JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Pregnancy Risk and Fetal Health Prediction using Machine Learning

[1]Dr.Harish B G, [2]Chetan Kumar G S , [3]Sushma N , [4]Sabeeha Banu N R ,
[5]Priyanka N B , [6]Shreyanka M N

[1] Professor, Department of Master of Computer Applications, UBDTCE, Davangere.
[2]Lecturer, Department of Master of Computer Applications, UBDTCE, Davangere.
[3,4,5,6] Student, Department of Master of Computer Applications, UBDTCE, Davangere.

University BDT College Of Engineering, Davangere-577004, Karnataka, India

Abstract : Developing a predictive system to assess the risk level ofpregnant women and classify fetal health can greatly contribute to the overall well-being of both the mother and the baby. The objective of this research is to utilize different machine learning classification models such as Naïve Bayes, K-Nearest Neighbor, Decision tree, and Random Forest to forecast potential complications during pregnancy, with the ultimate goal of reducing maternal mortality rates. The maternal health dataset was open access from the Kaggle website serves as a fundamental building block and various attributes were collected from expectant women. Datasets are analyzed then trained, and build the model. This model is hosted on a Flask web server, and users can predict the result using an easy-to-use GUI. A comparison of many machine learning classification algorithms reveals that the Decision Tree Algorithm has more accuracy in terms of fetus health classification and maternal risk prediction, with a numerical value of 93%.

Keywords **- Fetal, Fetal Health, Maternal Risk, Kaggle, Machine Learning, Flask Web Server.**

I. INTRODUCTION

The health of the fetus and the mother during pregnancy is a crucial concern for healthcare professionals. Pregnancy complications may be caused by the conditions of women have before pregnancy or develop during pregnancy. Estimating the effect of pregnancy complications on the outcomes for both mothers and neonates is challenging due to the wide spectrum of conditions involved, each possessingvarying degrees of severity. Around 99% of maternal fatalities are predicted to occur in underdeveloped, poornations annually, averaging 358,000. The implementation ofearly screening is anticipated to lead to a decrease in maternalmortality rates.

Leveraging machine learning algorithms for monitoring fetal health and prediction of maternal risk holds promise for enhancing prenatal care and mitigating negative outcomes during pregnancy. This project seeks to investigate the application of machine learning techniques in monitoring fetal health and predicting maternal risk throughout pregnancy. The primary objective is to analyze the risk factors associated with the risk level within the dataset while also identifying well-performing models capable of accurately predicting maternal complications.

The project involved collecting data from various websites. The data included information on maternal health, fetal movements, heartrate, and other vital signs. The data were pre-processed to remove any inconsistencies and outliers. The pre-processed data were then used to train various machine learning classification algorithms.

Employing machine learning algorithms for monitoring fetal health and predicting maternal risk offers numerous benefits. It enables healthcare professionals to make precise predictions and detect potential risks in the early stages. This facilitates timely interventions, ultimately improvingoutcomes for both the mother and the fetus. Nonetheless, there are several challenges associated with this approach. These include the requirement for extensive data sets, the complexity of the models, and the potential for false positivesand false negatives.

The project demonstrates the potential of machine learning algorithms for fetal health monitoring and maternal risk prediction. The trained models were able to health and maternal risk, highlighting the potential for this approach to improve prenatal care. We assess the accuracy of various machine learning models are random forest, decision tree, k-nearest-neighbor, and naïve Bayes. Calculations lead us to the conclusion that the Decision tree has the highest accuracy among them.

II. LITERATURE SURVEY

The primary objective is to investigate the risk factors associated with preterm birth by utilizing various algorithms and classification models.

[1] To predict the risk level, data mining techniques employing both neural network and decision tree algorithms were employed.The findings reveal that factors such as multiple births and hemorrhage during pregnancy are identified as the top two risk factors.

In recent decades, the World Health Organization (WHO) and Pan-American Health Organization (PAHO) have made efforts to combat mortality and Severe Maternal Morbidity. To address this issue, a comprehensive action plan was proposed for the years 2012-2017. This plan primarily focuses on enhancing information systems and monitoring.

maternal health in countries across the region. Reducingmaternal mortality has become a key national objective, withinitiatives including epidemiological surveillance, statistical data analysis, and the identification of risk factors aimed at achieving this goal and bringing about a decrease in maternal mortality rates

In the study conducted by Nanda et al. [3], a model isproposed for predicting gestational diabetes mellitus during the first trimester of pregnancy. The model incorporates biomarkers and selected maternal features. By employing logistic regression, the authors achieved a 74.1% accuracy intheir predictions, with a false positive rate of 20%.

In the study conducted by Farran et al. [4], logistic regression, KNN (k-nearest neighbors), multifactor dimensionality reduction, and SVM (support vector machines) were utilized for predicting diabetes. These techniques were implemented satisfactorily, yielding comparable results.

Galih Malela Damaraji and Adhistya Erna Permanasari [5] published in 2020. In this, it can be concluded that premature birth and hypertension threats are the most observed factors.

It's worth noting that the field of predicting risk levels in pregnancy and fetal health classification is constantly evolving, with researchers continually exploring new techniques and approaches to improve accuracy and reliability. Therefore, it's important to stay updated with the latest research in this field for the most current and acurate information.

III.PROPOSED METHODOLOGY

The presented paper introduces a maternal healthcare model designed to monitor the well-being of pregnant women and fetal health. The model proposed consists of four interconnected processes:

(1)The health data of the patients are obtained from the Kaggle website, ensuring a continuous stream of information.

(2)The collected data is then stored in both a local server anda cloud server, ensuring redundancy and accessibility.

(3)The stored data undergo classification and prediction using a machine learning algorithm, leveraging the knowledge acquired from trained data.

(4)The predicted results are subsequently transmitted to boththe source, including hospitals and emergency services, and the destination, which refers to the patients themselves, facilitating timely actions and interventions.

The proposed methodology outlines the process of merging data from diverse sources to create a consolidated medical dataset. This merged dataset is then utilized to prepare for analysis and prediction using machine learning algorithms. Incases where the prediction accuracy is unsatisfactory, hyperparameter tuning is performed on machine learning classifiers. The accuracy of different models is compared, and the best-performing prediction model is chosen for subsequent analysis and implementation.

1.Data Collection

Initially, we collect datasets from the Kaggle website for maternal risk prediction and fetal classification system. The dataset consists of 6 features and atotal of 1014 patient records which are used for predicting maternal risk and the dataset consists of 22 features and a total of 2126 patient records which are used for fetal classification.

We utilize a dataset comprising individuals who have conducted analyses and tests for identifying maternal risks and classifying fetuses. In the dataset, patients are represented by rows, while factors or attributes (features) under examination are represented by columns, forming a matrix structure.

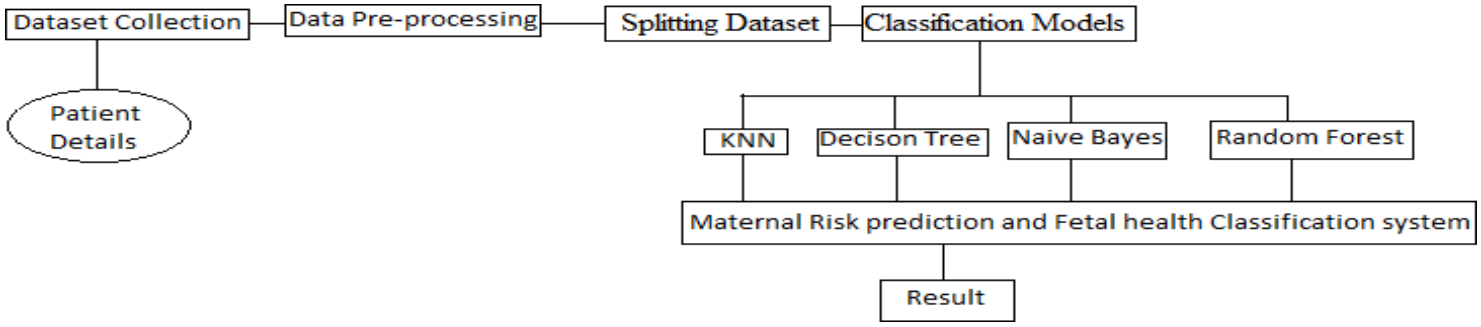


Fig 1: Architecture of prediction system

2.Pre-processing of data

Data pre-processing plays a vital role in the development of a machine learning model. Initially, the data may be in an untidy or incompatible format, leading to potentially erroneous outcomes. During the data pre-processing phase, we convert the data into the necessary format. Ensuring the dataset's reliability and accuracy, this step becomes crucial in addressing factors like noise, duplicates, and missing values present in the data.

Data pre-processing encompasses a range of activities that include collecting datasets, partitioning datasets, data cleaning, factorization, missing values imputation, etc. The purpose of data pre-processing is to enhance the accuracy of the model by refining the data before it is used for analysis or modeling.

3.Splitting Dataset

Once the dataset is collected, we proceed to partition it into training data and testing data. The effectiveness of the dataset classification is influenced by the quality of the training and testing phases. To achieve more accurate results, we divide the entire dataset into two parts: the majority portion (80%) is allocated for training purposes.while the remaining portion (20%) is reserved for testing the trained model.

4.CLASSIFICATION MODELS

The datasets are clustered based on the variables and criteria of Decision Tree (DT), KNN, Naive Bayes, and Random Forest regressor features. Subsequently, the classifiers are utilized on each clustered dataset to evaluate their performance and accuracy.

A. K-NEAREST NEIGHBOR (KNN)

K-nearest neighbor classifier (KNN) is a non-parametric supervised machine learning algorithm that operates on the principle of proximity. It classifies objects by considering the classes of their nearest neighbors. While commonly used for classification tasks, KNN can also be applied to regression problems.

The prediction for a new point is determined by its similarity to the points in the training set. By averaging the observations within the neighborhood, KNN effectively approximates the relationship between independent variables and the continuous outcome, providing an intuitive estimation.

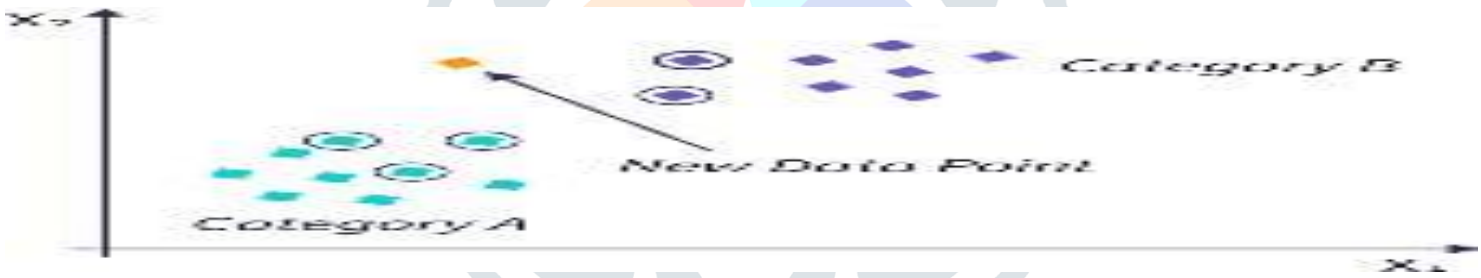


Fig 2: KNN Classifier

B. Decision Tree

A decision tree is a versatile supervised learning algorithm employed for classification and regression tasks. It derives its name from its visual representation, resembling a tree structure with branching pathways. By recursively dividing the data into subsets based on various feature values, it is a predictive model that learns from input features to make decisions or predictions. Each split creates a node in the tree, with branches representing the possible outcomes.

The decision tree classifier uses a series of if-else conditions to traverse the tree and classify new instances based on the learned patterns. It is known for its interpretability, as the decision-making process can be easily visualized and understood.

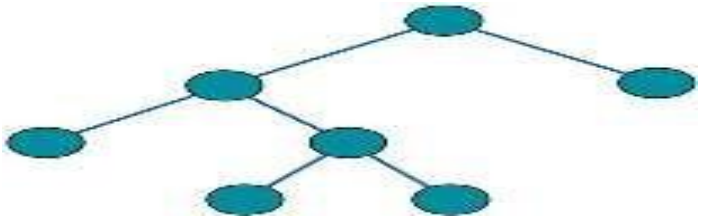


Fig 3: Decision Tree Classifier

C. Naïve Bayes

Naive Bayes is a widely used classification algorithm employed for classifying tasks. It operates based on Bayes' theorem to calculate the probability of a specific event happening, given prior knowledge. The "naive" aspect of Naive Bayes assumes that all features are independent of oneanother, simplifying the computational process.

D. RANDOM FOREST

A Random Forest Regressor is used for regression tasks. It employs an ensemble approach, where numerous decision trees are combined to generate predictions. In the Random Forest Regressor, an ensemble of decision trees is constructed. Each tree within the Random Forest Regressor is trained using a randomized subset of the training data and a randomized subset of input features. During prediction, theoutput of each individual tree is averaged to obtain the final prediction. It improves upon the weaknesses of a singledecision tree by reducing overfitting and improving generalization.

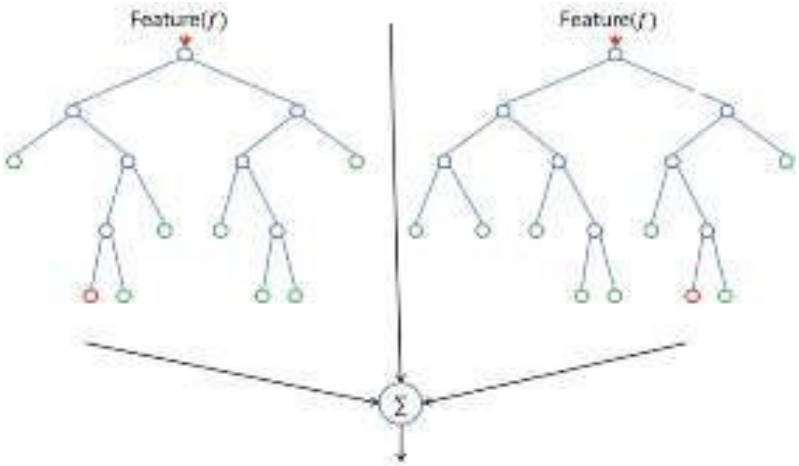


Fig 4: Random Forest Regressor

IV. SYSTEM VIEW

The proposed working model stands out as advantageous and time-efficient in comparison to other systems. It effectively reduces treatment costs by offering timely initial diagnosis. One can detect the maternal risks involved and fetus classification using our computer-aided system or website or application. After pre-processing the data, various classification techniques will be employed to predict the accuracy. The accuracy measure will be used to compare the performance of different classifiers. By utilizingthe proposed system, women can conveniently monitor their condition on a daily basis without incurring substantial costs in terms of both money and time.

V. RESULT AND DISCUSSION

After receiving input data from the system, the ML algorithm will be able to identify the statistics and deliver thefirst output in the design of various intermediates to locate the most accurate result. We compared the testing data and actualdata to get the accuracy of our project.

A website has been created with the purpose of generating test data and serving as a virtual platform to connect with doctors. The system is capable of efficiently matching specific symptoms with suitable doctors, making it convenient to find medical professionals.

Additionally, patients have the option to provide feedback ontheir experience with the doctors and communicate directly with them, which is particularly crucial during pregnancy.

1.Below fig refers to the home screen of maternal risk and fetal health classification.

2.Fig refers to the data flow for admin page so that patient canlog in with user name and password

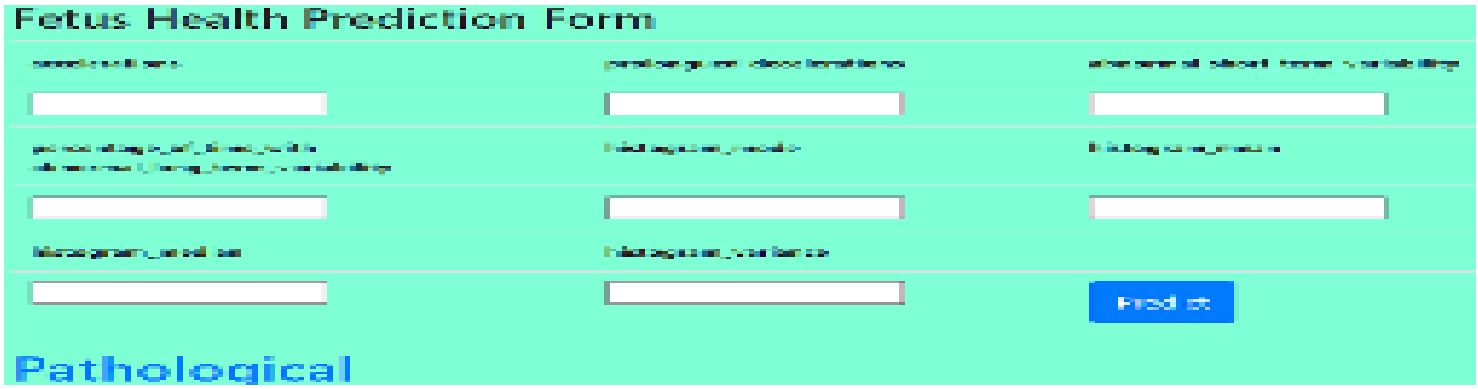
3.Fig refers to the home page of both maternal risk and fetalhealth. So that patient can check any of these which they wish

4.Fig shows the page to input the attributes to predict the typeof risk

5..Fig shows the output of high risk. By the help of attributevalues this will predict patient is having high maternal risk

6.Fig shows the window to input the attributes to predict the fetal health.

7.Fig shows the output of type of fetal health. By the help of attributes values this will predict patient is having Pathological.



Confusion Matrix- A confusion matrix is a performance evaluation tool used in machine learning and predictive modeling to assess the accuracy of a classification algorithm. In the context of predicting the health of a fetus, a confusion matrix can be constructed to analyze the performance of a model that predicts the fetus's health status using specific input features or data.

A confusion matrix for fetal health prediction consists of a square matrix with rows and columns representing the predicted and actual classes or categories of fetal health conditions. Typically, there are three categories or classes used to represent the health status, such as "normal," "suspect," and "pathological." Also, three categories are used to represent a maternal risk, such as “Low risk”, “Mid risk”, and “High risk”.

8.Fig shows the confusion matrix of the decision tree whereit has the most accuracy among all the classification models.

The following results demonstrate the accuracy in correctly identifying maternal risk levels and fetal health for the givendataset of parameters for various classification algorithms. Out of all the machine learning algorithms considered, the Decision Tree algorithm yields the highest accuracy for accurate classification in both the training set data. By employing the decision tree, it becomes

straightforward to determine whether pregnant women are at risk or not, assessthe fetal health, and gauge the severity of the risk.

Upon conducting multiple machine learning models for testing and training, it is observed that the Decision Tree classifier exhibits significantly higher efficiency and accuracy compared to other algorithms.

Below table shows the accuracy of various classification models for the Maternal Risk:

Algorithm	Accuracy
KNN	88
Random Forest	89
Decision Tree	93

Table shows the accuracy of various classification models forthe Fetus heath:

Algorithm	Accuracy
KNN	90
Naïve Bayes	84.5
Decision Tree	93

VI. CONCLUSION AND FUTURE ENHANCEMENT

Machine learning techniques are of immense importance in disease diagnosis, offering substantial benefits. They enable the early prediction of maternal risk and assessment of fetal classification, thereby facilitating the implementation of appropriate treatment procedures for patients. In this project, various classification methods employed in medical diagnosis are investigated, with a specific emphasis on their accuracy. It is intriguing to note that the decision tree classifier achieves the highest accuracy among the models KNN, Random Forest Regressor, Naïve Bayes, and Decision Tree.

To increase the accuracy of these models and address their shortcomings, additional research and development are required. In conclusion, predicting maternal health risks using machine learning models has the potential to significantly enhance maternal health outcomes and lower mortality rates.

REFERENCES

[1] Chen H-Y, Chuang C-H, Yang Y-J, Wu T-P. Exploring the risk factors of preterm birth using data mining. Expert Syst Appl. 2011;38(5):5384–87

[2] Organization, W.H., UNICEF.: Revised 1990 estimates of maternal mortality: a new approach. World Health Organization (2021)

[3] Nanda, S., Savvidou, M., Syngelaki, A., Akolekar, R., Nicolaides, K.H.: Prediction of gestational diabetes mellitus by maternal factors and biomarkers at 11 to 13 weeks. Prenat.Diagn. 31(2), 135–141 (2021)

[4] Farran, B., Channanath, A.M., Behbehani, K., Thanaraj, T.A.: Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity: machine-learning algorithms and validation using national health data from kuwaita cohort study. BMJ Open 3(5), e002457 (2020)

[5] Galih Malela Damaraji and Adhistya Erna Permanasari published in 2020. A Review of Expert System for Identification Various Risk in Pregnancy.

[6] Haaga, J.G., Wasserheit, J.N., Tsui, A.O., et al.: Reproductive Health in Devel- oping Countries: Expanding Dimensions, Building Solutions. National Academies Press, Washington,DC(2021)

[7] Morales-Osorno, B., Martinez, D.M., Cifuentes-Borrero,R.: Extreme maternal morbidity in Clinica Rafael Uribe Uribe, Cali, Colombia, from January 2003 to May 2006. Revista Colombiana de Obstetricia y Ginecolog´ıa 58(3), 184–188 (2019)

Muhammad Nazrul Islam, Sumaiya Nuha Mustafina,Tahasin Mahmud & Nafiz Imtiaz Khan- Machine learning to predict pregnancy outcomes: a systematic review, synthesizing framework and future research agenda.

[8] Patel RR, Murphy DJ. Forceps delivery in modern obstetric practice. BMJ. 2004;328(7451):1302–05.

[9] D. Coustan, R., "Diabetes in pregnancy," Clinical Maternal-Fetal Medicine Online (CMFO), pp.16-1, 2021.

[10] Marino Martinez, C.A., Fiesco, V., Carolina, D., et al.: Caracterizaci´on de la morbilidad materna extrema en el Instituto Materno Infantil-Hospital la Vic- toria/Characterization of extreme morbidity disease in the Instituto Materno Infantil-Hospital la Victoria. Ph.D. thesis, Universidad Nacional de Colombia

[11] Neocleous, C.K., Anastasopoulos, P., Nikolaides, K.H., Schizas, C.N., Neokleous, K.C.: Neural networks to estimatethe risk for preeclampsia occurrence. In: Inter- national JointConference on Neural Networks, IJCNN 2009, pp. 2221– 2225. IEEE (2019)

[12] Carty, D.M., Siwy, J., Brennand, J.E., Zurbig, P., Mullen,W., Franke, J., McCulloch, J.W., North, R.A., Chappell, L.C., Mischak, H., et al.: Urinary pro- teomics for prediction of preeclampsia. Hypertension 57(3), 561–569(2021)

[13] Thomas, Grégoire, Louise C. Kenny, Philip N. Baker, and Robin Tuytten. "A novel method for interrogating receiver operating characteristic curves for assessing prognostic tests."Diagnostic and prognostic research 1, no. 1 (2017): 1-9.

[14] van Doorn, Sander, Timo B. Brakenhoff, Karel GM Moons, Frans H. Rutten, Arno W. Hoes, Rolf HH Groenwold,and Geert Jan Geersing. "The effects of misclassification in routine healthcare databases on the accuracy of prognostic prediction models: a case study of the CHA2DS2-VASc scorein atrial fibrillation." Diag.