Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 9, September : 2023

PREGNANT WOMEN HEALTH RISK PREDICTION USING ENSEMBLE LEARNING

INDIGIBILLI SAHITHI,
M.Tech., Department of IT, Andhra University College of Engineering, Andhra University, Visakhapatnam, India. isahithi999@gmail.com
PROF. KUNJAMNAGESWARA RAO,
Department of IT&CA, Andhra University College of Engineering, Andhra University, Visakhapatnam, India. kunjamnag@gmail.com
DANGETY SOWJANYA,
Research Scholar Department of CSE, Andhra University College of Engineering, Andhra University, Visakhapatnam, India. sowjid91@gmail.com

ABSTRACT

Maternal health is an important aspect of women's health during pregnancy, childbirth, and the postpartum period. Specifically during pregnancy, different health factors like age, blood disorders, heart rate, etc. can lead to pregnancy complications. This study aims predict maternal health risks using health data records. According to WHO, women between the ages of 10 and 19 and those over 35 are considered high-risk pregnancies and Experts say the best time to get pregnant is between your late 20s and early 30s. Using this data we identify pregnancy risk levels, Similarly the risk by each factor BP, body temperature, heart rates, in these conditions the risk is divided which we divide is said to be low middle high. In these three conditions are taking as input. Using this data we implemented a risk prediction system for pregnant women. This can predict the health of the pregnant woman and protect the mother and child. Experiments involve using a dataset of 1014 samples collected from maternal health care, hospitals, and community clinics. The Maternal Health Exploratory Data Analysis (MHEDA) Class imbalance is resolved using the synthetic minority oversampling technique. In this proposed model Supervised Machine Learning Algorithms are DTC (Decision Tree Classifiers), SVM (Support Vector Machines), ETC (Extra Tree Classifiers), And Deep learning approach RCNN (Regions Convolutional Neural Networks) are implemented. It is identified that technique RCNN proposes best result with an accuracy of around 89%.

Key Words: LSTM, BI-LSTM, Machine Learning, Risk Score, DTC (Decision Tree Classifiers), SVM (Support Vector Machines), ETC (Extra Tree Classifiers), RCNN.

1. INTRODUCTION:

Due to limited information about maternal health care during and after pregnancy, pregnant women frequently pass away from difficulties, especially in rural areas and low-middle-class households in developing nations. Women who pass away from any cause connected to or made worse by pregnancy during pregnancy, during childbirth, or within the first 42 days after their pregnancy has been terminated. According to the World Health Organization (WHO), it is a serious public health issue around the world since 830 women die every day from pregnancy- and childbirth-related avoidable causes. At least 40% of all pregnant women will have some sort of pregnancy problem. About 15% of women will experience a potentially fatal condition that needs post-natal treatment. Around 60 million women are in pain. About 60 million women suffer from complications from pregnancy, also known as maternal morbidity. For more than 15 million women these morbidities are long-term and often debilitating. Most maternal deaths have place in low- and middle-income nations, when access to high-quality health care is scarce. The most frequent direct causes of maternal issues include excessive blood

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 9, September : 2023
loss, high blood pressure, infections, unsafe abortions, diabetes, thyroid disease, epilepsy, pregnancy-related blood disorders, poorly controlled asthma, and unsafe abortions. Additionally, consuming illegal substances, drinking alcohol, and smoking cigarettes can all endanger a pregnancy. To save the lives of the more than 500,000 women who die each year from complications during pregnancy and childbirth, maternal health must be improved. Poverty, illiteracy, and inadequate nutrition are risk factors for maternal death. The WHO has set a target to lower the worldwide maternal mortality ratio to fewer than 70 per 100,000 live births by 2030. Reducing maternal mortality is a global health priority. Comprehensive and integrated measures that enhance access to high-quality maternal healthcare services, boost women's education and empowerment, and address social and economic inequality must address the root causes of maternal death.

Direct Causes: Direct causes of maternal mortality are those that are straightway related to gestation or child birth. The most common direct causes of motherly mortality include

- Severe bleeding (postpartum hemorrhage)
- Infections (similar as sepsis and tetanus)
- High blood pressure during gestation (preeclampsia and breakdown)
- Complications from delivery (such as obstructed labor or ruptured uterus)
- Unsafe abortion

Indirect Causes: Indirect causes of maternal mortality are those that aren't directly related to the pregnancy or parturition but are complicated by the physiological effects of pregnancy. Common indirect causes include

- Pre-existing medical conditions
  (such as diabetes, pregnancy disease, HIV/AIDS)
- Malnutrition and anemia
- Inability to obtain good maternal healthcare services

2. LITERATURE SURVEY

Deep learning-based models have been used for a variety of medical applications, including disease prediction, according to Raza A, Siddiqui HUR, Munir K**.** They noted that they have been applied to images, health records, and time-series data. By exposing hidden patterns in medical data, these models help researchers and medical professionals make quick and accurate diagnoses. In view of emerging applications, appropriateness, and efficiency of deep learning models, this work employs deep neural models to identify health concerns related to pregnancy. It also suggested a brand-new deep neural network architecture called DT-BiLTCN for feature extraction, which combines decision trees, Bidirectional Long Shortterm Memory (BiLSTM), and Temporal Convolutional Network (TCN). using n to train machine learning models in the future. Extra tree classifiers (ETC), support vector machines (SVM)**[1]**. Marzia Ahmed and Mohammod Abul Kashem et al. presented a model for tracking the health of the fetus and pregnant women. IoT environments have collected and shared data both locally and to analyze medical data and identify risk levels. The suggested process was followed when combining data from various sources. Weka and Python machine learning algorithms have been integrated to analyze and forecast extra information from a medical data set. The three risk categories that have been considered are low risk, mid risk, and high risk. There were 1014 entries total, of which 406 were deemed to be low risk, 336 were deemed to be mid risk, and 272 were deemed to be high risk. The decision tree offers a maximum accuracy of 97% since groups of machine learning algorithms have been created using both Python and Weka. The classifier was adjusted using Grid Search CV, a hyperparameter tuning method, to get the best values. Additionally, a few statistical and data mining techniques, such the Chisquare test, Info gain, Gain ratio, etc., have been employed to pinpoint important variables. Every woman's life is delicate during pregnancy, and the health of pregnant women fluctuates for a variety of reasons, according lakshmi.B.N, Dr.Indumathi.T.S, and Dr.NandiniRavietal[**2**]. Using two classification algorithms decision tree classification algorithm created a project to anticipate the current health complications inflicting on a pregnant woman. These methods were chosen because they are frequently used to perform the classification and prediction tasks in data mining. The investigation is separated into four phases: data gathering, preprocessing, analysis, and model evaluation. Data collection is the initial phase of the study. The study is divided into four phases, the first being the data collection phase, second the data preprocessing phase, the third data analysis phase and lastly the model evaluation phase. A total of seventeen parameters are considered for the study like, age, present state, date of conceiving, present month of pregnancy, present trimester, pregnancy parity, history of pre-eclampsia, history of gestational diabetes, family history of preeclampsia, family history of gestational diabetes, Blood Pressure (BP), Presence of gestational diabetes, height, previous weight, present weight and weight gain. So, the accuracy of each parameter is crucial for pregnant women's health. The total accuracy obtained is 59%**[3].**

3. IMPLEMENTATION STUDY

Even the pregnancy condition is being highlighted as a quiet killer that kills a woman without showing any overt symptoms. Growing concern about the illness and its effects is a result of the disease's nature. Therefore, efforts to foresee the potential occurrence of this fatal disease in the past continue. As a result, various technologies and procedures are frequently tested to meet the needs of modern health. Machine learning methods can be extremely helpful in this situation. Despite the fact that pregnancy risk can take many distinct forms, there isa common set of fundamental risk variables that ultimately determine whether or not someone is at risk for pregnancy risk. By gathering information from many sources, organizing it into categories that make sense, and then performing analysis.
From Figure 1: Structure of supervised learning Supervised algorithms to give the necessary input and output data as well as to provide feedback on the accuracy of the predictions, which is crucial during algorithm training, they need a data researcher or data analyst with experience in machine learning. The characteristics or features that the model should include while making predictions are chosen by data scientists. After the training is over, the algorithm will use the fresh data to apply what it has learnt. Problems with regression and classification fall under the heading of supervised learning difficulties. There is a classification challenge when the output variable is a category, such as "red" or "blue" or "illness" and "no disease, "Regression: A regression problem arises when the output variable, such as "dollars" or "weight, "has a real value.

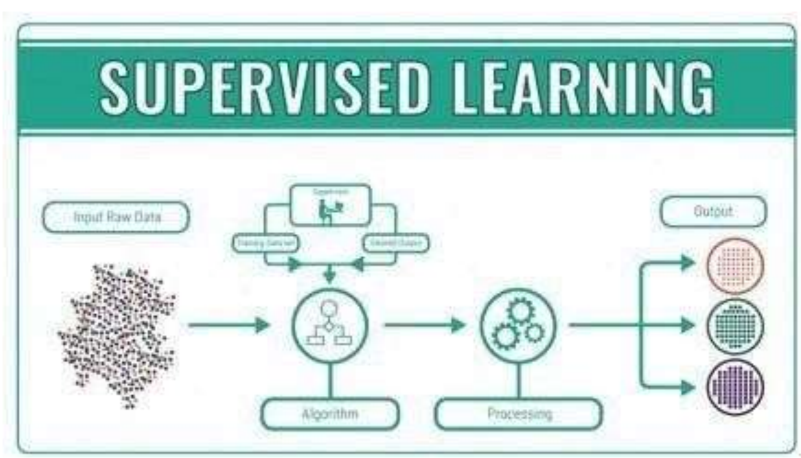

Figure 1: Structure of supervised learning

MHEDA refers to the process of discovering dataset patterns, hypothesis tests, and assumption checks by utilizing the graphical data representations and statistics summary of the dataset. MHEDA helps to summarize the main

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 9, September : 2023

dataset characteristics and features relation analysis. The feature relation analysis and data visualization methods help in the proposed model's pre- diction process. Results of statistical dataset features analysis.

4. METHODOLOGY:

In this project we can used Supervised Machine Learning Algorithms are DTC (Decision Tree Classifiers), SVM (Support Vector Machines), ETC (Extra Tree Classifiers), And Deep learning approach R-CNN (Regions with Convolutional Neural Networks).

DECISION TREE: Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome**.**

SVM (Support Vector Machine): Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

Extremely Randomized Trees Classifier (Extra Trees Classifier): is a type of ensemble learning technique which aggregates the results of multiple de-correlated decision trees collected in a "forest" to output it's classification result. In concept, it is very similar to a Random Forest Classifier and only differs from it in the manner of construction of the decision trees in the forest. Each Decision Tree in the Extra Trees Forest is constructed from the original training sample. Then, at each test node, Each tree is provided with a random sample of k features from the feature-set from which each decision tree must select the best feature to split the data based on some mathematical criteria (typically the Gini Index). This random sample of features leads to the creation of multiple de-correlated decision trees.

R-CNN: Object detection is the process of finding and classifying objects, One deep learning approach, regions with convolutional neural networks (R-CNN), combines rectangular region proposals with convolutional neural network features. R-CNN is a two-stage detection algorithm. The first stage identifies a subset of regions in an image that might contain an object. The second stage classifies the object in each region.

ALGORITHM:

Step1: Start

Step2: Data Preprocessing; In this step2 is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model.

Step3: Data Exploration; In this step3 refers to the initial step in data analysis in which data analysts use data visualization and statistical techniques to describe dataset characterizations, such as size, quantity, and accuracy, in order to better understand the nature of the data.

Step4: Model Creation; In this step3 a machine learning model(Decision Tree, SVM, Extra Tree Classification) or Deep Learning model(R-CNN) used a file that has been trained to recognize certain types of patterns. You train a model over a set of data, providing it an algorithm that it can use to reason over and learn from those data**.**

Step5: Training and Testing Training data is the initial dataset you use to teach a machine learning application to recognize patterns or perform to your criteria, while testing or validation data is used to evaluate your model's accuracy. You'll need a new dataset to validate the model because it already "knows" the training data.

Step6: The algorithms that gives the highest accuracy for finding pregnancy women risk prediction.

Step7: Stop

5. RESULS AND ANALYSIS:

Table-1 is representing the averaged results of applied learning models using the proposed DT-BiLTCN feature extraction technique.

Model	Accuracy%	Precision%	Recall%	F1 score%
SVC	60	63	60	61
ETC	84	86	84	85
DT	81	84	81	82
RCNN	89	97	93	95

Table 1: Shows the accuracy score, precision, recall, f1 score of every algorithm

Table 1 shows accuracy score of Machine learning algorithms SVC is 60% and Precision 63%, Recall 60% and f1 score 61% another Machine learning algorithms ETC is 84% and Precision 86%, Recall 84% and f1 score 85% another Machine learning algorithms DT is 84% and Precision 84% ,Recall 81% and f1 score 82%. And Deep learning approach RCNN gives highest accuracy is 89%, Precision 97% Recall 93% and f1 score 95%. It is identified that technique RCNN proposes best result with an accuracy of around 89%.

6. CONCLUSION AND FUTURE WORK

In this work, detection of Maternal Health Risk was attempted using various machine learning and deep learning techniques. Various performance evaluation metrics were used to analysis the performance of the models implemented for Maternal Health Risk detection on Maternal Health Risk Dataset. When we compare results between machine learning and deep learning algorithms, In those deep learning algorithms are performing well.In this proposed model Supervised Machine Learning Algorithms are DTC (Decision Tree Classifiers), SVM (Support Vector Machines), ETC (Extra Tree Classifiers), And Deep learning approach RCNN (Regions Convolutional Neural Networks) are implemented. It is identified that technique RCNN proposes best result with an accuracy of around 89%.

7. REFERENCES

[1]. Finlayson K, Crossland N, Bonet M, Downe S. What matters to women in the postnatal period: a meta-synthesis of qualitative studies. PloS one. 2020; 15(4):e0231415. https://doi.org/10.1371/journal.pone.0231415 PMID: 32320424

[2]. Turban JL, King D, Kobe J, Reisner SL, Keuroghlian AS. Access to gender-affirming hormones during adolescence and mental health outcomes among transgender adults. PLos one. 2022; 17(1): e0261039. https://doi.org/10.1371/journal.pone.0261039 PMID: 35020719

[3] Mercan Y, Tari Selcuk K. Association between postpartum depression level, social support level and

breastfeeding attitude and breastfeeding self-efficacy in early postpartum women. PloS one. 2021; 16 e0249538. https://doi.org/10.1371/journal.pone.0249538 PMID: 33798229

[4] WHO U. Maternal mortality ratio (modeled estimate, per 100,000 live births) |Data; 2019. Available from: https://data.worldbank.org/indicator/SH.STA.MMRT?end=2017&start=2000&view=chart.

[5] Max Roser and Hannah Ritchie. Maternal Mortality—Our World in Data;. Available from: https:// ourworldindata.org/maternal-mortality.

[6] Mehboob R, Gilani SA, Khalid S, Hassan A, Alwazzan A. Maternal Mortality Ratio in Low Income Developing Countries. In: Global Women's Health. IntechOpen; 2021.

[7] Kaur S, Singla J, Nkenyereye L, Jha S, Prashar D, Joshi GP, et al. Medical Diagnostic Systems Using Artificial Intelligence (AI) Algorithms: Principles and Perspectives. IEEE Access. 2020; 8:228049–228069. https://doi.org/10.1109/ACCESS.2020.3042273

[8] Manifold B, Men S, Hu R, Fu D. A versatile deep learning architecture for classification and label-free prediction of hyperspectral images. Nature machine intelligence. 2021; 3(4):306–315. https://doi.org/ 10.1038/s42256-021-00309-y PMID: 34676358

[9] Zeng M, Li M, Fei Z, Wu FX, Li Y, Pan Y, et al. A Deep Learning Framework for Identifying Essential Proteins by Integrating Multiple Types of Biological Information. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2021; 18(1):296–305. PMID: 30736002

[10] Berrar D, Dubitzky W. Deep learning in bioinformatics and biomedicine. Briefings in Bioinformatics. 2021; 22(2):1513–1514. https://doi.org/10.1093/bib/bbab087 PMID: 33693457

[11] Theis J, Galanter WL, Boyd AD, Darabi H. Improving the In-Hospital Mortality Prediction of Diabetes ICU Patients Using a Process Mining/Deep Learning Architecture. IEEE Journal of Biomedical and Health Informatics. 2022; 26(1):388–399. https://doi.org/10.1109/JBHI.2021.3092969 PMID: 34181560

[12] Wang D, Mo J, Zhou G, Xu L, Liu Y. An efficient mixture of deep and machine learning models for COVID-19 diagnosis in chest X-ray images. PloS one. 2020; 15(11):e0242535. https://doi.org/10.1371/ journal.pone.0242535 PMID: 33201919