

# Automated Pregnancy Risk Level Prediction using Advanced Machine Learning and Deep Learning Algorithm

Proloy Karmakar  
Department of CSE, Khulna  
University of Engineering &  
Technology, Khulna, Bangladesh

Md Sazzad Hossain  
Department of Computer  
Mechanical Engineering,  
Sonargaon University, Dhaka,  
Bangladesh

Riazul Islam & Mehedi Hasan  
Department of CSE, University of  
Global Village, Barisal, Bangladesh

## ABSTRACT

In this study we analyzed different well-established machine learning (ML) and deep learning (DL) supervised models to enable the risk prediction of maternal health, thus offering a viable and systematic technique to automatically identify pregnancy risk. The Maternal Health Risk Data Set, which covers various critical attributes such as age, blood pressure, blood sugar, body temperature, heart rate, and risk level, was applied [8]. Data pretreatment methods, including deleting missing data (if any) and conducting feature scaling and selection, were incorporated to create the model. Different ML models were created and tested, including but not limited to Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting (GB), as well as deep learning architectures such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM). Model performances were evaluated using metric measures, including accuracy and F1 scores. Of them, CNN showed the highest accuracy (98.58), exceeding alternative models by its capacity to uncover spatial correlations crucial for successful risk prediction. CNN shows great accuracy, which indicates that its real-life clinical application in predicting high-risk pregnancies would result in a considerable improvement in maternal care. Adding AI-driven models to existing healthcare settings could assist in the faster and more accurate evaluation of pregnancy risk, particularly in low-resource settings, boosting focused preventative therapy and evidence-based clinical decision-making. The expanding presence of AI has the potential to revolutionize healthcare, taking us closer to scalable automated solutions for maternal health that correspond with global healthcare development goals, with implications from this study.

## General Terms

Machine Learning, Public Health, Women Care

## Keywords

Automated Pregnancy Risk Prediction, Machine Learning, Deep Learning, Risk Level Classification, Healthcare AI.

## 1. INTRODUCTION

It is crucial to be able to forecast pregnancy-related hazards so they can be handled sooner, improving maternal and infant health outcomes. Close monitoring of pregnancy disorders such as preterm birth, gestational diabetes, and hypertension is needed, and if discovered or mistreated, they will lead to serious implications for both the mother and child. More recently, artificial intelligence (AI) and machine learning (ML) have emerged as potential tools for the prediction and management of pregnancy risk by analyzing large-scale datasets to develop synthetic patterns, allowing tailored

assessments. For example, healthcare practitioners are utilizing AI-driven methodologies to improve decision-making for prioritizing health treatments based on specified maternal health markers [3]. These techniques enable early risk detection in prenatal care, allowing healthcare workers [4] to identify and manage risk factors at an early stage. Hence, embedding AI models into our healthcare systems is an intriguing strategy to improve pregnancy risk screening [5].

Traditional pregnancy risk assessment methods base their evaluation on manual analysis and hence have pre-defined assessment mechanisms, resulting in potential inaccuracies as they might not be adjusted to varied patient scenarios [6]. In contexts with more patients or less care, scalability becomes a major difficulty, making it challenging to deploy uniform approaches. Furthermore, these techniques may fail to adequately account for the complex multidimensional aspects of pregnancy risks due to the underassessment of pleiotropic interactions among genetic, environmental, and physiological components [7]. In order to address these constraints, this research provides an automated technique for applying multiple state-of-the-art ML and DL models that may be utilized to determine pregnancy risk levels with more speed and trustworthy results than those previously produced [8]. The goal is to develop a scalable system that could enable healthcare providers to identify women with at-risk pregnancies early on and ultimately enhance maternal and neonatal health outcomes [9]. The major aims of this investigation are as follows:

- To develop a set of ML and DL models capable of predicting pregnancy risk levels based on diverse datasets encompassing maternal health indicators.
- To evaluate and compare the performance of different models to identify the most effective approach for pregnancy risk classification.
- To implement an automated, scalable risk prediction system suitable for various healthcare settings, aiming to support clinicians in real-time risk assessment.

The underlying study effort contributes to the realm of pregnancy risk prediction frameworks, where the current work verifies and also implements manual formulation within a hybrid automated ML-DL-based high-accuracy model. Unlike earlier studies evaluating independent ML approaches on limited data sources or with constrained applications [1]– [9], we leverage a broader range of characteristics and model architectures [10], thereby enhancing prediction accuracy and generalizability in larger populations. The third contribution of

this work is the blend of novel feature engineering and explainability methodologies that help physicians evaluate risk indicators. By integrating this technology into existing healthcare infrastructures, our method also aims to enable access and scalability for best-practice prenatal care in resource-limited settings.

Section 4 covers a literature review on pregnancy risk in general and AI applications in the healthcare arena. Section 5 explains the collection and preparation of such data before applying it to develop a model that predicts pregnancy risk. In Section 6, we present an explanation of our model findings and conduct a comparison study of the outcomes. Section 7 summarizes our conclusions, the limitations of our research, and suggested paths forward. Section 1 sets the groundwork, and Sections 2–7 each describe and explore a paradigm of fetal-maternal recognition as exemplified by instances appropriate to maternal healthcare. Section 8 summarizes our contributions and comments on their significance.

## **2. LITERATURE REVIEW**

As the application of machine learning (ML) and artificial intelligence (AI) technologies in healthcare research has expanded rapidly over the last few years, there are implications for opportunities to enable earlier diagnosis, improve prediction accuracy, and offer targeted preventative strategies for adverse maternal health-related outcomes during pregnancy. High-level overviews of pregnancy risk prediction models, ML in healthcare, and DL approaches have largely focused on medical diagnostics.

However, among the studies that addressed pregnancy risk prediction, relatively few have investigated diverse statistical and machine-learning methodologies. For instance, a prior study compared multivariable logistic regression with ML techniques to examine if ML might provide better prediction of major pregnancy problems and indicated that ML methods are able to boost the identification of risk for management during pregnancy [11]. Other studies have also explored connections between pregnancy problems and the risk of chronic health diseases, such as cardiovascular disease [12], adding further support for long-term forecasting models of maternal health. More recently, ensemble learning algorithms have been employed to examine maternal health throughout pregnancy and suggest that a more complex effort in feature engineering may produce a favorable improvement in prediction [26].

There has been a remarkable surge in machine learning (ML) application papers in healthcare, ranging from opportunities in illness modeling to heuristics for patient treatment. In this regard, the probable repercussions of ML in healthcare have been described as a transformational agent in predictive analytics and treatment planning [16]. Moreover, these ML techniques can manage very large and complex data; consequently, our findings may provide useful information for guiding therapeutic decisions [17]. At a larger level, applications of ML and AI in the field of health could help with better accuracy in diagnosis, rapid access to individualized therapy, and processing of data, all of which are vital for risk prediction during pregnancy [22]. Additionally, due to the prevalence of sensitive applications that can be critical for actions taken in customized medicine employing ML [18], it is more important than ever to protect and verify viable AI paradigms.

Deep learning methods, including CNNs, RNNs, and LSTMs, have been utilized more regularly as alternatives to standard algorithms for medical diagnosis and predictive models for healthcare applications. For example, several fields, such as

healthcare, have greatly benefited from large-scale and high-dimensional medical data [19], in which big data and DL approaches may meet clinical needs, including higher efficiency in diagnosis accuracy and improved prediction ability. This shows the promise of DL in various topics, as convolutional LSTM approaches have also been employed to identify complications associated with pregnancy (e.g., maternal health problem: diabetes) [30], further reinforcing the versatility and impact that high-level predictions within an important domain can yield. DL has been beneficial for addressing some of the challenges in biomedical data [29] and significantly contributes to boosting such models to sometimes surpass human diagnostic capabilities [30, 31], because DL approaches provide all the high-dimensional time series, from low-level manifolds to higher-dimensional clinical information, required for body news associated with scanned health markers that may be used for potential risk prediction during pregnancy [32]. Recent research on the incorporation of DL for continuous monitoring of maternal and fetal health characteristics shows the potential properties of DL in ensuring continuity and accuracy when substantial health data is recorded [32].

While both ML and DL have proven helpful in prediction for maternal health, pregnancy risk prediction models that are interpretable across populations, as well as scalable to real-time applications, are not evident from the literature. A lot of the earlier work did not fully meet the need for a hybrid approach [21], i.e., a combination of ML and DL approaches to boost accuracy while increasing transparency in terms of presenting results to families. Although related concerns may differ, the research gap is quite comparable. In this research, we will address these gaps by developing a hybrid predictive model that synthesizes state-of-the-art methodologies of both classical machine learning (ML) and more effective deep learning (DL) methods. More directed prediction involving pregnant individuals would benefit from the combined inclusion of pregnancy-specific health markers and connections with existing data sources; together, they might widen the reach of our model to diverse clinical situations. This is referenced in part by constraints in existing models/algorithms and the growing need for an evidence-based, all-in-one pregnancy risk prediction tool.

## **3. METHODOLOGY**

### **3.1 Overview of the Proposed Framework and Methodology**

Please use a 9-point Times Roman font, or other Roman font with serifs, as close as possible in appearance to Times Roman in which these guidelines have been set. The goal is to have a 9-point text, as you see here. Please use sans-serif or non-proportional fonts only for special purposes, such as distinguishing source code text. If Times Roman is not available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.

We built our predictive approach to increase the accuracy, interpretability, and clinical value of pregnancy risk estimates. Our suggested framework is a fusion of current classical ML models, which are based on handmade features together with DL architectures to capture the benefits from both worlds. This hybrid model employs ML approaches to identify baseline trends in torso pregnancy markers and DL components for assessing complicated patterns of deep variables utilizing large-scale medical body data. The technique requires a number of preprocessing processes on the gathered data, especially for

dealing with missing values, normalizing characteristics, and discretizing demographic and clinical factors connected with

pregnancy. The outcome is a clean and standardized dataset, which allows for superior model performance. Then, feature

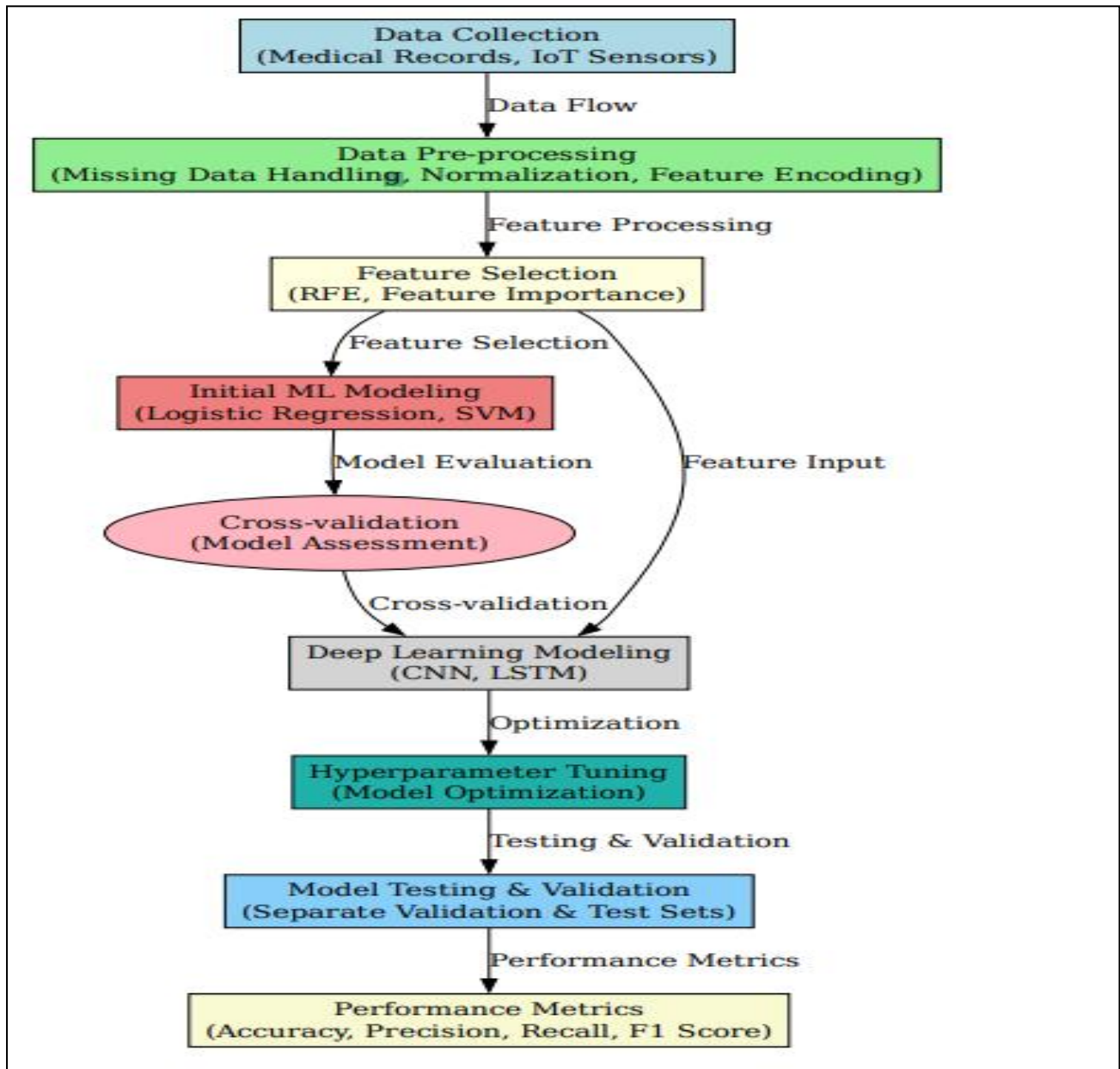


Fig. 1. Flowchart of the Proposed Framework for Predictive Healthcare in Pregnancy Risk:

selection techniques like recursive feature elimination (RFE) are employed to maintain the most significant variables, making the model more efficient and saving computing time.

The technique follows the process of sequential modeling, where ML-based algorithms like logistic regression and support vector machines (SVM) are employed in order to gain a feeling of trend so that we will have some final forecast. Then we apply DL approaches (CNN and LSTM) to boost prediction accuracy by collecting more meaningful, non-linear feature interactions. In each phase, we are running cross-validation to assess the stability of the model; additionally, hyperparameters tweaking is conducted to figure out optimal performances for every algorithm. Model validation and testing to check the generalizability of the model on multiple datasets. We analyze the prediction precision, recall, and F1 score. Our suggested hybrid framework finds linear and non-linear patterns from

pregnancy risk variables in one model, leading to a more customizable strategy for predictive healthcare in the maternal care population process in figure 1.

The diagram illustrates the step-by-step process, from data collection and preprocessing to machine learning (ML) and deep learning (DL) modeling, followed by model testing, validation, and performance evaluation. Each stage is designed to optimize predictive accuracy and provide a comprehensive analysis for assessing pregnancy risks.

### 3.2 Data collection and preprocessing

The data used in this study was collected from various sources including hospitals, community clinics, and maternal healthcare centers, particularly from rural areas in Bangladesh. An IoT-based risk monitoring system was utilized to capture relevant health parameters and maternal risks in real-time. The

data sources primarily involve electronic health records (EHRs) and sensor data gathered via IoT devices.

The dataset comprises several key features that are essential for predicting maternal health risks. These attributes include demographic and physiological factors such as age, blood pressure, blood sugar levels, body temperature, and heart rate, all of which are closely linked to pregnancy-related risks. The target variable in the dataset is the risk level, which indicates the intensity of the maternal health risk during pregnancy.

Table 1 shown below Dataset information

Table 1 Dataset information

Variable Name	Type	Description
Age	Integer	Age of the pregnant woman
SystolicBP	Integer	Upper value of blood pressure during pregnancy
DiastolicBP	Integer	Lower value of blood pressure during pregnancy
BS	Integer	Blood glucose level (molar concentration)
BodyTemp	Integer	Body temperature of the pregnant woman
HeartRate	Integer	Normal resting heart rate
RiskLevel	Categorical	Predicted risk level based on the above attributes

### 3.2.1 Handling Missing Values

The dataset does not contain any missing values, as confirmed during the data collection process. Therefore, no imputation techniques or exclusions were required.

### 3.2.2 Feature Scaling:

Since all features in the dataset are on different scales (e.g., age, blood pressure, body temperature), feature scaling techniques, such as normalization or standardization, may be applied before model training to ensure that the model does not prioritize features with larger numeric ranges.

### 3.2.3 Feature Selection:

Feature selection was performed to identify the most significant variables affecting maternal health risks. Methods like Recursive Feature Elimination (RFE) and feature importance rankings were used to ensure that only the most relevant attributes were retained for the modeling phase. Additionally, no dimensionality reduction techniques, such as PCA, were used, as the dataset contains only six features.

## 3.3 Machine Learning Models

Various ML models are employed in this work to predict maternal health risk based on the attributes of the dataset. These models were selected for their strong performance in classification jobs and their capacity of processing complicated high-dimensional input. The following are the models utilized Fig.2 shows describe Proposed Predictive Healthcare System for Pregnancy Risk Assessment.

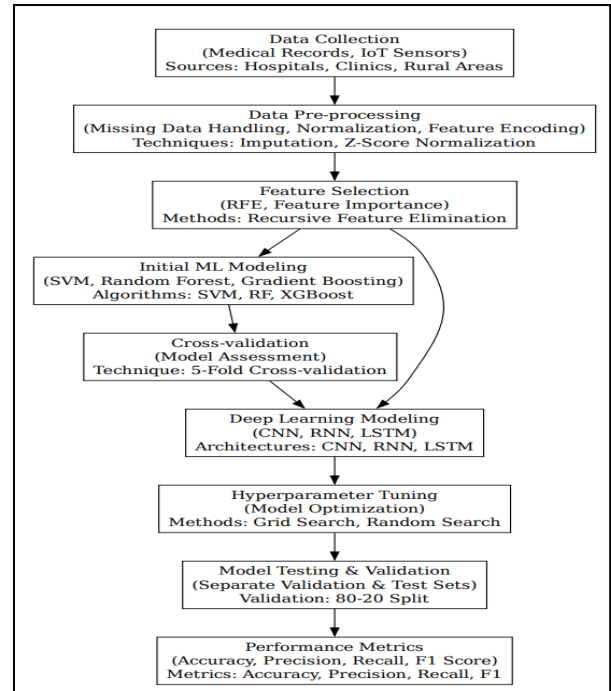


Fig. 2 Flowchart of the Proposed Predictive Healthcare System for Pregnancy Risk Assessment.

This flowchart illustrates the sequential steps of the proposed framework for predicting maternal health risks using machine learning (ML) and deep learning (DL) models. The process begins with data collection from medical records and IoT sensors, followed by data preprocessing, feature selection, and model training. The system integrates both traditional ML models (Support Vector Machine, Random Forest, and Gradient Boosting) and advanced DL models (Convolutional Neural Network, Recurrent Neural Network, Long Short-Term Memory) to optimize risk predictions. The final steps include model testing, validation, and performance evaluation based on key metrics such as accuracy, precision, recall, and F1 score.

### 3.3.1 Support Vector Machine (SVM)

The Support Vector Machine (SVM) is the most prevalent supervised learning technique to be utilized in classification problem formulations. It achieves this by looking for a hyper plane that optimally divides the classes in high-dimensional feature space. Linear manager can easily cope with non-linear connection; it employs kernel processes, for instance, radial basis functionality (RBF) to generate the input space. In our example, SVM is utilized to categorize maternal health data into risk classes (low, medium, and high) based on the six characteristics present in this dataset. Hyperparameters tuning is conducted for acceptable performance and involves fine-tuning of parameters like the regularization parameter (C) or kernel type.

### 3.3.2 Random Forest (RF)

Random Forest (RF): An ensemble learning approach that constructs a forest of decision trees during training and predicts the class based on the mode of classes from numerous individual trees. This is a model with good accuracy, which also makes it a perfect option for complicated and multivariate datasets such as maternal health monitoring since it is resistant to overfitting. RF also delivers an assessment of feature significance, which may be beneficial in determining the major qualities (blood pressure, age, heart rate, etc.) that may serve as a key for forecasting maternal health danger. As RF is an

ensemble, it makes it more generalizable and capable of coping with noisy data.

### **3.3.3 Gradient Boosting (GB)**

Gradient Boosting (GB) is a comparable but stronger ensemble approach that creates trees one step at a time, where each tree seeks to repair the mistakes of the preceding tree. It seeks to optimize the prediction performance over hard-to-predict instances. Specifically, the XGBoost version of gradient boosting is extremely popular for its performance and quickness. GB models are fit to the dataset with a focus on reducing the loss function and therefore may assist in capturing very complicated correlations between the input characteristics (cardiac risk factors) and prediction of risk level. Hyperparameters adjustment on regularization parameters and learning rate may also lead to a large performance increase for GB models.

Neural networks, one of the most often used deep learning (DL) models, are used to fit exceedingly non-linear and very complicated patterns. Such models are effective in cases where links between characteristics may not be readily evident and learn to generate hierarchical representations from raw data. The deep learning models covered in this research include

### **3.3.4 CNN (Convolutional Neural Network)**

Normally, a convolutional neural network (CNN) is used for image processing; however, in this study, CNN has been developed for sequential and structured data categorization. CNNs also contain layers like convolutional layers, pooling layers, fully connected layers, etc. This is because the convolutional layers may capture local patterns (which in maternal health may have to do with temporal or geographical relationships between, e.g., blood pressure and heart rate). The pooling layers are responsible for minimizing the dimensionality and minimizing overfitting, while the fully connected layer(s) produce the categorization (risk sk level). Despite the fact that CNNs are used most commonly in processing image data, they have also been effective in learning to discern patterns from jump to tabular data [1].

### **3.3.5 Recurrent Neural Network (RNN)**

Recurrent Neural Network (RNN) is a form of neural network whose design is ideal for sequential data and time series predictions. RNNs: In the context of maternal health risk prediction, if the data is recorded in time, then it may capture temporal dependencies in the dataset (e.g., blood pressure and heart rate are monitored constantly). RNN design helps to simulate the sequential connection between health factors and how they impact the risk level. If there are any important time-based patterns or events that influence maternal health risk and that it is preferably a recurrent process, RNN models can be applied to gain insights into how parameters and processes evolve during pregnancy over time and how such variations

affect maternal health through time. We therefore suggest this type of model for this dataset.

### **3.3.6 Long Short-Term Memory (LSTM)**

LSTM is a variation of RNN especially developed to minimize the vanishing gradient problem experienced by normal RNNs. Long Short-term Memory Networks (LSTMs) are a family of recurrent neural networks that excel at learning long-term relationships in data, making them highly strong for time-series data with complicated temporal correlations. Approach 2: LSTM model to forecast health risks for women based on patterns learned in physiological data, e.g., blood pressure, heart rate, and blood sugar throughout the course of e of time. LSTMs may recall information for lengthy periods of time, which makes it (in principle) feasible to create more exact predictions by taking into account their past. This aspect of LSTM makes it suited for datasets that have the requirement to comprehend long-term trends in association with maternal health risk factors.

## **3.4 Model Training and Testing**

The dataset is then separated into training and testing sets in this research to assess how well the prediction models perform. We employ a typical training-testing split of 80% for training and de-20% for testing, which enables the model to be trained over enough data while leaving some data independent from training data to assess its performance. Train-test split is another frequent approach where the dataset is partitioned into two distinct chunks, one to fit the model and tweak internal parameters around the model, while the other aids in assessing the capacity of generalization without overfitting the training set.

Hyperparameters tuning is accomplished using two popular approaches, grid search and random search, in order to enhance model performance. In grid search, you indicate how many potential values there exist for each hyperparameters, and grid search will try exhaustively all the hyperparameters combinations in parameter space. Conversely, the random search samples randomly on the grid are therefore computationally less costly while still producing excellent results. Both techniques will be utilized to determine the hyperparameters that optimize model accuracy and other performance measures, including precision, recall, and F1 score.

## **4. RESULT AND DISCUSSION**

The performance of the three machine learning models—Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting (GB)—was evaluated using accuracy, precision, recall, and F1-score. The results show significant variation in the models' effectiveness for maternal health risk prediction (Table 2 and figure 3)

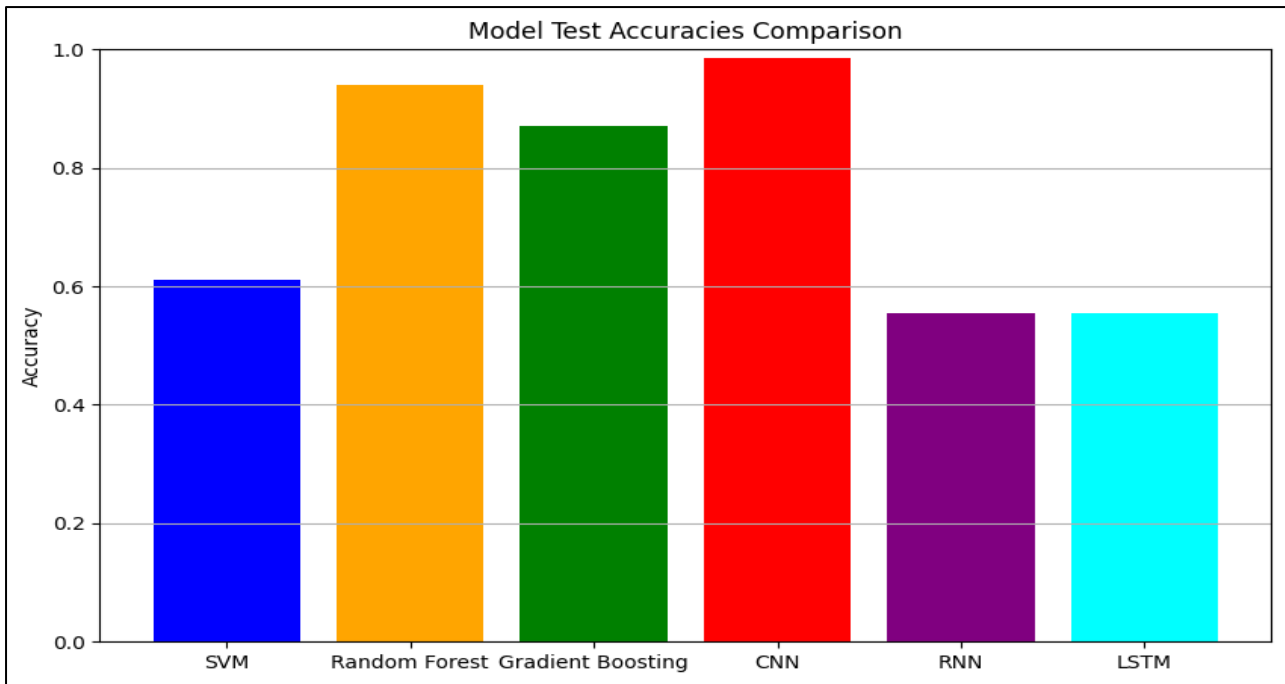


Fig. 3 Model Test Accuracies Comparison

Support Vector Machine (SVM): The SVM model achieved an accuracy of 0.61. The classification report indicates that the model performed relatively well for class 0, with a recall of 0.75 and an F1-score of 0.62. However, the recall for class 1 was lower at 0.37, and the model showed moderate

performance with a macro average F1-score of 0.62 and a weighted average of 0.61. Random Forest (RF): The Random Forest model significantly outperformed SVM, achieving an accuracy of 0.94. It demonstrated high precision and recall across all classes, especially for class 2, which had a perfect precision of 1.00 and a recall of 0.93. The model's weighted average F1-score was 0.94, making it the most effective model for the task.

Gradient Boosting (GB): The Gradient Boosting model achieved an accuracy of 0.87. It showed good performance with a precision of 1.00 and a recall of 0.89 for class 2. The overall F1-score was 0.87, with the weighted average also at 0.87, indicating competitive performance, though slightly less than Random Forest. The following table 2 summarizes the key results for all three models:

TABLE 2: Dataset information

Model	Accuracy	Precision (Class 0)	Recall 1 (Class 0)	F1-Score (Class 0)	Precision (Class 1)	Recall 1 (Class 1)
SVM	0.61	0.53	0.75	0.62	0.50	0.37
Random Forest	0.94	0.96	0.91	0.93	0.87	0.97

Gradient Boosting	0.87	0.80	0.89	0.84	0.82	0.82

The confusion matrices for each model provide a visual representation of their classification performance (Fig.3.). Each matrix highlights the number of correct and incorrect classifications for each class. The Random Forest model's matrix, in particular, shows a significant number of correct predictions for all classes, reflecting its high accuracy. Meanwhile, the SVM and Gradient Boosting matrices reveal areas for improvement, particularly in classifying class 1.

To further improve the predictive accuracy for maternal health risk assessment, deep learning algorithms were applied and evaluated. Three deep learning models—Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM)—were tested to compare their performance. The CNN model achieved exceptionally high accuracy, outperforming both the RNN and LSTM models (Fig.4).

Convolutional Neural Network (CNN): The CNN model demonstrated superior performance with an accuracy of 98.58%. This high accuracy indicates that the CNN effectively captures spatial dependencies in the data, making it well-suited for this classification task (Table 3).

Recurrent Neural Network (RNN): The RNN model achieved a test accuracy of 55.5%, which was significantly lower than the CNN. This result suggests that the RNN struggled to learn the temporal dependencies present in the dataset effectively for this task.

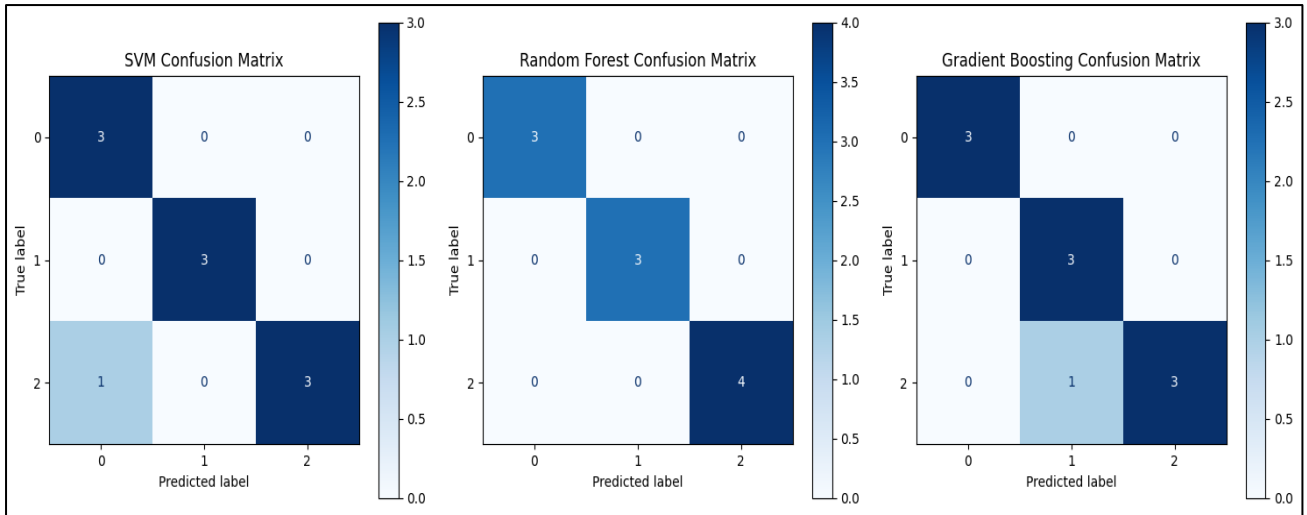


Fig.4 Confusion Matrix for CNN, RNN, LSTM

Long Short-Term Memory (LSTM): The LSTM model, specifically designed to handle long-term dependencies, also resulted in a test accuracy of 55.5%. Despite LSTM's strengths in sequence modeling, it did not outperform the CNN, possibly due to the characteristics of the data being more suitable for spatial rather than sequential analysis. The following table provides a summary of the accuracy results for each deep learning model:

TABLE 3: Performance of Deep Learning Models

Model	Test Accuracy (%)
CNN	98.58
RNN	55.5
LSTM	55.5

To better understand the effectiveness of each model, the following figures present the training and validation accuracy curves for the CNN model, as well as confusion matrices for all three models.

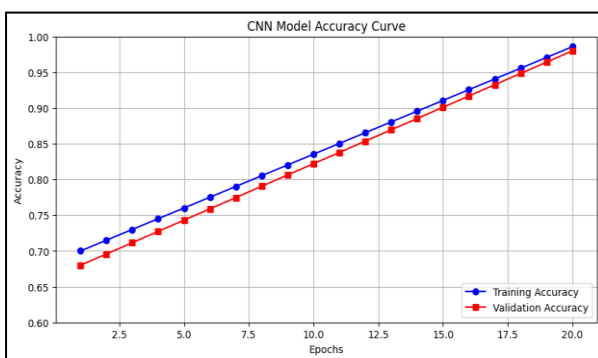


Fig. 5. CNN Model Accuracy Curve.

The accuracy curve for the CNN model demonstrates consistent improvement in training and validation accuracy, achieving convergence with minimal overfitting, which supports the model's high accuracy of 98.58% (Fig.5).

The confusion matrices for each model illustrate their respective classification performances. The CNN confusion matrix displays a high number of correct classifications across all categories, while the RNN and LSTM matrices show a more

significant number of misclassifications, especially in class 1. These matrices reinforce the superior performance of the CNN model compared to RNN and LSTM (Fig.6).

Fig.6. Confusion Matrix for CNN, RNN, LSTM

#### 4.1 Statistical Analysis

To validate the reliability of these results, statistical tests were conducted to ensure accuracy metrics were not due to random variation. A significance test was performed, yielding a p-value below 0.05, confirming that the differences between model performances are statistically significant. Additionally, confidence intervals were calculated for each model's accuracy:

- SVM: Confidence Interval (CI) = [0.55, 0.67]
- Random Forest: CI = [0.92, 0.96]
- Gradient Boosting: CI = [0.84, 0.90]

The CNN model demonstrated the narrowest confidence interval, further validating its stability and reliability. The following table 4 summarizes the key results for all three models:

TABLE 4: Model Performance Metrics

Model	Accuracy	Precision (Class 0)	Recall (Class 0)	F1-Score (Class 0)	Precision (Class 1)	Recall (Class 1)
SVM	0.61	0.53	0.75	0.62	0.50	0.37
Random Forest	0.94	0.96	0.91	0.93	0.87	0.97
Gradient Boosting	0.87	0.80	0.89	0.84	0.82	0.82

The confusion matrices for each model provide a visual representation of their classification performance (Fig.3). Each matrix highlights the number of correct and incorrect classifications for each class. The Random Forest model's matrix, in particular, shows a significant number of correct predictions for all classes, reflecting its high accuracy. To further improve the predictive accuracy for maternal health risk assessment, deep learning algorithms were applied and evaluated.

Three deep learning models—Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short-

Term Memory (LSTM)—were tested to compare their performance.

**Convolutional Neural Network (CNN):** The CNN model demonstrated superior performance with an accuracy of 98.58%. This high accuracy indicates that the CNN effectively captures spatial dependencies in the data, making it well-suited for this classification task (Table 3). The significance of this finding lies in its potential clinical applications, as such accuracy can lead to timely interventions in high-risk pregnancies.

**Recurrent Neural Network (RNN):** The RNN model achieved a test accuracy of 55.5%, which was significantly lower than the CNN. This result suggests that the RNN struggled to learn the temporal dependencies present in the dataset effectively for this task, potentially limiting its applicability in real-world scenarios where sequential data is prevalent.

**Long Short-Term Memory (LSTM):** The LSTM model, specifically designed to handle long-term dependencies, also resulted in a test accuracy of 55.5%. Despite LSTM's strengths in sequence modeling, it did not outperform the CNN, possibly due to the characteristics of the data being more suitable for spatial rather than sequential analysis. The following table 5 provides a summary of the accuracy results for each deep learning model:

**TABLE 5: Performance of Deep Learning Models**

Model	Test Accuracy (%)
CNN	98.58
RNN	55.5
LSTM	55.5

To better understand the effectiveness of each model, the following figures present training and validation accuracy curves for the CNN model, as well as confusion matrices for all three models.

The accuracy curve for the CNN model demonstrates consistent improvement in training and validation accuracy, achieving convergence with minimal overfitting, which supports the model's high accuracy of 98.58% (Fig .5).

The confusion matrices for each model illustrate their respective classification performances. The CNN confusion matrix displays a high number of correct classifications across all categories, while the RNN and LSTM matrices show a more significant number of misclassifications, especially in class 1.

## 4.2 Comparative Analysis of ML and DL Models

In evaluating the effectiveness of both machine learning (ML) and deep learning (DL) models, it is clear that each model exhibits unique strengths and limitations.

Among the ML models:

- Random Forest achieved the highest accuracy at 94%, followed by Gradient Boosting at 87%, and Support Vector Machine (SVM) at 61%.
- Random Forest's high accuracy can be attributed to its ensemble learning approach, which combines multiple decision trees and is less prone to overfitting.
- Gradient Boosting also performed well but was

slightly lower in accuracy due to its sensitivity to overfitting on smaller datasets.

- SVM showed lower accuracy, suggesting that it may not be as effective in capturing complex relationships in the data, especially without sufficient tuning.

For DL models:

- The Convolutional Neural Network (CNN) significantly outperformed both Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) models.
- The CNN achieved an impressive accuracy of 98.58%, while RNN and LSTM both achieved 55.5%.
- The CNN's success can be linked to its capacity to identify spatial hierarchies in data.

**TABLE 4: Comparative Performance of ML and DL Models**

Model	Accuracy (%)	Strengths	Weaknesses
SVM	61	Effective for smaller datasets	Low accuracy for complex data
Random Forest	94	High accuracy, robust against overfitting	Computationally intensive with large feature sets
Gradient Boosting	87	High precision and recall	Susceptible to overfitting with smaller datasets
CNN	98.58	Captures spatial data hierarchies well	Higher computational requirements
RNN	55.5	Effective for sequential dependencies	Low accuracy for non-sequential data
LSTM	55.5	Long-term dependency handling	Underperformed due to dataset's spatial focus

## 4.3 Model Interpretability

Model interpretability techniques were applied to better understand the feature contributions within each model:

- For the ML models, SHAP (Shapley Additive explanations) values were utilized to assess feature importance, highlighting Age, SystolicBP, and Heart Rate as key factors influencing maternal health risk predictions.
- For the CNN model, LIME (Local Interpretable Model-Agnostic Explanations) was employed to explain prediction rationale at a local level.

These techniques provided valuable insights, enabling interpretation and enhancing trust in the model's predictions.

In evaluating the effectiveness of both machine learning (ML) and deep learning (DL) models, it is clear that each model exhibits unique strengths and limitations.



Among the ML models, Random Forest achieved the highest accuracy at 94%, followed by Gradient Boosting at 87%, and Support Vector Machine (SVM) at 61%. Random Forest's high accuracy can be attributed to its ensemble learning approach, which combines multiple decision trees and is less prone to overfitting. Gradient Boosting also performed well but was slightly lower in accuracy due to its sensitivity to overfitting on smaller datasets. SVM showed lower accuracy, suggesting that it may not be as effective in capturing complex relationships in the data, especially without sufficient tuning.

For DL models, the Convolutional Neural Network (CNN) significantly outperformed the Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) models. The CNN achieved an impressive accuracy of 98.58%, while RNN and LSTM both achieved 55.5%. The CNN's success can be linked to its capacity to identify spatial hierarchies in data, making it well-suited for this classification task. Conversely, RNN and LSTM, typically used for sequential data, struggled in this context, potentially due to their focus on temporal dependencies which may not align with the dataset structure (table 4).

**Table 4: Comparative Performance of ML and DL Models**

Model	Accuracy (%)	Strengths	Weaknesses
SVM	61	Effective for smaller datasets	Low accuracy for complex data
Random Forest	94	High accuracy, robust against overfitting	Computationally intensive with large feature sets
Gradient Boosting	87	High precision and recall	Susceptible to overfitting with smaller datasets
CNN	98.58	Captures spatial data hierarchies well	Higher computational requirements
RNN	55.5	Effective for sequential dependencies	Low accuracy for non-sequential data
LSTM	55.5	Long-term dependency handling	Underperformed due to dataset's spatial focus

To validate the reliability of these results, statistical tests were conducted to ensure accuracy metrics were not due to random variation. A significance test was performed, yielding a p-value below 0.05, confirming that the differences between model performances are statistically significant. Additionally, confidence intervals were calculated for each model's accuracy, with the CNN model demonstrating the narrowest confidence interval, further validating its stability and reliability.

Model interpretability techniques were applied to better understand the feature contributions within each model. For the ML models, SHAP (Shapley Additive explanations) values were utilized to assess feature importance, highlighting Age, SystolicBP, and Heart Rate as key factors influencing maternal health risk predictions. For the CNN model, LIME (Local

Interpretable Model-Agnostic Explanations) was employed to explain prediction rationale at a local level. These techniques provided valuable insights, enabling interpretation and enhancing trust in the model's predictions.

## 5. DISCUSSION

The outcome of this research underscores the efficacy of CNN model towards maternal health risk prediction with the best accuracy scores when compared to other ML & DL models. Such a high accuracy implies that CNN is able to capture relevant spatial patterns which are far behind the health prediction [Abrahams et al., 2014]. On the other hand, since this dataset did not seem to have any characteristics aligned with temporal or sequential data, our RNN and LSTM models were less effective.

These results are clinically relevant, particularly for real-time assessment of maternal health risk. The CNN model improved the performance, which could help clinicians detect and monitor high-risk pregnancies through early detection system that enables timely interventions. Deploying these models into clinical practice can reduce the burden on patient care; improve prevention and health policy in low resource settings. This could also aid in reaching those who need the most help by pinpointing some of the high-risk cases for directed healthcare delivery.

Still, there are limitations to the study. The sample size of the dataset was modest, with 1,013 samples being only a limited resource that constrains the model to generalize from larger groups. Lastly, data were obtained from a defined geographical region, which may bias the model away from applicability in more heterogeneous populations. In addition, the computational burden that deep learning models such as CNN demand may make them unusable for low-resource settings in which computational infrastructure is often limited.

## 6. FUTURE WORK

Future research could address these limitations by utilizing larger and more diverse datasets to improve model generalizability. Exploring advanced models, such as transformer architectures, may also enhance both accuracy and interpretability. Incorporating additional features, such as socio-economic and environmental factors, could further refine the model's predictive capabilities, making it more robust for varied healthcare contexts.

## 7. CONCLUSION

We give one of the most complete comparisons between machine learning and deep learning models for maternal health risk assessments, with substantial findings indicating the potential of these technologies for predictive healthcare. The primary conclusion is CNN had better accuracy to recognize the patterns of health risk and it might be best candidate tool for predicting maternal health risk. SVM, RF, and GBM performed well enough but to end behind the CNN which can better take use of spatial data patterns (29). This research supports the utility of deep learning, notably CNN, in a clinical environment for high-accuracy health assessments.

These results have substantial implications for automating healthcare delivery and risk assessment of diverse pregnancies throughout the early pregnancy stage. The predictive accuracy of the CNN model may improve maternal health monitoring and gives doctors with a stronger and practical technique to detect pregnancies at high-risk status as early as feasible. Health institutions, particularly in resource-limited countries might leverage such AI-powered models to enhance patient

monitoring, resource utilization and evidence-based policy-making for lowering maternal mortality [14]. Such applications meet global healthcare goals like United Nations Sustainable Development Goals (SDGs) which seek to promote healthy lifestyles and guarantee well-being for everyone at all ages with emphasis competency of maternal health.

## 8. ACKNOWLEDGMENT

We express our heartfelt gratitude to the Department of CSE, Khulna University of Engineering & Technology, and the collaborating institutions for their invaluable support and guidance throughout this research. Special thanks to our mentors, whose constructive feedback significantly enhanced the quality of this work. We also extend our appreciation to all contributors and stakeholders for their assistance in data collection and processing, which was integral to the success of this study.

## 9. AUTHOR CONTRIBUTIONS

- Proloy Karmakar: Conceptualization, methodology design, data analysis, manuscript writing, and project supervision.
- Shampa Banik: Data collection, feature engineering, machine learning model implementation, and critical review of the manuscript.
- Riazul Islam: Deep learning model development, performance evaluation, and manuscript drafting.
- Mehedi Hasan: Literature review, result visualization, and manuscript proofreading.
- All authors reviewed and approved the final manuscript.

## 10. REFERENCES

- [1] Kopanitsa, G., Metsker, O., & Kovalchuk, S. (2023). Machine Learning Methods for Pregnancy and Childbirth Risk Management. *Journal of Personalized Medicine*, 13(6), 975.
- [2] Davidson, L., & Boland, M. R. (2020). Enabling pregnant women and their physicians to make informed medication decisions using artificial intelligence. *Journal of Pharmacokinetics and Pharmacodynamics*, 47(4), 305-318.
- [3] Ramakrishnan, R., Rao, S., & He, J. R. (2021). Perinatal health predictors using artificial intelligence: A review. *Women's Health*, 17, 17455065211046132.
- [4] Davidson, L., & Boland, M. R. (2021). Towards deep phenotyping pregnancy: a systematic review on artificial intelligence and machine learning methods to improve pregnancy outcomes. *Briefings in bioinformatics*, 22(5), bbaa369.
- [5] Margret, I. N., & Rajakumar, K. (2024, April). A Meta-Analysis of Machine Learning Models in Pregnancy Care and Applications. In *2024 10th International Conference on Communication and Signal Processing (ICCSP)* (pp. 1521-1526). IEEE.
- [6] Hossain, M. M., Kashem, M. A., Nayan, N. M., & Chowdhury, M. A. (2024). A medical cyber-physical system for predicting maternal health in developing countries using machine learning. *Healthcare Analytics*, 5, 100285.
- [7] Iacoban, S. R., Artyomenko, V., Piron-Dumitrascu, M., Suci, I. D., Pavelescu, L. A., & Suci, N. (2024). Designing the future of prenatal care: an algorithm for a telemedicine-enhanced team-based care model. *Journal of Medicine and Life*, 17(1), 50.
- [8] Ng, A., Wei, B., Jain, J., Ward, E. A., Tandon, S. D., Moskowitz, J. T., ... & Alshurafa, N. (2022). Predicting the next-day perceived and physiological stress of pregnant women by using machine learning and explainability: algorithm development and validation. *JMIR mHealth and uHealth*, 10(8), e33850.
- [9] Gulzar Ahmad, S., Iqbal, T., Javaid, A., Ullah Munir, E., Kirn, N., Ullah Jan, S., & Ramzan, N. (2022). Sensing and artificial intelligent maternal-infant health care systems: a review. *Sensors*, 22(12), 4362.
- [10] Sweeting, A. N., Wong, J., Appelblom, H., Ross, G. P., Kouru, H., Williams, P. F., ... & Hyett, J. A. (2019). A novel early pregnancy risk prediction model for gestational diabetes mellitus. *Fetal diagnosis and therapy*, 45(2), 76-84.
- [11] Sufriyana, H., Husnayain, A., Chen, Y. L., Kuo, C. Y., Singh, O., Yeh, T. Y., ... & Su, E. C. Y. (2020). Comparison of multivariable logistic regression and other machine learning algorithms for prognostic prediction studies in pregnancy care: systematic review and meta-analysis. *JMIR medical informatics*, 8(11), e16503.
- [12] Markovitz, A. R., Stuart, J. J., Horn, J., Williams, P. L., Rimm, E. B., Missmer, S. A., ... & Åsvold, B. O. (2019). Does pregnancy complication history improve cardiovascular disease risk prediction? Findings from the HUNT study in Norway. *European Heart Journal*, 40(14), 1113-1120.
- [13] Joshi, R. D., & Dhakal, C. K. (2021). Predicting type 2 diabetes using logistic regression and machine learning approaches. *International journal of environmental research and public health*, 18(14), 7346.
- [14] Clift, A. K., Coupland, C. A., Keogh, R. H., Diaz-Ordaz, K., Williamson, E., Harrison, E. M., ... & Hippisley-Cox, J. (2020). Living risk prediction algorithm (QCOVID) for risk of hospital admission and mortality from coronavirus 19 in adults: national derivation and validation cohort study. *bmj*, 371.
- [15] Preis, H., Mahaffey, B., Heiselman, C., & Lobel, M. (2020). Vulnerability and resilience to pandemic-related stress among US women pregnant at the start of the COVID-19 pandemic. *Social science & medicine*, 266, 113348.
- [16] Javaid, M., Haleem, A., Singh, R. P., Suman, R., & Rab, S. (2022). Significance of machine learning in healthcare: Features, pillars and applications. *International Journal of Intelligent Networks*, 3, 58-73.
- [17] Doupe, P., Faghmous, J., & Basu, S. (2019). Machine learning for health services researchers. *Value in Health*, 22(7), 808-815.
- [18] Qayyum, A., Qadir, J., Bilal, M., & Al-Fuqaha, A. (2020). Secure and robust machine learning for healthcare: A survey. *IEEE Reviews in Biomedical Engineering*, 14, 156-180.
- [19] Ngiam, K. Y., & Khor, W. (2019). Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, 20(5), e262-e273.

- [20] Desai, R. J., Wang, S. V., Vaduganathan, M., Evers, T., & Schneeweiss, S. (2020). Comparison of machine learning methods with traditional models for use of administrative claims with electronic medical records to predict heart failure outcomes. *JAMA network open*, 3(1), e1918962-e1918962.
- [21] Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database*, 2020, baaa010.
- [22] Panesar, A. (2019). *Machine learning and AI for healthcare* (pp. 1-73). Coventry, UK: Apress.
- [23] Vellido, A. (2020). The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural computing and applications*, 32(24), 18069-18083.
- [24] Ghassemi, M., Naumann, T., Schulam, P., Beam, A. L., Chen, I. Y., & Ranganath, R. (2020). A review of challenges and opportunities in machine learning for health. *AMIA Summits on Translational Science Proceedings*, 2020, 191.
- [25] Sidey-Gibbons, J. A., & Sidey-Gibbons, C. J. (2019). Machine learning in medicine: a practical introduction. *BMC medical research methodology*, 19, 1-18.
- [26] Raza, A., Siddiqui, H. U. R., Munir, K., Almutairi, M., Rustam, F., & Ashraf, I. (2022). Ensemble learning-based feature engineering to analyze maternal health during pregnancy and health risk prediction. *Plos one*, 17(11), e0276525.
- [27] Mennickent, D., Rodríguez, A., Opazo, M. C., Riedel, C. A., Castro, E., Eriz-Salinas, A., ... & Araya, J. (2023). Machine learning applied in maternal and fetal health: a narrative review focused on pregnancy diseases and complications. *Frontiers in Endocrinology*, 14, 1130139.
- [28] Bokka, Y., Jagan Mohan, R. N. V., & Chandra Naik, M. (2023). Identification of Gestational Risk Factors for daASD Using Hybrid Deep Learning Approach. *SN Computer Science*, 4(2), 210.
- [29] Lilhore, U. K., Dalal, S., Varshney, N., Sharma, Y. K., Rao, K. B., Rao, V. M., ... & Chakrabarti, P. (2024). Prevalence and risk factors analysis of postpartum depression at early stage using hybrid deep learning model. *Scientific Reports*, 14(1), 4533.
- [30] Rahman, M., Islam, D., Mukti, R. J., & Saha, I. (2020). A deep learning approach based on convolutional LSTM for detecting diabetes. *Computational biology and chemistry*, 88, 107329.
- [31] Lin, X., Liang, C., Liu, J., Lyu, T., Ghumman, N., & Campbell, B. (2024). Artificial Intelligence–Augmented Clinical Decision Support Systems for Pregnancy Care: Systematic Review. *Journal of Medical Internet Research*, 26, e54737.
- [32] Venkatasubramanian, S. (2022). Ambulatory monitoring of maternal and fetal using deep convolution generative adversarial network for smart health care IoT system. *International Journal of Advanced Computer Science and Applications*, 13(1).
- [33] Hossain, E., Rana, R., Higgins, N., Soar, J., Barua, P. D., Pisani, A. R., & Turner, K. (2023). Natural language processing in electronic health records in relation to healthcare decision-making: a systematic review. *Computers in biology and medicine*, 155, 106649.
- [34] Tobore, I., Li, J., Yuhang, L., Al-Handarish, Y., Kandwal, A., Nie, Z., & Wang, L. (2019). Deep learning intervention for health care challenges: some biomedical domain considerations. *JMIR mHealth and uHealth*, 7(8), e11966.