7 attributes including the class attribute. Three categorical risks have been considered and collected as low risk, mid risk, and high risk. The total data size is 1014 where 406 was classified as low-risk level, 336 in mid and 272 was in high-risk level[5] .The details statistics of the attributes is tabulated in table 1.The independent attributes considered are age, body temperature, blood sugar, heart rate, systolic and diastolic blood pressure.

Table 1 Statistical detail of the attributes

	Age	**Systolic BP**	**Diastolic BP**	**Blood Sugar**	**Body Temp**	**Heart Rate**
Mean	29.87	113.20	76.46	8.73	98.67	74.30
SD	13.47	18.40	13.89	3.29	1.37	8.09
Min	10.00	70.00	49.00	6.00	98.00	7.00
Max	70.00	160.00	100.00	19.00	103.00	90.00

The correlation analysis of the attributes in done using Pearson correlation factor as shown in figure 1.The attributes are ranked based on the correlation factor. The prediction the risk is done based on the attributes rank using major prediction algorithm namely Random forest, neural network and k NN. An ensemble Meta algorithm AdaBoost is also built for the dataset in order to improve the accuracy.

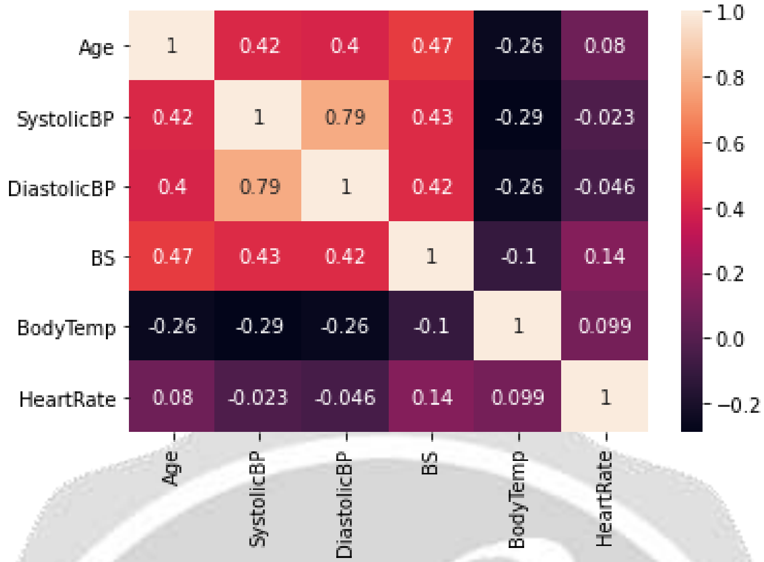

Fig 1: Correlation between the attributes

4. RESULTS AND DISCUSSION

The correlation shows that there is strong positive correlation between the attributes age and blood sugar. Both the Blood pressures are associated positively with the blood sugar level. From the correlation values a strong association between blood sugar, BP and age can be observed.

The prediction model is applied for different set attributes. Standard 10 fold cross validation process is done in all the experiments. This process allows the classifiers divides the data into 9 and 1 fold in which, 9 folds of data is used for training and 1 unused fold for testing. The first models applied is kNN where the number of neighbors considered are 15, the random forest is implemented by taking about 10 trees and neural network model is built with 100 neuron hidden layers ,ReLu activation function. The AdaBoost model is implemented with tree as base estimator and 50 estimators. The prediction of maternal risk is done by using same classifiers by using all the attributes and selected attributes. The Blood sugar, systolic and diastolic blood pressures are the top three attributes selected based on info gain, correlation. Table 2 shows the accuracy of the classifiers for the selected features and the accuracy of classifiers with all features. The Roc analysis of the models with all the features is shown in figure 2 , the analysis is done by taking high risk as the target class and 27% of target probability with cost as false positive =500 and false negative =500. Performances of the classifiers with all the features are tabulated in table 3.

Table 2: performance of the prediction model with top 3 attributes

Model	AUC	**Accuracy**	F1	Precision	Recall
kNN	0.84	0.68	0.68	0.68	0.68
Random Forest	0.91	0.77	0.77	0.77	0.77
Neural Network	0.81	0.66	0.62	0.68	0.66
AdaBoost	0.91	0.77	0.77	0.77	0.77

Table 3: performance of the prediction model with all the attributes

Model	**AUC**	**Accuracy**	**F1**	**Precision**	**Recall**
kNN	0.83	0.64	0.64	0.64	0.64
Random Forest	0.94	0.82	0.83	0.83	0.82
Neural Network	0.85	0.69	0.68	0.68	0.69
AdaBoost	0.96	0.86	0.86	0.86	0.86

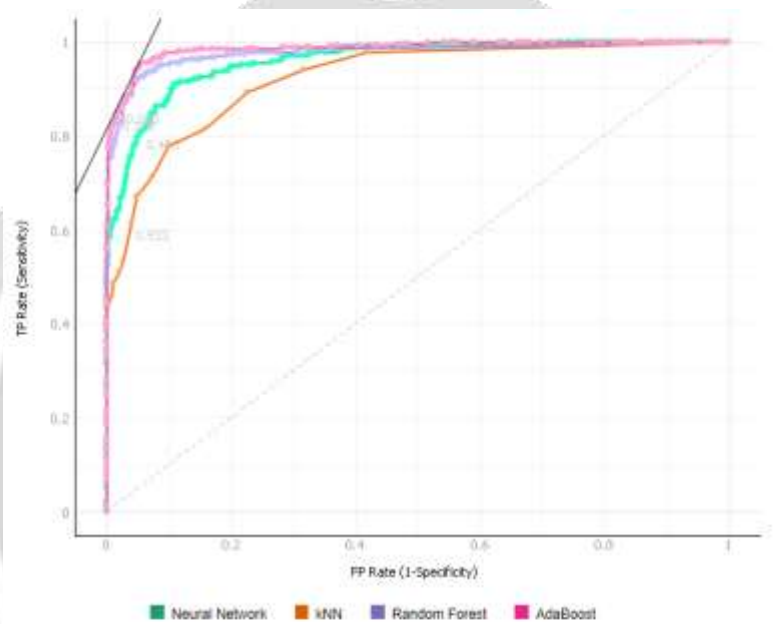

Fig 2 ROC analysis taking high risk as the target class

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 77% is obtained with top 3 attributes by AdaBoost and random forest prediction model, whereas the highest accuracy of 86% is obtained with all the 6 attributes by AdaBoost model.

5. CONCLUSION AND FUTURE WORK

A different approach has been done to find out the significant risk factor for analyzing, classifying and predicting the intensity of risk. The ensemble model AdaBoost is giving highest accuracy for the dataset. The association analysis between the attributes are also done using Pearson correlation analysis
In the future, more attributes can be added to make the vast analysis as well as research on maternal and fetal health and immunology. A better hybrid model can be implemented, which can improve the accuracy further.

6. REFERENCES

[1] M. Ahmed, M. A. Kashem, M. Rahman, and S. Khatun, "Review and analysis of risk factor of maternal health in remote area using the Internet of Things (IoT)," *InECCE2019*, pp. 357-365, 2020.

[2] L. Say, D. Chou, A. Gemmill, Ö. Tunçalp, A.-B. Moller, J. Daniels, *et al.*, "Global causes of maternal death: a WHO systematic analysis," *The Lancet global health,* vol. 2, pp. e323-e333, 2014.

[3] M. S. H. Talpur, M. H. Shaikh, R. A. Buriro, H. S. Talpur, F. Talpur, H. Shafi, *et al.*, "Internet of things as intimating for pregnant women's healthcare: an impending privacy issues," *TELKOMNIKA Indonesian Journal of Electrical Engineering,* vol. 12, pp. 4337-4344, 2014.

[4] R. Mustafa, S. Ahmed, A. Gupta, and R. C. Venuto, "A comprehensive review of hypertension in pregnancy," *Journal of pregnancy,* vol. 2012, 2012.

[5] M. Ahmed 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, pp. 1-6.

[6] V. Santhi, K. Ramya, A. Tarana, and G. Vinitha, "IOT based wearable health monitoring system for pregnant ladies using cc3200," *International Journal of Advanced Research Methodology in Engineering & Technology,* vol. 1, pp. 56-59, 2017.

[7] K. Ahmed 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,* vol. 7, pp. 155-160, 2014.

[8] K. Ahmed, T. Jesmin, and M. Z. Rahman, "Early prevention and detection of skin cancer risk using data mining," *International Journal of Computer Applications,* vol. 62, 2013.

[9] K. Ahmed, M. A. Habib, T. Jesmin, M. Z. Rahman, and M. B. A. Miah, "Prediction of breast cancer risk level with risk factors in perspective to bangladeshi women using data Mining," *International Journal of Computer Applications,* vol. 82, 2013.

[10] G. Thomas, L. C. Kenny, P. N. Baker, and R. Tuytten, "A novel method for interrogating receiver operating characteristic curves for assessing prognostic tests," *Diagnostic and prognostic research,* vol. 1, pp. 1-9, 2017.

[11] S. van Doorn, T. B. Brakenhoff, K. G. Moons, F. H. Rutten, A. W. Hoes, R. H. Groenwold, *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,* vol. 1, pp. 1-9, 2017.

[12] M. Raihan, S. Mondal, A. More, M. O. F. Sagor, G. Sikder, M. A. Majumder, *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, pp. 299-303.

[13] R. Mehta, 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,* vol. 138, pp. 20-4, 2016.

[14] K. H. Abegaz and E. M. Habtewold, "Trend and barriers of antenatal care utilization from 2000 to 2016 Ethiopian DHS: a data mining approach," *Scientific African,* vol. 3, p. e00063, 2019.