

A Breakdown of Feature Importance Ranking and Trimming (FIRT) for Maternal Health Risk Prediction Using Physiological Factors and Machine Learning

Shilpa Mahajan

Computer science and Engineering
The NorthCap University
Haryana, India
shilpa@ncuindia.edu

Garima Sharma

Computer science and Engineering
The NorthCap University
Haryana, India
garimasharma@ncuindia.edu

Abdullah Ayub Khan

Faculty of Informatics and Computing
Universiti Sultan Zainal
Abidin, Malaysia
abdullah.khan00763@gmail.com

Asif Ali Laghari

Software College
Shenyang Normal University
Shenyang, China
asiflaghari@synu.edu.cn

Angela M. Alves

Poli.TIC
CTI
Campinas, Brazil
amalves@cti.gov.br

Gabriel G. de Oliveira

FEEC
UNICAMP
Campinas, Brazil
ggomes@unicamp.br

Vania Vieira Estrela

Department of Telecommunications
UFF
Niteroi, Brazil
vania.estrela.phd@ieee.org

Abstract—This study examines the nuts and bolts of Machine Learning (ML) when identifying maternal health concerns using previously sampled indicators. The Feature Importance Ranking and Trimming (FIRT) exploits ANOVA F-tests, Mutual Information (MI), and "Recursive Feature Elimination" (RFE) to reduce the feature vector (FV) entries (variables) and assess model soundness. The shorter, more expressive FV improves focal predictor identification, interpretability, and training efficiency, while reducing overfitting. Key methods for measuring FV entries' importance include (a) measuring a feature's total node impurity reduction (e.g., by evaluating entropy) across all possibilities, where a higher total reduction indicates greater appeal, (b) shuffling FV entries and measuring the breakdown in model accuracy, where a suggestive drop in performance indicates a highly noteworthy feature, and (c) gauging how much accuracy must be This study predicted maternal health risks using a Kaggle dataset from several sources. This study selected and ranked features following data preprocessing. The results show that FIRT, the ANOVA F-test, and RFE performed better than MI. After model training and testing, feature selection occurred. The models used in this study include K-nearest Neighbors (KNN), Random Forest (RF), "Gradient Boosting" (GB), "Decision Tree" (DT), "Support Vector Classifier" (SVC), and "Logistic Regression." The RF delivered the best accuracy (87%) and performance. Random forests have been the most accurate and robust models for predicting maternal well-being risk. Additional feature-selection rationales for ML models can boost future usefulness and efficacy. A large dataset spanning multiple states and nations can boost forecast accuracy.

Index Terms—Maternal health risks, machine learning, dimensionality reduction, feature selection, feature ranking, RFE, KNN, SVC, ANOVA F-test.

I. INTRODUCTION

MATERNAL Health (MH) encompasses women's well-being throughout pregnancy, childbirth, in addition

to the postpartum period and supports the foetus's good health. There are various health risks associated with women throughout and after childbirth, which can bring about critical conditions, even mortality in some cases. Various disorders, viz., preeclampsia, hypertension, high sugar levels, etc., are high risks to maternal health. It is indispensable to promptly understand and weigh MH risks. These risks could have long-term chronic effects on women, as the risk of cardiac diseases and metabolic diseases increases later in their lives.

As per the WHO, roughly 140 million births ensue annually, and the percentage guarded by qualified health professionals has gone up from 58% in 1990 to 81% in 2019. This advancement is primarily attributable to more births occurring at health facilities. The Mortality Ratio (MR) owing to pregnancy complications has also gone down by 38% from the last twenty years. The mean annual MR decay is only 3%, and hence progress is very slow [9]. The Maternal MR (MMR) in India during 2018-2020 has been reported in the newest "Sample Registration System" (SRS) data, totalling 97/100000 live births. This figure amounts to a decrease of 33 points, compared to 130/ 100000 live births for the period of 2014-2016 [18].

Risk factors, e.g., abnormal blood sugar (BS) levels, hypertension, thyroid dysfunction, body temperature, and heart rate, can lead to critical conditions during childbirth [4], [5], [19]. Out of these, the hypertensive illnesses are the most frequent, affecting 6–8% of all pregnancies [14]. The previously stated factors could pose high risks, and some could lead to long-term effects and diseases. Although health facilities have improved, many complications persist in areas where accurate risk assessment resources are lacking. The

areas that lack basic needs for maternal healthcare are more predisposed to these health complications. The MR may be decreasing, but countless serious health complications, like internal bleeding, diabetes, preeclampsia, etc., can affect a woman's life before, during, and after childbirth, besides impacting the foetus's life. Maternal emotional stability and nutritional supplementation are essential for predicting risks, e.g., autism spectrum disorder in offspring [19]. MH is a paramount component of universal healthcare. Timely risk prediction of maternal health can enhance clinical decision-making and outcomes.

Machine Learning (ML) models are widely used to predict perinatal complications, viz., preeclampsia, gestational diabetes, preterm birth, perinatal mortality, and foetal malformations. These models utilize data from "Electronic Medical Records" (EMR), medical imagery, biological markers, and sensor-obtained data to recognize at-risk pregnancies and enable early intervention [3], [7], [8]. ML models enable early identification of high-risk pregnancies, support timely clinical interventions, inform resource allocation [13], as well as predict MH risks. Previous work, such as that by M. Ukrit et al. [3], has applied a variety of ML algorithms but has suffered from low model efficiency and suboptimal performance. Many studies, e.g., S. Aggarwal, N. Maitra [1], and Sweeting et al. [16], relied on small or imbalanced datasets, yielding overfitting and poor generalizability.

Orthodox algorithms, namely, "Logistic Regression" (LR), "K-nearest Neighbour" (KNN), "Support Vector Classifier" (SVC), and "Random Forest" (RF) have formed the foundation of many models [12], [17]. Recent research has shifted towards more complex methods, viz., "Artificial Neural Networks" (ANNs) and "Deep Learning" (DL) [17]. These works achieve high accuracy, but the models are very computationally intensive, limiting their practical feasibility and deployment. The existing literature also faces issues of biased performance metrics [2], negligence, and potentially exaggerated model performance.

Decision Trees (DTs) and Neural Networks (NNs) have been employed to forestall pregnancy complications. Some existing models have accurately acknowledged high-risk pregnancies through vital signs and synthetic health data for prediction [1], [17]. Risk prediction models emphasize factors, like advanced maternal age, usage of assisted reproductive technology, and lifestyle factors like smoking and obesity, that heighten the prospect of adverse outcomes [12], [16]. Studies suggest that multivariable models incorporating symptoms like oxygen saturation and chest pain/dyspnea are more effective at predicting adverse aftermaths than single predictors [1].

Many prediction models are prone to bias and lack sufficient internal and external validation. This limits their clinical applicability and highlights the need for methodological improvements and comprehensive validation studies. Effective prediction models often require integrating diverse data sources, including patient medical history, demographics, literacy, and lifestyle factors. While several models show promise, their utility in clinical settings for patient counselling and care

optimization remains to be fully evaluated.

This paper evaluates ML tactics centered on feature selection for predicting MH risk into low-, mid-, and high-level categories, including RFs, KNNs, SVCs, DTs, and LR, and compares MH risk assessments [4]. Most available ML models are hampered by inadequate validation and feature selection, leading to reduced accuracy and clinical utility. Henceforth, constructing an efficient, scalable ML framework to accurately predict MH complications utilizing appropriate physiological measurements is imperative. In response to these challenges, the present investigative effort also gauges performance via metrics, such as Accuracy, Precision, Recall, and F1-score.

The main contribution of this investigation is to expose how Feature Vector (FV) entries (aka, health variables) behave regarding importance ranking and dimensionality reduction. FV entries may affect the assessment of MH risks. Feature Importance Ranking and Trimming (FIRT) relies on variable-importance techniques (e.g., ANOVA F-test, "Mutual Information" (MI), and "Recursive Feature Elimination" (RFE)) to identify the most explanatory features for predicting MH risk. ML feature selection streamlines computation, boosts learning accuracy, and deepens data and learning model knowledge.

This study leveraged a well-curated MH dataset comprising originally 1,200 records. The diverse dataset enhances applicability and mitigates demographic bias when applied to real-world patients. FIRT ameliorates model accuracy and shrinks computational time.

This paper's organization follows: Section 2 discourses the literature review and identifies gaps. Section 3 portrays the study's materials and methodology, in which each FV holds seven important physiological entries. Section 4 describes the proposed methodology and associated minutiae diagrammatically. Section 5 encompasses experimental results and discussion. Conclusions appear in Section 5.

II. RELATED WORK

In the academic literature, numerous researchers have investigated different ML models for predicting MH risks to support welfare decision-making. This section reviews numerous studies focused on classifying MH risks and diseases and highlights existing research gaps. Previous research has shown that a multitude of ML algorithms can forecast MH risks. However, despite promising outcomes, there are still areas for further investigation and improvement [15].

In previous work, M. Ukrit and others used ML to predict MH risks, but performance was low due to the model's low efficiency [11]. To overcome this, they castoff multiple algorithms and developed a more efficient model for MH risk prediction.

Decidedly, old-style ML measures (like LR, KNN, SVC, and RF) have served as the basis for solutions. [6], [12], [17]. Yet, various investigators have been exploiting ANNs and DL to improve model performance [17]. Although ANNs achieve high accuracy, they still have limitations, such as limited feasibility and an inability to handle unstructured data [17].

Further, the datasets' size and quality matter for the accuracy of predictive models. A few analyses, such as those by S. Agrawal and N. Maitra (2016) and M. Ukrit et al. (2024), were affected by small sample sizes, leading to overfitting and limited generalizability [1]. There needs to be a larger diversity of datasets that enable the construction of models that work in practical applications. Additionally, data costs and accessibility continue to pose challenges, as revealed by A. Sweeting et al. (2018), thereby hindering the adoption of advanced approaches [16].

Md Assaduzzaman, Abdullah Al Mamun, and Md Zahid Hasan (2023) tested different ML approaches, including CatBoost, RF, XGBoost (XGB), DT, and Gradient Boosting (GB), where RF achieved the highest accuracy of 90% [2]. This study reported biased performance metrics, indicating overestimated accuracy scores. Given that more fresh data will be reachable and the nature of our modeling will not allow for biased trends, the study emphasizes the need for unbiased evaluations of our model capabilities [2].

Although accuracy is a significant performance metric, quantitative efficiency, qualitative feasibility, and qualitative applicability are also considered in model selection. Good models are worth nothing if they require costly computational resources or do not generalize across the dataset.

To create a reliable and effective framework, this study has addressed all previously identified research gaps, employed feature selection, and trained the model, thereby augmenting accuracy and overall model performance.

III. MATERIALS AND METHODS

This section overviews the materials and procedures implemented in this research, outlines the study's dataset, analyzes feature engineering and ML methods for predicting MH risks, and illustrates the multiple steps of the developed methodology in greater detail.

A. Dataset Description

The dataset stems from an IoT-based MH monitoring system [10]. It contains physiological data from hospitals, community clinics, and MH facilities. The dataset lacks geographic locations, demographic segmentation, and socioeconomic attributes of the subjects in Table 1. Missing information limits the extent to which the dataset can be deemed typical or applicable beyond the sample included in it. Future efforts will validate the model across diverse populations and more clinically curated datasets to ensure reliability and applicability in clinical practice.

This dataset consists of several feature vectors, each one having seven entries: Age, systolic BP, diastolic BP, BS, Body Temp, Heart Rate, and Risk Level, as described in Table 2. The risk level includes low-, mid-, and high-risk levels. The dataset contains 1,200 instances divided across 7 attributes. These features help one identify the most risk-causing factor in females.

Table 3 describes the dataset, which captures a wide range of patient records. It depicts various physiological factors such

as age, BS, Body Temp, Heart Rate, systolic BP, diastolic BP, and the corresponding Risk Levels.

B. Feature Selection Techniques Relying on FIRT

This section discusses feature selection and the ML techniques used to classify MH risks by briefly introducing the basic concepts underlying each algorithm.

Feature Importance Ranking and Trimming (FIRT) performs dimensionality reduction of the entries (variables) of a feature vector (FV) relying on their contributions to model accuracy. The new FV is shorter and more representative, aiming at better identifying critical predictors, improving model interpretability, reducing overfitting, and enhancing training efficiency. Key methods for measuring FV entries' Importance can be evaluated by (a) measuring the total reduction in node impurity (e.g., by evaluating entropy) that a feature provides across all possibilities, so that a higher total reduction indicates greater importance, (b) shuffling FV entries and measuring the resulting provide a breakdown in model accuracy, noting that a significant drop in performance indicates a highly important feature, and (c) gauging how much accuracy is lost when a feature is removed or randomized.

1) *Assessment Metrics* : Feature selection identifies the most significant and relevant features in a dataset, thereby refining model performance and accuracy. The number of features can be curbed via various techniques:

- **a) ANOVA F-Test:** This statistical scheme compares variances across feature groups and their associations with the target variable. It is a filter-based method. The F-statistic tests the between group variance; it is defined by:

$$F = \frac{\text{Variancebetween groups}}{\text{Variance within groups}} \quad (1)$$

A higher value of F means features are more significant in classification, and a lower value of F means features are less significant.

- **b) Recursive Feature Elimination (RFE):** It is a wrapper method that iteratively eliminates the least significant features employing a selected ML model. This method finds the optimal subset of features for prediction. It is highly efficient because it operates recursively.
- **c) Mutual Information (MI):** This technique also comes under the filter method of feature selection. It is an information-theoretic method that measures the dependence between each FV entry and the target variable. It works for both numerical and categorical data and can seize nonlinear relationships. The MI is

$$I(X; Y) = \sum_{x \in N} \sum_{y \in N} p(x, y) \log \left(\frac{p(x, y)}{p(x) \cdot p(y)} \right) \quad (2)$$

Where $I(X; Y)$ is the MI between X and Y. Likewise, $p(x)$ and $p(y)$ are the marginal probabilities of X and Y, correspondingly, with $p(x, y)$ is the joint probability of X and Y.

TABLE I
DESCRIPTION OF ATTRIBUTES IN THE DATASET.

Features	Description
Age	The age of the woman at the time of her pregnancy
Systolic BP	The upper limit value of a pregnant woman's blood pressure (systolic blood pressure) is in mmHg.
Diastolic BP	The lower limit value of a pregnant woman's blood pressure (diastolic blood pressure) is in mmHg.
BS	The blood glucose concentration in peripheral blood, expressed in mmol/L.
Body Temp	The individual's body temperature in degrees Celsius.
HeartRate	The average heart rate over a certain time period at rest in the given units bpm.
Risk Level	Estimated value of risk intensity level during pregnancy based on prior variables.

TABLE II
DATASET OVERVIEW.

	Age	systolicBP	DiastolicBP	BS	BodyTemp	HeraRate	RiskLevel
1009	22	120	30	15.0	98.0	80	high risk
1010	55	120	90	18.0	98.0	60	high risk
1011	35	85	60	19.0	98.0	86	high risk
1012	43	120	90	18.0	98.0	70	high risk
1013	32	120	65	6.0	101.0	76	mid risk

2) *Machine Learning Techniques*: ML enables computers to learn predictive models or make decisions without explicit programming through Supervised Learning (SL), Unsupervised Learning (UL), and Reinforcement Learning (RL). This study classifies the MH risks employing SL methods, as follows:

- **a) Random Forest (RF)**: RF is an ensemble method that builds multiple DTs and merges them to raise accuracy and lessen overfitting. It makes use of bagging to train on a random subset of the dataset. RF handles large datasets efficiently. The RF prediction formula is

$$\hat{y} = \text{mode}(T_1(x), T_2(x), \dots, T_n(x)) \quad (3)$$

Where $T_1(x)$, $T_2(x)$, ..., $T_n(x)$ are the number of distinct DTs, which are trained on random subsets. $T_i(x)$ is the class prediction for input x , and y related to the mode (most recurrent value) of all predictions.

- **b) Gradient Boosting (GB)**: GB is also an ensemble learning that builds trees sequentially. First, a DT is trained on the data; subsequently, the next model is trained on the previous model using gradient descent. Then, all trees are combined to make a final prediction model. The GB is predicted as follows.

Where $F_m(x)$ is the boosted model after m number of iterations. $F_{m-1}(x)$ is the model from the previous iteration. $h_m(x)$ is the new weak learner, and is the learning rate (or step size) for this iteration.

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \quad (4)$$

- **c) Decision Tree (DT)**: A DT is a tree-like model for classification and regression that partitions the dataset at

each step through recursively picking the most informative feature, thereby reducing impurity. Entropy is the information gain from splitting, calculated as

$$H(X) = - \sum p(x) \log_2 p(x) \quad (5)$$

Where p is the proportion of class instances in the X dataset. Entropy is 0 when the dataset contains only instances from a single class. Entropy is maximum when classes are evenly mixed.

- **d) Support Vector Classifier (SVC)**: SVC is a nonlinear classification technique that finds the hyperplane (the optimal classification decision boundary). It separates the classes using a hyperplane that maximizes the margin. The margin is defined via support vectors and the kernel trick. In a linear SVC, the decision boundary or hyperplane is calculated as depicted in equation (6).

$$w \cdot x + b = 0 \quad (6)$$

Where w stands for the weight vector, x is the input FV, and b is the bias intercept term.

- **e) K-Nearest Neighbour (KNN)**: KNN is an instance-based technique that classifies a data point based on the majority class of its nearest neighbours. It computes the distance between the data point and all other training points. In this technique, we select a value for K , typically between 1 and 5, indicating the number of k -nearest neighbours. The Euclidean distance is calculated as

$$d(x, y) = \sqrt{\sum_n^i (x_i - y_i)^2} \quad (7)$$

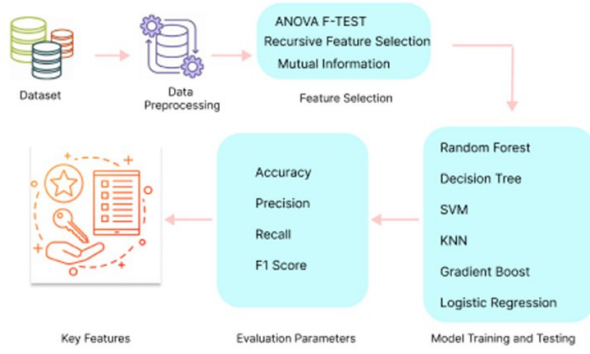


Fig. 1. Representation of different phases of the proposed methodology utilized in this study.

Where x and y form a pair set, KNN is a down-to-earth algorithm that performs capably on small datasets but is typically slower on larger ones.

- **f) Logistic Regression (LR):** LR is an originally linear binary classification model that can also work for multi-class classification, being simple, fast, and interpretable when applied to linear relationships. A sigmoid function that maps predictions to values within $[0, 1]$ can be pivotal to LR:

$$\sigma = \frac{1}{1 + e^{-x}} \quad (8)$$

where x is the input value, and e is Euler's number (approx. 2.718).

IV. PROPOSED METHODOLOGY

This section introduces the proposed approach with FIRT and ML for MH risk classification. In the methodology, the first step is data preprocessing, in which EDA (Exploratory Data Analysis) is done, then the second is feature selection on the dataset, the third is model testing and training, and the model evaluation and result discussion is done as depicted in Fig. 1.

A. Data Preprocessing

This phase preprocesses the dataset with respect to quantity and quality by passing it through multiple sub-stages to enhance the classification performance for attribute (object) types. Then, the counts, minimum, maximum, standard deviation, in addition to the 25th, 50th, and 75th percentiles of the total values for each attribute, were determined.

In a subsequent stage, null values were checked in the dataset. The dataset contains no null values, so no changes were necessary until now. After checking for null values, a label encoder is used to convert the Risk levels from categorical to numerical. With the help of label encoding, the low risk is converted to 0, the mid risk to 1, and the high risk to 2. To check how the features are correlated with each

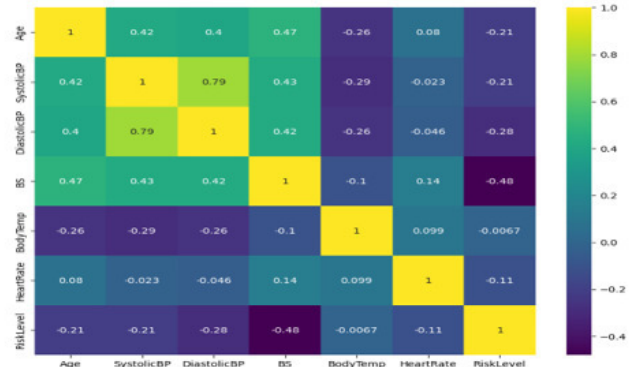


Fig. 2. Correlation heatmap.

other, a heatmap is generated (Fig. 2). The heatmap shows how the features are correlated with each other and with the target variable. Each cell shows the Pearson correlation coefficient:

- +1 for strong positive correlation;
- -1 for strong negative correlation;
- 0 for no linear correlation.

This heatmap shows that BS (Blood Sugar), Age, DiastolicBP, and SystolicBP have strong negative correlations, which help predict risk levels. BodyTempo and HeartRate have a very weak correlation, which makes them less useful for our data and model.

After the heatmap, outliers were detected using boxplots. It was found that two attributes, named BS and BodyTemp (Body Temperature), contain outliers. Outliers are the data items that meaningfully deviate from the rest of the data items or objects. To detect the outliers, a boxplot is used, as in Fig. 3(a).

A box plot is a graphical representation of the distribution of groups of numeric data depicted using boxes and lines. Each box limit indicates the central 50% of the data, with a central line marking the median. It demonstrates the locality, spread, and skewness of numerical data. Detecting outliers matters because they can skew the mean, affect measures of central tendency, and influence test results. To remove the outliers, the Inter-quartile Range (IQR) method is used, then visualised using a boxplot to check if any outliers are left, as shown in Fig. 3(b).

From Figure 3(a), it is clear that BS and BodyTemp contain outliers. These outliers were removed using the IQR method and depicted in Figure 3(b), where it can be seen that BS and BodyTemp have no more outliers in them. EDA also analyses information more clearly by exploring data, visualizing, and summarizing relationships between attributes and important characteristics. It has been done before applying ML models. In this study, EDA depicts risk levels in a pie chart (Fig. 4).

The risk-level rates indicate that 26.8% of patients are high risk (Figure 4). The mid-risk distribution rate is 33.1%, which is higher than that of the high-risk category. The largest proportion of patients falls into the low-risk category (40%), indicating that the dataset is reasonably balanced but not

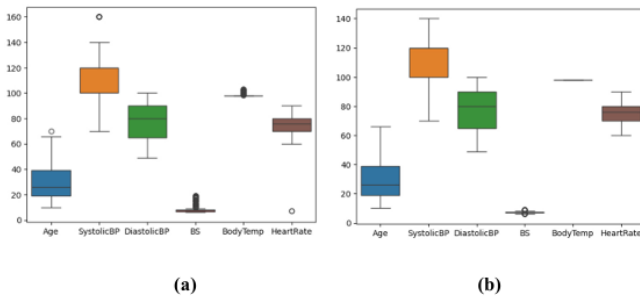


Fig. 3. (a) Boxplot to visualize outliers in each feature. (b) Boxplot after the outlier removal.

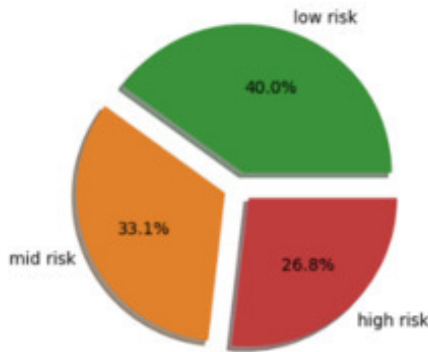


Fig. 4. Risk level distribution rate.

perfectly uniform. The high-risk percentage underscores the importance of engendering an accurate predictive model.

B. FIRT Strategy To Handle Feature Vector Entries

The original dataset had 7 feature entries per feature vector. The second step is feature selection, which involves selecting the optimal feature subset to improve outcomes and ranking the features. This step is optional because it is not always necessary, but it can aid model optimization.

- **ANOVA F-test:** It checks how the numerical features separate classes of a target variable or the categorical target. The FV entries selected using this method are ['Age', 'SystolicBP', 'DiastolicBP', 'BS', 'HeartRate'].
- **RFE:** This method recursively picks FV entries by eliminating the least influential features for each model. First, models were trained on each feature set using RFE, and the best features were selected. The features that were selected using this method are ['SystolicBP', 'DiastolicBP', 'BS', 'BodyTemp', 'HeartRate'].
- **MI:** It appraises the mutual dependence between the target variable and an FV entry in order to quantify the information a feature yields about other features. The features selected using this method are ['Age', 'SystolicBP', 'DiastolicBP', 'BS', 'HeartRate'].

Repeatedly emphasized that blood sugar (BS), diastolic BP, and systolic BP are the most relevant indicators of maternal

health risk. This is in conjunction with clinically verified medical literature. Increased glucose levels during diabetes are observed to be a common, serious condition during pregnancy because it might lead to preterm delivery, disproportionate foetal growth, and preeclampsia [5], [7]. Even moderate elevations in diastolic pressure can indicate vascular stress, which is the major contributor to adverse maternal outcomes. Both systolic and diastolic hypertension together form the diagnostic basis for the different types of hypertension in pregnancy, which affects nearly 8% of pregnancies globally [14]. The fact that these features emerged as strong predictors in our analytical methodology underscores their clinical significance and demonstrates that the model's learning has incorporated acknowledged risk factors in obstetric medicine. Combining the ML and clinical perspectives enhances the reliability and interpretability of our framework.

After feature selection, the dataset is ready for model training and testing, as the best features have been identified using these methods.

C. Model Training and Testing

This classification using ML algorithms with FIRT required chosen features for model training with training and test sets at 80%/20%. ML models comprising RF, GB, DT, SVC, KNN, and LR were employed to forecast MH risks.

D. Model Assessment and Performance Metrics

After training and testing, the performance of each model is evaluated using certain performance metric (Table 4).

V. EXPERIMENTAL RESULTS, DISCUSSION AND FUTURE WORK

A. Experimental Setup

In this section, the model results are discussed and compared with one another and with previous models. The visualization has also been done to gain deeper insight into the dataset and classification models. By comparing the performance of these models, one can assess how well they predict maternal health risks.

All methods and models were run on Windows 11 on a 12th-generation Intel Core i5 processor with Intel Iris Xe Graphics and 16 GB of RAM. All computations were executed in Google Colab using Python libraries (embracing Pandas, NumPy, Seaborn, Matplotlib, and scikit-learn). Table 4 reports the important highlights of the experimental setup.

The accuracy, precision, recall, and F1 score are reported in Table 5.

B. Discussion

ML metrics have been identified and compared for each model based on FIRT stage outputs, which relied on ANOVA, F-tests, RFE, and MI. Table 6 shows how accurately each model predicts MH risks.

Among all the models, the RF reached the highest accuracy of 87% across all feature selection methods, as shown in Table 6. Because of its ensemble-learning capabilities, it also

TABLE III
DESCRIPTION OF ATTRIBUTES IN THE DATASET.

Measure	Formula	Description
Accuracy	$\frac{TP+TN}{TP+FP+TN+FN}$	It gauges the fraction of correctly classified samples, both malware (TP) and benign (TN), over all samples.
Precision	$\frac{TP}{TP+FP}$	Helps indicate the proportion of samples predicted as malware that are malicious; high precision indicates low false positives.
Recall	$\frac{TP}{TP+FN}$	Represents the proportion of malware samples that were correctly identified as such. Supreme recall means micro false negatives.
F1-score	$2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$	It provides a balanced average of precision and recall and is effective in skewed datasets, where both precision and recall are required.

TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative

TABLE IV
COMPARISON OF PERFORMANCE BEFORE FEATURE SELECTION.

Model name	Accuracy	Precision	Recall	F1-score
Decision Tree	0.77	0.82	0.82	0.82
Random Forest (RF)	0.81	0.82	0.82	0.81
SVC	0.67	0.70	0.68	0.63
KNN	0.63	0.63	0.64	0.63
Gradient Boosting	0.73	0.74	0.74	0.75
Logistic regression	0.63	0.66	0.65	0.61

TABLE V
APPRAISAL OF EACH MODEL'S ACCURACY WITH RESPECT TO FEATURE SELECTION METHODS.

1	Decision Tree	0.80	0.80	0.77
2	Random Forest (RF)	0.87	0.87	0.86
3	SVC	0.74	0.78	0.74
4	KNN	0.70	0.70	0.70
5	Gradient Boosting	0.81	0.81	0.81
6	Logistic regression	0.65	0.65	0.64

handles complex data. This depicts the robustness and the generalization capability of the RF. The GB also performed well, achieving 81% accuracy across all three FV entries' ranking and selection approaches. The SVC model improved from 74% to 78% accuracy with RFE. The DT exhibited stable performance, whereas KNN accomplished a lower but consistent accuracy of 70%. LR has the lowest accuracy of 65%, indicating it performs poorly on nonlinear data, even after feature selection (Fig. 5).

In Figure 5, the RF model with FIRT outperformed all models in predicting MH risk. Secondly, the GB has run well. The least accuracy is shown by LR with all feature selection practices. These fallouts show that the RFE and ANOVA F-

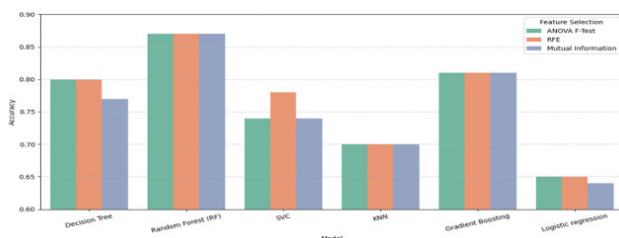


Fig. 5. Model accuracy comparison over feature selection methods.

TABLE VI
CLASS-WISE EVALUATION METRICS.

CLASS	PRECISION	RECALL	F1-SCORE
Low risk	0.89	0.92	0.90
Mid risk	0.85	0.82	0.83
High risk	0.84	0.80	0.82

Test are performing well for most of the models as compared to the MI. This shows that both the RFE and ANOVA F-test behaved excellently at identifying relevant features for MH risk prediction.

To develop a reliable baseline and assess the effectiveness of our breakdown, we focused on evaluating the performance of our best model, RF (accuracy 87%), against previously established benchmarks using similar MH datasets. For example, Assaduzzaman et al. [2] obtained a 90% accuracy using RF on a relevant MH dataset, and Togunwa et al. [17] reported 95% using a hybrid ANN-RF deep learning approach. Some results consider artificially optimistic metrics or suffer from a lack of generalizability due to dataset-centric constraints. Our study sheds light on existing models that exploit systematic feature selection (ANOVA F-test, RFE, MI) alongside comprehensive outlier exclusion and balanced metric evaluations.

The class-wise assessment shows that the framework performs well across all three chance levels, with the highest F1-score for low-risk (0.90) and steady performance for mid-risk (0.83) and high-risk (0.82), demonstrating solid generalization and calibrated expectation.

VI. CONCLUSION

This work proposes a novel methodology for breaking down the problem of identifying maternal health risks from a set of previously sampled indicators via AI/ML. Feature Importance Ranking and Trimming (FIRT) condenses the dimensionality of feature vector (FV) entries (variables) to reflect model correctness through ANOVA F-test, RFE, and MI. The new, shorter, more expressive FV boosts the identification of focal predictors, augments interpretability, lessens overfitting, and improves training efficiency. Key schemes for measuring FV entries' importance include (a) measuring the total node impurity reduction (e.g., by evaluating entropy) that a feature yields across all possibilities, where a higher total reduction signals greater appeal, (b) shuffling FV entries and measuring the breakdown in model accuracy, where a suggestive drop in performance points to a highly noteworthy feature, and

(c) gauging how much accuracy must be gained. For this purpose, this study used a Kaggle dataset from multiple sources to predict maternal health risks. In this study, feature selection and ranking occurred after data preprocessing. The outcomes indicate that among the three methods comprising FIRT, the ANOVA F-test and RFE accomplished better model performance, whereas mutual information (MI) did not. After training and testing the models, feature selection followed. This investigation's models are "K-nearest Neighbors" (KNN), "Random Forest" (RF), "Gradient Boosting" (GB), "Decision Tree" (DT), Support Vector Classifier (SVC), and "Logistic Regression" (LR). Among these, the RF attained the highest accuracy (87%) and the best overall performance. This shows that the random forest has been the best, most robust model for forecasting maternal well-being risks. To improve usefulness and effectiveness for future applications, additional feature-selection rationales can be applied to ML models. A large dataset can be collected across multiple states and countries to improve prediction accuracy. The work is limited by the lack of performance variance attributable to the single 80/20 train-test split used in the 1200-sample dataset, despite being greater than the ones used in other studies.

VII. FUTURE WORK

The FIRT strategy proposes a mental health risk analysis focused on qualitative interpretability and practical application in low-income clinical settings and edge computing. By prioritizing Manual Feature Engineering (HFE) over "black-box" Deep Learning models, the methodology utilizes domain knowledge to mitigate algorithmic biases and prevent hidden signals or surrogate variables (proxies), such as postal codes or hospital costs, from masking socioeconomic and racial disparities. While nominal accuracy may be lower than that of purely automated models, the design favors transparency and equity in healthcare, allowing specialists to focus on real clinical indicators [11].

Future efforts aim to refine the model's performance through k-fold cross-validation and aggregation techniques that maintain diagnostic confidence in remote areas. The application of Explainable AI (XAI) tools, such as SHAP and LIME, in SVC, ANN, and Random Forest models, demonstrates robust results in identifying risk zones, overcoming limitations of previous studies that lacked essential demographic data. Thus, HFE establishes itself as an ethical and crucial resource for reducing bias and improving fairness in high-risk decisions in health and humanitarian aid, transforming complex data into actionable and reliable predictions [11].

REFERENCES

- [1] Agrawal, S., Maitra, N.: Prediction of adverse maternal outcomes in preeclampsia using a risk prediction model. *The Journal of Obstetrics and Gynecology of India* **66**(Suppl 1), 104–111 (2016)
- [2] Assaduzzaman, M., Al Mamun, A., Hasan, M.Z.: Early prediction of maternal health risk factors using machine learning techniques. In: 2023 international conference for advancement in technology (ICONAT). pp. 1–6. IEEE (2023)
- [3] Bertini, A., Salas, R., Chabert, S., Sobrevia, L., Pardo, F.: Using machine learning to predict complications in pregnancy: a systematic review. *Frontiers in bioengineering and biotechnology* **9**, 780389 (2022)
- [4] Bogale, D.S., Abuhay, T.M., Dejene, B.E.: Predicting perinatal mortality based on maternal health status and health insurance service using homogeneous ensemble machine learning methods. *BMC Medical Informatics and Decision Making* **22**(1), 341 (2022)
- [5] Kang, B.S., Lee, S.U., Hong, S., Choi, S.K., Shin, J.E., Wie, J.H., Jo, Y.S., Kim, Y.H., Kil, K., Chung, Y.H., et al.: Prediction of gestational diabetes mellitus in asian women using machine learning algorithms. *Scientific Reports* **13**(1), 13356 (2023)
- [6] Kolukisa, B., Yavuz, L., Soran, A., Bakir-Gungor, B., Tuncer, D., Onen, A., Gungor, V.C.: Coronary artery disease diagnosis using optimized adaptive ensemble machine learning algorithm. *International Journal of Bioscience, Biochemistry and Bioinformatics* **10**(1), 58–65 (2020)
- [7] Mennickent, D., Rodriguez, A., Farias-Joffre, M., Araya, J., Guzmán-Gutiérrez, E.: Machine learning-based models for gestational diabetes mellitus prediction before 24–28 weeks of pregnancy: a review. *Artificial Intelligence in Medicine* **132**, 102378 (2022)
- [8] Mennickent, D., Rodríguez, A., Opazo, M.C., Riedel, C.A., Castro, E., Eriz-Salinas, A., Appel-Rubio, J., Aguayo, C., Damiano, A.E., Guzman-Gutierrez, E., et al.: Machine learning applied in maternal and fetal health: a narrative review focused on pregnancy diseases and complications. *Frontiers in Endocrinology* **14**, 1130139 (2023)
- [9] Myatra, S.N., Gangakhedkar, G.R., Savarkar, S.: Global health and women: Looking beyond maternal health. In: *Diversity, Gender Equity and Inclusion in Critical Care and Perioperative Medicine: A New Guidance for Team Excellence*, pp. 33–48. Springer (2025)
- [10] Organization, W.H.: Guide for conducting national and subnational programme reviews for maternal, newborn, child and adolescent health. World Health Organization (2024)
- [11] Patel, S.S.: Explainable machine learning models to analyse maternal health. *Data & Knowledge Engineering* **146**, 102198 (2023)
- [12] Pawar, L., Malhotra, J., Sharma, A., Arora, D., Vaidya, D.: A robust machine learning predictive model for maternal health risk. In: 2022 3rd international conference on electronics and sustainable communication systems (ICESC). pp. 882–888. IEEE (2022)
- [13] Raza, A., Siddiqui, H.U.R., Munir, K., Almutairi, M., Rustam, F., Ashraf, I.: Ensemble learning-based feature engineering to analyze maternal health during pregnancy and health risk prediction. *Plos one* **17**(11), e0276525 (2022)
- [14] Schaefer, V., Hilfiger-Kleiner, D., Keil, C.: Hypertension in pregnancy. *Cardiovascular Medicine* **26**(3), 80–86 (2023)
- [15] Singh, A., Prakash, N., Jain, A.: Particle swarm optimization-based random forest framework for the classification of chronic diseases. *IEEE Access* **11**, 133931–133946 (2023)
- [16] Sweeting, A.N., Appelblom, H., Ross, G.P., Wong, J., Kouru, H., Williams, P.F., Sairanen, M., Hyett, J.A.: First trimester prediction of gestational diabetes mellitus: a clinical model based on maternal demographic parameters. *Diabetes research and clinical practice* **127**, 44–50 (2017)
- [17] Togunwa, T.O., Babatunde, A.O., Abdullah, K.U.R.: Deep hybrid model for maternal health risk classification in pregnancy: synergy of ann and random forest. *Frontiers in Artificial Intelligence* **6**, 1213436 (2023)
- [18] Ukrit, M.F., Jeyavathana, R.B., Rani, A.L., Chandana, V.: Maternal health risk prediction with machine learning methods. In: 2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE). pp. 1–9. IEEE (2024)
- [19] Wei, Q., Xiao, Y., Yang, T., Chen, J., Chen, L., Wang, K., Zhang, J., Li, L., Jia, F., Wu, L., et al.: Predicting autism spectrum disorder using maternal risk factors: a multi-center machine learning study. *Psychiatry Research* **334**, 115789 (2024)