ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 9, September 2023

Maternal Health Risk Analysis by using Exploratory Data Analysis and Machine Learning Algorithms

[1] Ayesha Siddika, [2] Mim Sultana

[1] PhD Student, Department of CSE, Dhaka University of Engineering and Technology (DUET), Dhaka, Bangladesh
[2] BSc., Student, Department of CSE, World University of Bangladesh (WUB) Dhaka, Bangladesh
Corresponding Author Email: [1] ayeshareba.cse@gmail.com, [2] mimsultana199@gmail.com

Abstract— ***Maternal patients, such as when a mother thinks her blood pressure may rise, her glucose may rise, or fall, and there is a lot of data that can harm the baby. Here we analyzed that data and trained it with different machine learning algorithms. These are Xgboost, Decision Tree, Random Forest and Support Vector Machine, and Naive Bayes Algorithms. Maternal health risk analysis has been done by many, but no one has produced results that are more accurate than 88%. But 3 of our models show a better performance compared to the existing ones. We have 94% accuracy in those three models; these are Xgboost, Decision Tree and Random Forest Algorithm. At the same time, SVM has 72% accuracy and Naive Bayes has 64% accuracy.***

Index Terms—Maternal Health, Data Analysis, Machine Learning, AI, Xgboost, Decision Tree, Random Forest and Support Vector Machine, and Naive Bayes Algorithms.

I. BACKGROUND

With important contributing factors such as racial and ethnic health disparities, access to care, and the existence of other health issues, the nation's maternal mortality and morbidity rates have gradually climbed over the past 20 years. When a mother thinks, for example, her blood pressure may rise, her blood sugar levels may rise or fall, and there is a lot of information that can harm the baby. The rates of maternal mortality and morbidity have increased recently. So, in this research, we tried to create a system that uses machine learning to assess patient data and forecast who is at risk. A subfield of artificial intelligence (AI) called machine learning may be devoted to creating software programs that learn from data and gradually increase their accuracy without being explicitly programmed to do so. An algorithm in data science might be a series of statistical processing processes. In machine learning, algorithms are ”trained” to look for patterns and features within vast volumes of data in order to make judgments and predictions based on fresh information. Because it analyses more data, an algorithm at a higher level will produce predic- tions and decisions that are more accurate. Maternal Health Risk Analysis Using EDA and Machine Learning is the topic of this thesis. Here, use these facts to analyze. In this paper, we use machine learning to train it and then display the results. To complete this work, five algorithms used EDA and machine learning. These five algorithms are the Gaussian Naive Bayes, Support Vector Machine (SVM), Xgboost, Random Forest, and Decision Tree. In this system, these algorithms can produce the best and most accurate outcomes.

II. OBJECTIVES

A. To analyze the risk to a mother's health by using exploratory data analysis and machine learning algorithms
B. To compare the analysis result with some existing results to find out the better prediction.

III. SCOPE OF STUDY

There are numerous studies that employ Exploratory Data Analysis (EDA) and machine learning to analyze the risk to a mother's health. Maternal health risk analysis has been done by many, but no one has produced results that are more accurate than 88%. But more accuracy should be achieved by focusing on this topic because it is so crucial for both mother and child. These research papers make use of a variety of approaches. But here we are trying to apply five different techniques in this research at a time. After gathering data, they manually checked and found that there was some missing data. This is the initial attempt to fill in some missing data rows. The dataset is then converted to a numerical representation because SVM requires numbers as inputs. That was the obstacle. Nobody uses those algorithms on the same dataset at the same time. Doctors and patients can both use this and also benefit from this.

IV. LITERATURE REVIEW

The main part of this work was to evaluate the effect of breastfeeding on long-term and short-term) maternal health outcomes. In this work, they used PubMed, Cochrane Library and CABI databases and explored subgroup analysis and meta-regression. They found breastfeeding greater than 12 months was associated with reduced risk of breast and ovarian carcinoma by 26% and 37%, respectively.

Breastfeeding was associated with 32% lower risk of type 2 diabetes. This review supports the hypothesis that breastfeeding is protective against breast and ovarian carcinoma, and exclusive breastfeeding and predominant breastfeeding increase the duration of lactation amenorrhea [1].

In this review and analysis, the main aim was to investigate the impact of male involvement on maternal health outcomes of women in developing countries. They focused on four electronic databases and grey literature sources were searched (up to May 2013). And they found that fourteen studies met the inclusion criteria. Male involvement was significantly associated with reduced odds of postpartum depression (OR=0.36, 95% CI 0.19 to 0.68 for male involvement during pregnancy; OR=0.34, 95% CI 0.19 to 0.62 for male involvement post partum), and also with improved utilization of maternal health services (skilled birth attendance and postnatal care). Male involvement during pregnancy and at post partum appeared to have greater benefits than male involvement during delivery [2].

A meta-analysis and review of pregnancy complications and behavioral risk factors associated with's low birth weight and other poor outcomes which occur during adolescent pregnancy was undertaken using the published literature. Methods: Stud- ies were eligible for inclusion if they: 1) utilized a clearly defined sample of teenagers 2) provided numeric data on complications of interest or the proportions needed to compute this information 3) included a control or comparison group.Many behavioral risk factors (smoking, drinking and drug use) appeared to be less prevalent among teenage gravidas, par- ticularly when the young women were ethnic minorities. An increased risk of preterm delivery was associated with young maternal age in both developed and developing countries [3].

The objective of this study was to estimate the effects of individual, community and district level characteristics on the utilisation of maternal health services with special reference to antenatal care (ANC), skilled attendance at delivery and post- natal care (PNC). This study was designed as a cross sectional study. Data from 15,782 ever married women aged 15-49 years residing in Madhya Pradesh state of India who participated in the District Level Household and Facility Survey (DLHS-3) 2007-08 were used for this study. The results of this study showed that 61.7% of the respondents used ANC at least once during their most recent pregnancy whereas only 37.4% of women received PNC within two weeks of delivery [4].

Sensing and Artificial Intelligent Maternal Infant Health Care Systems. This article reviews the arable sensors and AI algorithms based on existing systems designed to predict the risk factors during and after pregnancy for both mothers and infants. This review covers sensors and an AI algorithm used in these systems and analyzes each approach with its features, outcomes, and novel aspects in chronological or- der. It also includes discussion on datasets used and extends challenges as they are future work directions for researchers. [6]

The aim of the paper is a decision support system based on a three-stage classifier that has been developed. Each stage acts as a filter and allows re-evaluation of the classification made in the previous stage in order to find diagnostic errors. This classifier has been implemented and tested for four different aid algorithms: Multilayer Perception, Deep Learning, Support Vector Machine, and NativesBayes [5].

A decision support system for predicting the treatment of ectopic pregnancies. The aim of the paper is that Ectopic pregnancy is an important cause of morbidity and mortality worldwide. An early diagnosis, as well as the choice of the most suitable treatment for the patient, is crucial to avoid possible complications. According to different factors, an ectopic pregnancy must be treated through surgery, using pharmacological treatment, or following conservative treatment. In this paper, a clinical decision support system based on artificial intelligence algorithms has been developed to help clinicians to choose the initial treatment to be followed by the patient [5].

Machine learning for maternal health: Predicting delivery location in a community health worker program in Zanzibar. They use program data from 38,787 women enrolled in Safer Deliveries, a community health worker program in Zanzibar, to build a generalizable prediction model that accurately predicts whether a newly enrolled pregnant woman will deliver in a health facility [6]. They use information collected during the enrollment visit, including demographic data, health char- acteristics, and current pregnancy information. They apply four machine learning methods: logistic regression, LASSO regularized logistic regression, random forest, and an artificial neural network; and three sampling techniques to address the imbalanced data: under sampling of facility deliveries, oversampling of home deliveries, and addition of synthetic home deliveries using SMOTE [6].

The timely monitoring pregnant women and fetus vital health parameters and effective treatment can make the dif- ference between life and death for both the mother and the baby. According to WHO, 830 women die every day due to the complexities of childbirth. The existing research works are focused on maternal health surveillance methods. The proposed work presents the comprehensive study on the methods employed to monitor the health of pregnant women and fetus. The advantages and disadvantages of the existing methods are discussed in more detail. The challenging issues in monitoring the health of the pregnant women and fetusare presented in a coherent manner. The unresolved challenging issues in this area are presented in detail, which serves as an open-ended research area concerning the problem of maternal and fetal health monitoring [7]

A Novel Health Monitoring Approach for Pregnant Women. The aim of the paper is to highlight the impact of parameter standardization on prediction accuracy achieved in the present research. The performance of the C4.5 decision tree classifica- tion algorithm selected for study in terms of accuracy obtained when applied to collected and standardized pregnancy data-set is also analyzed in the paper [8].

Currently, information and communication technology (ICT) allows health institutions to reach disadvantaged groups in rural areas using sensing and artificial intelligence (AI) technologies. Applications of these technologies are even more essential for maternal and infant health, since maternal and infant health is vital for a healthy society. Over the last few years, researchers have delved into sensing and artificially intelligent healthcare systems for maternal and infant health. Sensors are exploited to gauge health parameters, and machine learning techniques are investigated to predict the health conditions of patients to assist medical practitioners. Since these healthcare systems deal with large amounts of data, significant development is also noted in the computing platforms. The relevant literature reports the potential impact of ICT-enabled systems on improving maternal and infant health. This article reviews wearable sensors and AI algorithms based on existing systems designed to predict the risk factors during and after pregnancy for both mothers and infants. This review covers sensors and AI algorithms used in these systems and analyze each approach with its features, outcomes, and novel aspects in chronological order. It also includes discussion on datasets used and extends challenges as well as future work directions for researchers [9].

V. METHODOLOGY

A method is a structured procedure for bringing about a certain goal. This is the methodology that chooses in the study of this research methods. Input data, Preprocessing, Cleaned Data, Train Data, Evaluate model, Comparing Model, and results steps are included here.

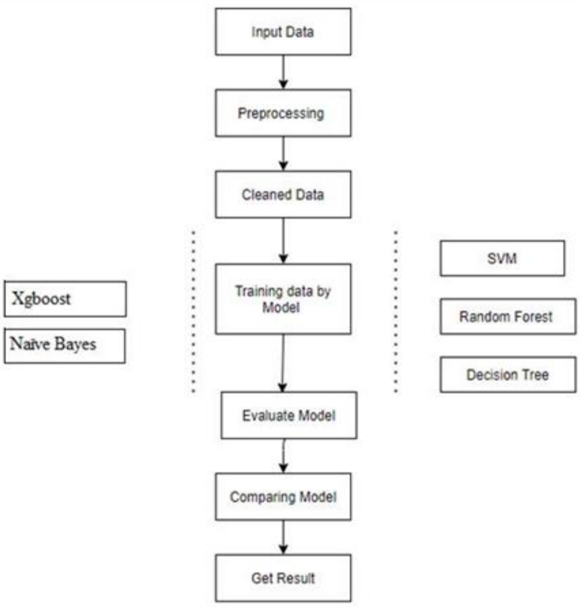

Figure 1. Proposed Methodology

A. Data Source and Collection

Obtain data from the webpage of the UCI Machine Learning Repository. The University of California, Irvine's Center for Machine Learning and Intelligent Systems is responsible for hosting and maintaining it. As a PhD student at UC Irvine, David Aha first developed it. Every dataset has a webpage with all the information that is currently available about it, including any pertinent research articles. The actual datasets are available for download as ASCII files, frequently in the practical CSV format. Well-researched datasets and real data are used. Author track age, systolic and diastolic blood pressure, body temperature, heart rate, and risk level attributes, when the patient's name is a distinguishing feature.

B. Data Prepossessing

Age, Systolic BP, Diastolic BP BS, Body Temp, Heart Rate, and Risk Level characteristic are the 5 attributes collected. Where a contributor to the Age, Systolic BP, Diastolic BP, BS, Body Temp, and Heart Rate columns

```
df = pd.read_csv('../input/maternal-health-risk-data/Maternal Health Risk Data
df.head()
```

	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

Figure 2. 5 attributes for data processing

C. Implemented Algorithm

A supervised learning method, which functions based on training and testing, was employed in this paper. The classification model is built using the training dataset. The outcome is then obtained by applying this created model to the testing dataset. The machine-learning algorithm will be swiftly illustrated in the following sections.

D. Testing and Result

Following implementation, there have a result that meets the criteria. The accuracy was 94 percent when compared it to another similar existing algorithm after that.

VI. WORKING PROCEDURE

Many pregnant women die from pregnancy issues as a result of a lack of information on maternal health care during and after pregnancy. It is more common in rural regions and among lower-middle-class families in emerging countries. During pregnancy, every minute should be observed to ensure the proper growth of the baby and the safe

ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 10, Issue 9, September 2023

delivery.Here we used Exploratory Data Analysis (EDA) and 5 algorithms Classification. Those are: Support Vector Machine (SVM), Xg- boost, Decision Tree, Random Forest, and Gaussian Naive Bayes.

Data Set Information: Data has been collected from different hospitals, community clinics, maternal health care's through the IoT based risk monitoring system.

Age: Age in years when a woman is pregnant.

Systolic BP: Upper value of Blood Pressure in mmHg, another significant attribute during pregnancy.

Diastolic BP: Lower value of Blood Pressure in mmHg, another significant attribute during pregnancy.

Blood Sugar: Blood glucose levels is in terms of a molar concentration, mmol/L.

Heart Rate: A normal resting heart rate in beats per minute. **Risk Level**: Predicted Risk Intensity Level during pregnancy considering the previous attribute.

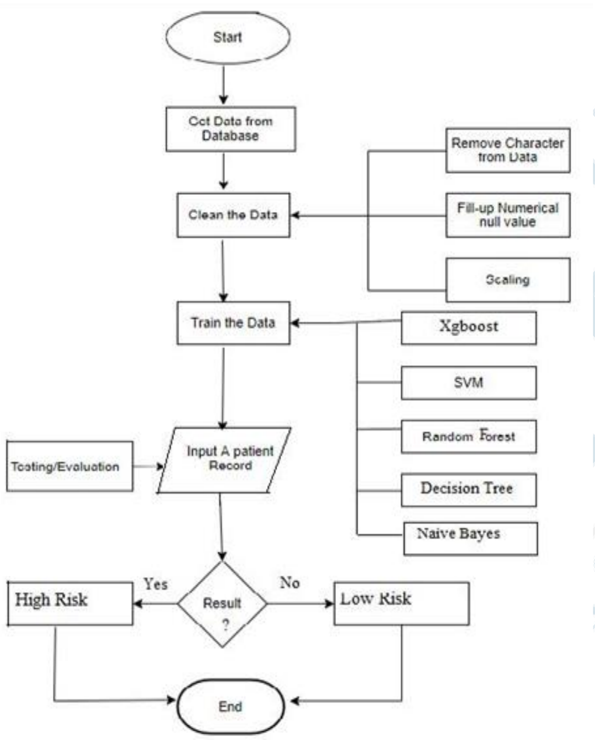

Figure 3. Flowchart of the executing method

VII. GRAPHICAL REPRESENTATION

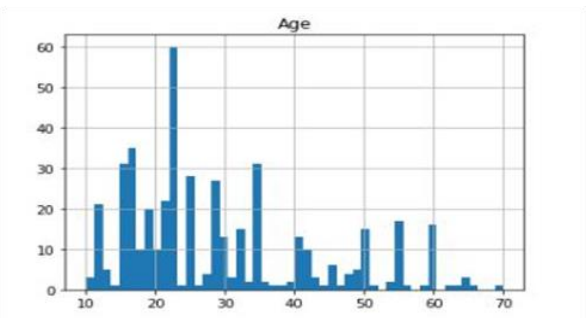

Figure 4. Graph of Age

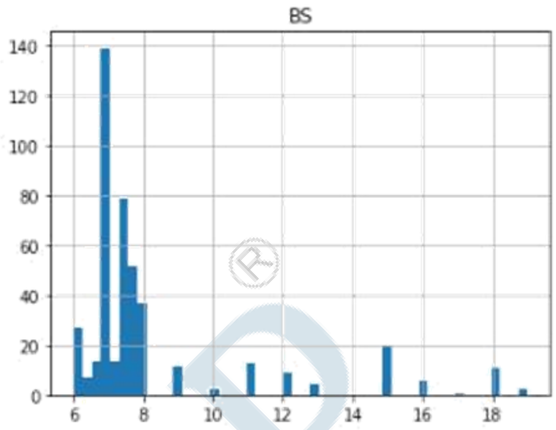

Figure 5. Graph of Systolic BS

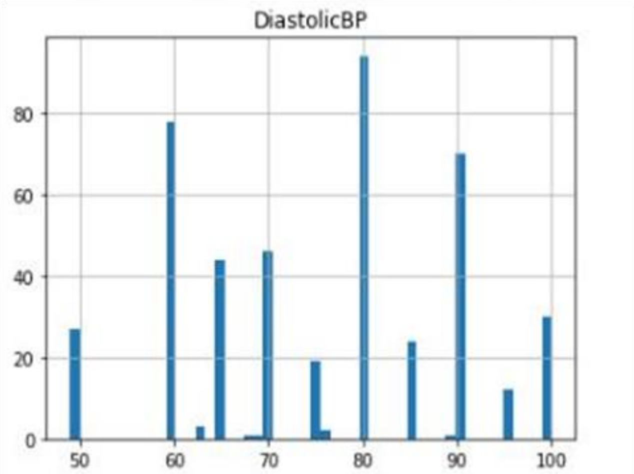

Figure 6. Graph of Diastolic BP

Using data preparation code then finding out the frequency. The quantity of waves that pass a set location in a predeter- mined period of time is known as frequency. Therefore, if a wave passes in 1/2 of a second, the frequency is 2 per second. The frequency is 100 times per hour if it takes 1/100 of an hour. In **Figure 4** showed the graph for frequency and age in years when a woman is pregnant. **Figure 5** showed the graphical view of frequency and Blood Sugar. Frequency and Systolic Blood Pressure are used to create this graph. The graph in **Figure 6** showed diastolic blood pressure. Here, we use frequency and BP to create this graph. The SystolicBP graph in **Figure 7** looks like this. Create the following graph using SystolicBP and frequency. **Figure 8** shows the body temperature of paternity period. Here, we use temperature and frequency to create this graph. Finding the frequency after using data preparation code. **Figure 9** shows the graph of heart rate of maternity period. Here, we use frequency and heart rate to create this graph. **Figure 10** is the graph of risk level of maternal health. The following graph based on frequency and risk level. Here, we use frequency and SystolicBP to create this graph.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 9, September 2023

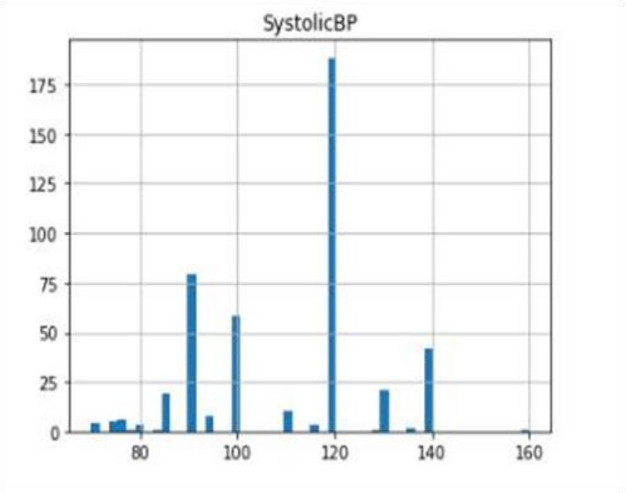

Figure 7. Graph of Systolic BP

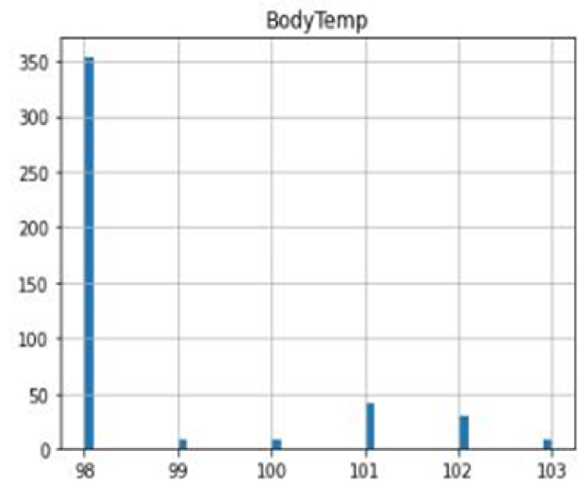

Figure 8. Graph of Body Temperature

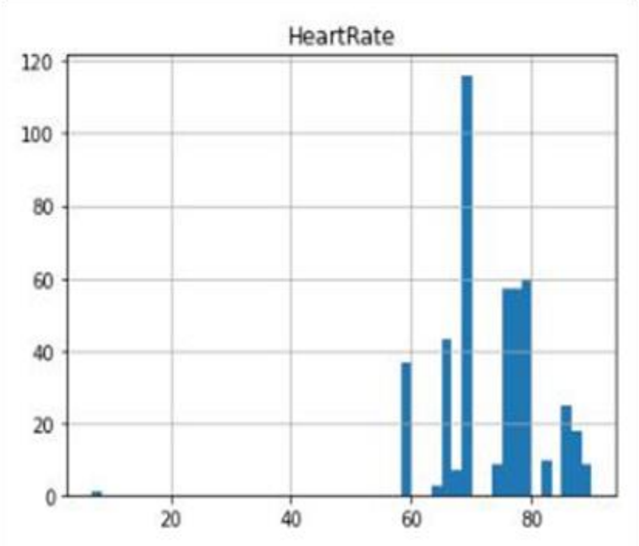

Figure 9. Graph of Heart Rate

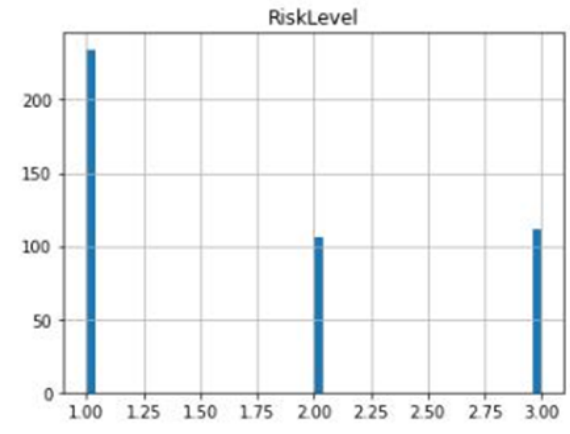

Figure 10. Graph of Risk Level

VIII. RESULT ANALYSIS

Accuracy data table have five algorithms. These are Xg-boost, Decision Tree, Random Forest and Support Vector Machine, and Naiıve Bayes Algorithms.

Table I: Dataset Accuracy

Selected Algorithm	**Data set accuracy rate**			
	Accuracy	***Precision***	***Recall***	***F1Score***
Xgboost	***94***	***0.67***	***0.83***	***0.74***
Decision Tree	***94***	***0.71***	***0.72***	***0.72***
Random Forest	***94***	***0.68***	***0.83***	***0.74***
SVM	***72***	***0.63***	***0.95***	***0.76***
Na¨ıve Bayes	***64***	***0.60***	***0.97***	***0.74***

A. Accuracy

Three algorithms has 94% accuracy, these are Xgboost, Decision Tree, and Random Forest. Support Vector Machine Algorithm has 72% accuracy and Naıve Bayes has 64% accuracy.

B. Precision

Xgboost Precision is 0.67, Decision Tree Precision is 0.71, Random Forest Precision is 0.68, Na¨ıve Bayes Precision is

0.60 and Support Vector Machine Algorithms Precision is 0.60.

$$Precision_{lr} = TP_{lr} / (TP_{lr} + FP_{lr}) \quad (1)$$

C. Recall

Xgboost Recall is 0.83, Decision Tree Recall is 0.72, Random Forest Recall is 0.83, Naıve Bayes Recall is 0.97 and Support Vector Machine Algorithms Recall is 0.95.

$$Recall_{lr} = TP_{lr} / (TP_{lr} + FN_{lr}) \quad (2)$$

D. F1Score

Xgboost F1Score is 0.74, Decision Tree F1Score is 0.72,Random Forest F1Scor e is 0.74, Na¨ıve Bayes F1Score is 0.74 and Support Vector Machine Algorithms F1Score is 0. 76.

$$F1Score_{lr} = 2 * Precision_{lr} * Recall_{lr} / (Precision_{lr} + Recall_{lr}) \quad (3)$$

Table II: Comparative Result

Author	***Algorithms***	***Accuracy***
Kourou et.al [32]	*ANN, BN, SVM, and DT*	*ANN best, 89%*
Ashima Singh	*BN, ANN, SVM, DT and RF*	*BN best,74%*
Naveen	*BN and SVM*	*Both are 72%*
Babita Rani	*KNN, NN*	*Accuracy 68%*
Radhanath Patra	*RBF*	*Accuracy 66%*
Muhammad Mubashir	*DNNs*	*Accuracy 88%*
Yasemin Gultepe	*RF,NB,LR*	*Best,87%*
Bijaya Kumar Hatuwal	*CNN*	*accuracy 86%*
Xing Chen	*SL*	*75% accuracy*
Sara Irfan	*k-means and DT*	*k-means,69%*

The table above shows the difference of author's papers. This paper is better than all these papers. This research achieved 94 percent accuracy using the algorithms (Xgboost, Random Forest and Decision Tree). There is no paper with a 94 percent accuracy rate.

IX. CONCLUSION

Maternal Health Risk Analysis Using EDA and Machine Learning is the subject of the research study. Five methods are used to forecast maternal health risk analysis. 1015 data points were used in this analysis. For the assessment of maternal health risk, the author employed the Xgboost, Decision Tree, Random Forest, Support Vector Machine, and Naive Bayes algorithms. Three algorithms, Xgboost, Decision Tree, and Random Forest Algorithm 'have an accuracy rate of 94 percent. SVM accuracy is 72 percent, whereas Naive Bayes accuracy is 64 percent. This unquestionably demonstrates how much better that work is than others. It's also going to be useful for analyzing the risk to the mother's health. I t saves the patients money and time. Doctors and patients both can use this and also get benefits from this.

Limitations:

The acceptability of this paper will enhance if use additional data. Here worked with only five algorithms. There are many more algorithms that have not worked yet. It is possible to learn more if work with these algorithms.

Future Work:

This work will be helpful to people. The method used for this work is machine learning. People can therefore readily determine whether a risk is high or low. Here, just 1015 data are used. Future work on it will incorporate more data. Here are just five of the employed algorithms. There are other additional algorithms that go unutilized. More algorithms will be added in the future.

REFERENCES

[1] Ranadip Chowdhury, Bireshwar Sinha, Mari Jeeva Sankar, Sunita Taneja, Nita Bhandari, Nigel Rollins, Rajiv Bahl, and Jose Martines. Breastfeed- ing and maternal health outcomes: a systematic review and meta-analysis. Acta paediatrica, 104:96–113, 2015.

[2] Judith Yargawa and Jo Leonardi-Bee. Male involvement and maternal health outcomes: systematic review and meta-analysis. J Epidemiol Community Health, 69(6):604–612, 2015.

[3] Theresa O Scholl, Mary L Hediger, and Daniel H Belsky. Prenatal care and maternal health during adolescent pregnancy: a review and meta- analysis. Journal of Adolescent Health, 15(6):444–456, 1994.

[4] Tej Ram Jat, Nawi Ng, and Miguel San Sebastian. Factors affecting the use of maternal health services in madhya pradesh state of india: a multilevel analysis. International journal for equity in health, 10(1):1–11, 2011.

[5] Alberto De Ramo´n Ferna´ndez MTeresa Prieto-Sa´nchez, Daniel Ruiz Ferna´ndez. A decision support system for predicting the treatment of ectopic pregnancies. International Journal of Medical Informatics, 129:198–204, 2019.

[6] Alma Fredriksson, Isabel R Fulcher, Allyson L Russell, Tracey Li, Yi- Ting Tsai, Samira S Seif, Rose N Mpembeni, and Bethany Hedt-Gauthier. Machine learning for maternal health: Predicting delivery location in a community health worker program in zanzibar. Frontiers in Digital Health, 4:855236, 2022.

[7] VN Rachana, CM Madhumitha, S Vanitha Mani, and Naveen Kumar HN. A comprehensive study on health monitoring systems for pregnant women and fetus.

[8] BN Lakshmi, TS Indumathi, and Nandini Ravi. A novel health monitoring approach for pregnant women. In 2015 International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), pages 324–328. IEEE, 2015.

[9] Saima Gulzar Ahmad, Tassawar Iqbal, Anam Javaid, Ehsan Ullah Munir, Nasira Kirn, Sana Ullah Jan, and Naeem Ramzan. Sensing and artificial intelligent maternal-infant health care systems: A review. Sensors, 22(12):4362, 2022.