Classification of Maternal Health Risk Features Using Machine Learning Algorithms

Sunday Moses Benjamin
Faculty of Engineering Environment and Computing Coventry University
Coventry, England
sundaym@uni.coventry.ac.uk

Ademola Adenike Deborah
Faculty of Engineering Environment and Computing Coventry University
Coventry, England
ademolaa2@uni.coventry.ac.uk

Abstract—**This paper applies machine learning classification algorithms such as Decision Tree, Random Forest, K-Nearest Neighbors (KNN), Support Vector Machines (SVM) and Naïve Bayes algorithms to classify maternal health risk features. The models are evaluated based on performance measurements and the most effective algorithm is considered on factors such as accuracy, precision, recall, and f1-score.**

Keywords—machine learning; maternal health risk; k-nearest neighbors; algorithm; model; f1-score

I. Introduction

A. Background and Problem Statement

Maternal health is a critical component of global healthcare and requires accurate and timely predictions of health risks to safeguard the well-being of pregnant women, and children.

B. Research Question

This paper focuses on answering the question: "Which classification algorithm is most effective for predicting maternal health risks, considering not only accuracy and reliability but also social, ethical, and legal considerations?"

C. Aim and Objectives

The aim is to evaluate and compare various classification algorithms (Decision Tree, Random Forest, KNN, SVM and Naïve Bayes) for predicting maternal health risks. Objectives include:

- Implementation and comparison of the algorithms based on accuracy, precision, recall, and f1-score.
- Understanding the social, ethical, and legal implications of applying these models in maternal healthcare.

D. Research History

Machine Learning-based maternal health risk prediction model for IoMT framework (Subhash Mondal, 2023) utilized five ML classifiers, namely RF, DT, KNN, Logistic Regression, and Support Vector Machine, but the integration of these algorithms, with a focus on social, ethical, legal, and professional considerations is relatively unexplored. This paper contributes to the application of machine learning in maternal health, emphasizing responsible and effective technology use.

II. THE DATASET AND ETHICAL CONSIDERATIONS

The dataset used in this paper comes from data collected from different hospitals, community clinics, maternal health cares from the rural areas of Bangladesh through the IoT based risk monitoring system (UCI Machine Learning Repository, 2023). TABLE I. displays the summary of the data set describing name, role, type, units and whether any variable contains missing values.

TABLE I. VARIABLES TABLE

Variable Name	Role	Type	Units	Missing values
Age	Feature	Integer		no
SystolicP	Feature	Integer		no
DiastolicBP	Feature	Integer		no
BS	Feature	Integer	mmol/L	no
Body Temp	Feature	Integer	F	no
Heart rate	Feature	Integer	bpm	no
Risk Level	Target	Categorical		no

A. Variable names and Descriptions

Age: The age of the pregnant woman.
SystolicBP: Systolic blood pressure measured in mmHg.
DiastolicBP: Diastolic blood pressure measured in mmHg.
BS: Blood sugar level.
BodyTemp: Body temperature measured in degrees Fahrenheit.

HeartRate: Heart rate measured in beats per minute.
RiskLevel: The level of health risk classified into categories such as 'high risk', 'low risk', and 'mid risk' (UCI Machine Learning Repository, 2023)

B. *Social, Ethical, and Legal Considerations*

- Social Considerations: The use of machine learning in healthcare can lead to disparities in care, especially if the data lacks representation from diverse populations.
- Ethical Considerations: Issues related to consent, privacy, and the potential for misuse of data must be addressed.
- Legal Considerations: Compliance with healthcare regulations, such as HIPAA (Health Insurance Portability and Accountability Act) in the United States or GDPR (General Data Protection Regulation) in the European Union, is necessary. (Schroeder, 2009).

III. DATASET PREPARATION

A. *Class Imbalance*

Class imbalance happens when there are significantly lesser training examples in one class compared to other class (Aida Ali, 3013).

In this paper, the dataset has three classes with the following distribution:

TABLE II. CLASS DISTRIBUTION

CLASS	Name	Percentage (%)
0	high risk	26.82
1	low risk	40.04
2	mid risk	33.145

Although there is a slight imbalance, it is not severe. This is considered during model evaluation.

B. *Categorical Target Variable Encoding and Feature Normalization*

Machine Learning algorithms accept only numeric values as input. To use categorical data for Machine Learning purposes, the data needs to be encoded into numeric values such that each categorical feature is represented with a number.

Some categorical variable encoding techniques include:

- One Hot Coding: It compares each level of the categorical variable to a fixed reference level.
- Ordinal Coding: In ordinal encoding, an integer is assigned to each category, provided the number of existing categories is known.
- Sum Coding: Sum coding compares the mean of the dependent variable for a given level to the overall mean of the dependent variable over all the levels (Kedar Potdar, 2017).

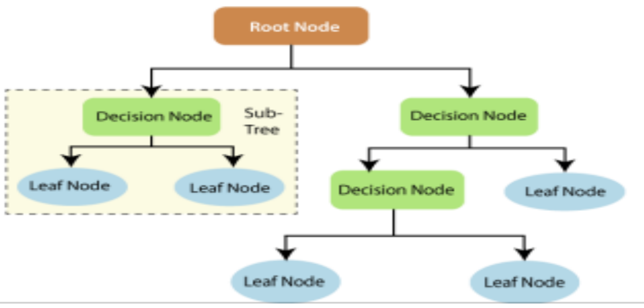
The categorical target variable in the dataset is standardized by removing the mean and scaling to unit variance and features are also normalized.

IV. CLASSIFICATION TECHNIQUES

A. *Decision Tree*

Decision tree is one of the most widely used classifiers in statistics and machine learning. Decision tree is a hierarchical design that implements the divide-and-conquer approach. It can be directly converted to a set of simple if-then rules. Fig. 1 illustrates the structure of a Decision Tree (Mohamed, 2017).

Fig. 1 Decision Tree



B. *Random Forest (RF)*

Random Forest is another classification technique based on decision tree; it is a collection of a group of tree predictors. Each tree depends on the values of a vector independently with the same distribution over all trees in the forest (Ashfaq Ahmed K, 2013).

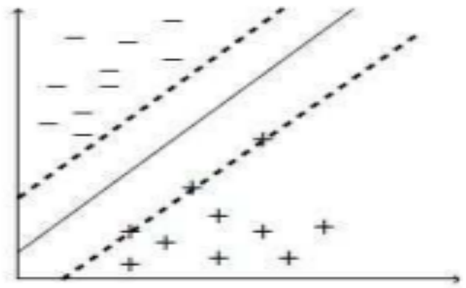
C. *K-Nearest Neighbour (KNN)*

K- Nearest-Neighbor is an example of instance-based leaning and it is often used for classification where the task is to classify the unseen examples based on the database stored. The observations are presented in a d-dimensional space, where d is the number of attributes or characteristics which the observation has (Mohamed, 2017).

D. *Support Vector Machines (SVM)*

SVM learning is one of many machine learning methods. Compared to other machine learning methods, SVM is very powerful at recognizing subtle patterns in complex datasets. SVM can be used to recognize handwriting, recognize fraudulent credit cards, and identify a speaker (Shujun Huang, 2018).

Fig. 2 Linearly Separable Classification



V. APPLICATION OF TECHNIQUES AND RESULTS

The dataset is split into train and test sets, with 80 percent used for training, while the remaining 20 percent is used for testing. After splitting, the training set contains 811 samples and 6 features, and the testing set contains 203 samples and 6 features.

An initial pass for training and prediction was carried out. A simple but straightforward loop structure is used to tune a single hyperparameter per classification method observed to show signs of possible overfitting or those with room for improvement.

The models were evaluated against:

- Accuracy
- Precision
- Recall
- F1-Score

A. *All Models Initial Pass*

Decision Tree, Random Forest, K-Nearest Neighbors, Support Vector Machines and Naïve Bayes classification algorithms were applied, with their default values and the result tabulated as shown in TABLE III.

TABLE III. SUMMARY OF PERFORMANCE METRICS FOR DIFFERENT CLASSIFICATION ALGORITHMS

Algorithm	Set	Accuracy	Precision	Recall	F1-Score
Decision Tree	Train	0.9358	0.9357	0.9358	0.9357
	Test	0.8177	0.8209	0.8177	0.8178
Random Forest	Train	0.9358	0.9363	0.9358	0.9358
	Test	0.8128	0.8185	0.8128	0.8133
K-Nearest Neighbors	Train	0.7620	0.7601	0.7620	0.7587
	Test	0.6354	0.6327	0.6354	0.6264
Support Vector Machines	Train	0.7274	0.7245	0.7274	0.7155
	Test	0.6798	0.7009	0.6798	0.6577
Naïve Bayes	Train	0.6091	0.5993	0.6091	0.5735
	Test	0.5763	0.5979	0.5763	0.5287

Decision Tree and Random Forest show the best performance among the algorithms, with high accuracy, precision, recall, and F1-score on both training and testing data.

K-Nearest Neighbors and Support Vector Machines have moderate performance, with better metrics in training than in testing, but lower accuracies.

Naive Bayes demonstrates the lowest performance across all metrics, indicating it may not be the best fit for this dataset or problem, and as a result was excluded from subsequent analysis.

B. *Decision Tree (DT)*

By varying the max_depth parameter and applying hyperparameter tuning to Decision Tree, the machine learning model was trained, and the results presented in Fig 3, Fig. 4, Fig. 5, and TABLE IV.

Fig. 3 DT Accuracy vs. Max Depth

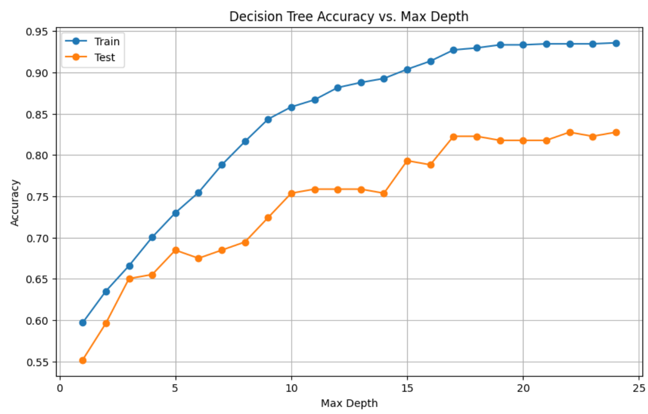


Fig. 4 DT Confusion Matrix

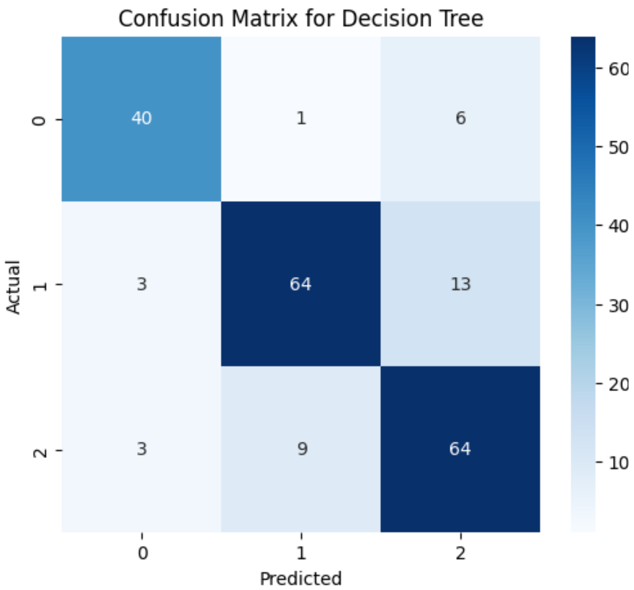


Fig. 5 Original Data vs. Predicted Data for DT

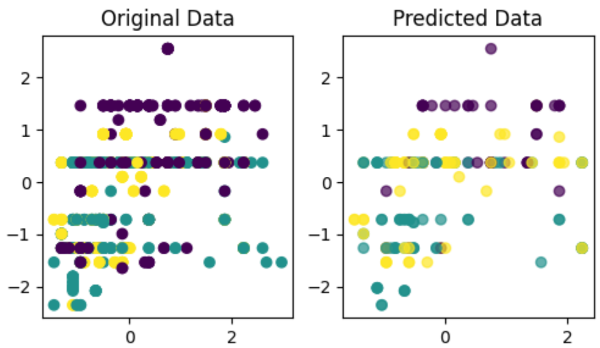


TABLE IV. DECISION TREE PERFORMANCE METRICS

Algorithm	Set	Accuracy	Precision	Recall	F1-Score
Decision Tree	Train	0.9346	0.9345	0.9346	0.9344
	Test	0.8177	0.8225	0.8177	0.8183

The results show that the hyperparameter tuning of the Decision Tree model resulted in marginal improvements in the model's ability to generalize to new data. While the improvements in accuracy, precision, recall, and F1-score on the test set are modest, they indicate that the tuning process made the model more effective at predicting unseen data without overfitting to the training data.

C. *Random Forest*

By varying the max_depth parameter and applying hyperparameter tuning to Random Forest, the machine learning model was trained, and the results presented in Fig. 6, Fig. 7, Fig. 8 and TABLE V.

Fig. 6 Random Forest Accuracy vs. Max Depth

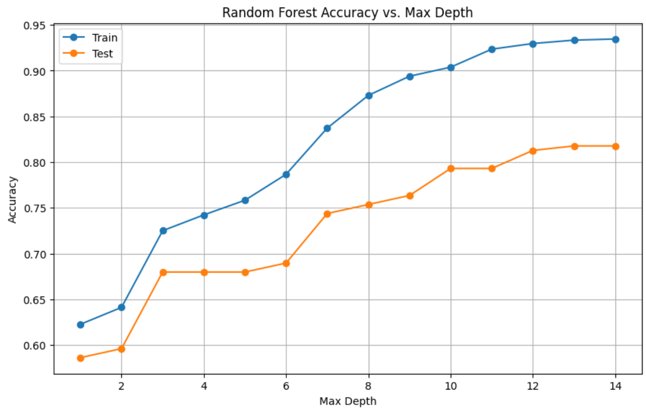


Fig. 7 RF Confusion Matrix

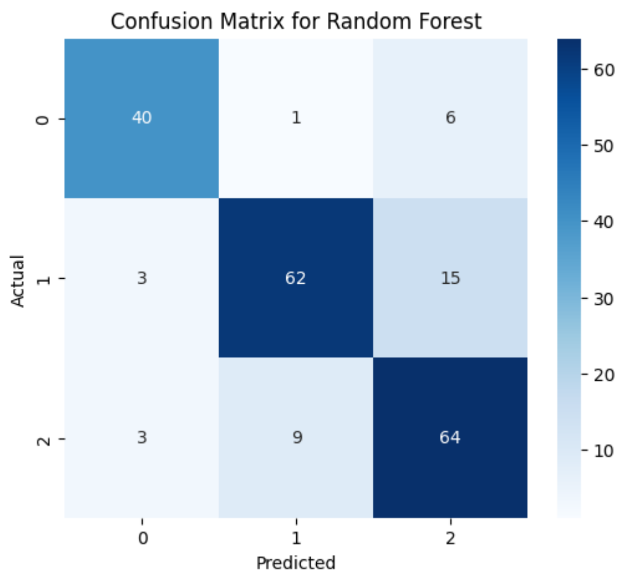


Fig. 8 Original Data vs. Predicted Data for RF

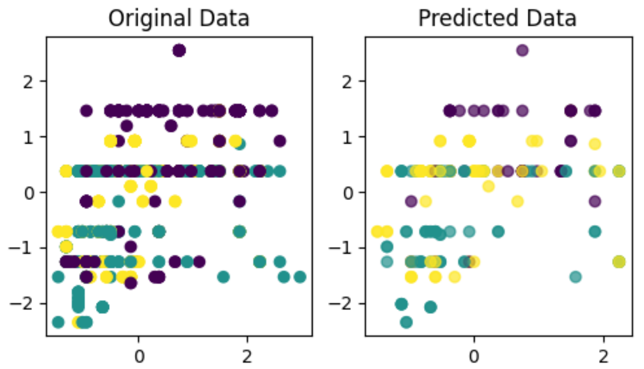


TABLE V. RANDOM FOREST PERFORMANCE METRICS

Algorithm	Set	Accuracy	Precision	Recall	F1-Score
Random Forest	Train	0.9346	0.9345	0.9346	0.9344
	Test	0.8177	0.8225	0.8177	0.8183

The results show that the hyperparameter tuning of the Random Forest model on the test data remains robust, indicating good generalization capabilities. This, therefore, suggests that tuning was successful in optimizing the model's performance while maintaining a balance between overfitting and underfitting.

D. *K-Nearest Neighbors(KNN)*

By varying the n_neighbors parameter and applying hyperparameter tuning to K-Nearest Neighbors, the machine learning model was trained, and the results presented in Fig. 9, Fig. 10, Fig. 11, and TABLE VI.

Fig. 9 K-Nearest Neighbors Accuracy vs. n_neighbors

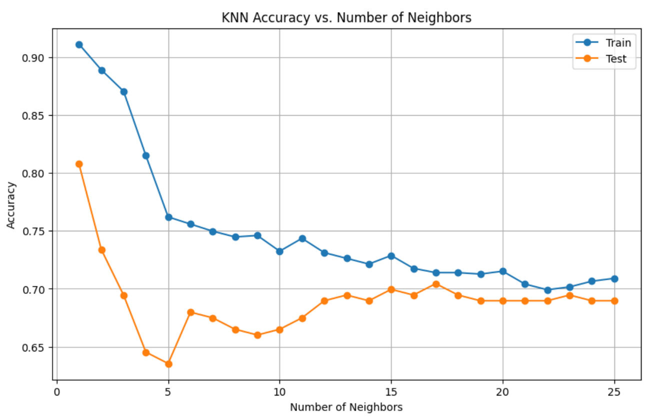


Fig. 10 KNN Confusion Matrix

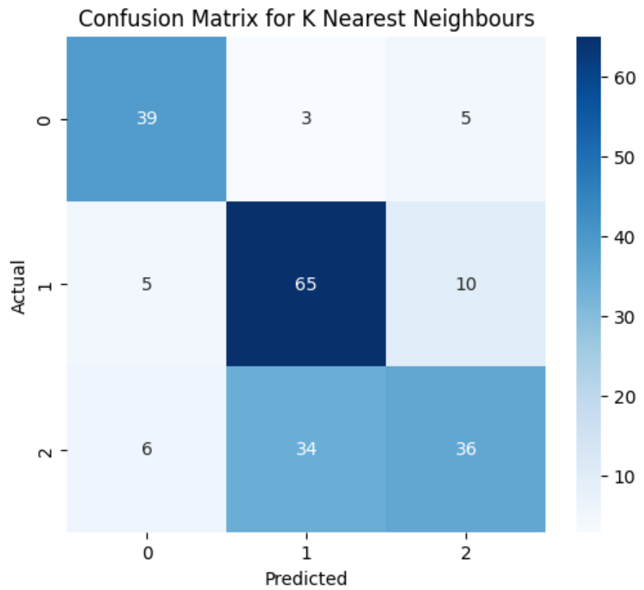


Fig. 11 Original Data vs. Predicted Data for KNN

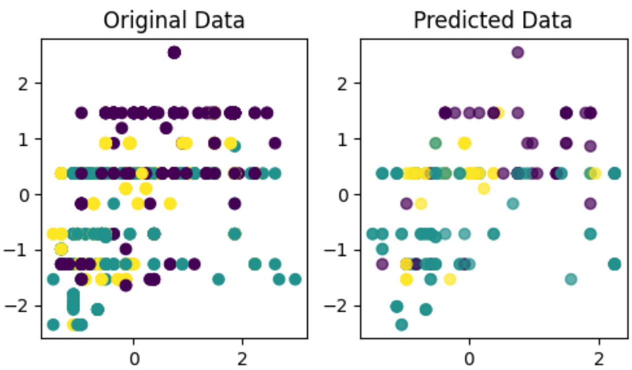


TABLE VI. KNN PERFORMANCE METRICS

Algorithm	Set	Accuracy	Precision	Recall	F1-Score
K-Nearest Neighbors	Train	0.6896	0.7051	0.7090	0.7038
	Test	0.6896	0.6959	0.6896	0.6799

The results show that the hyperparameter tuning of the KNN model demonstrates that while the model is reasonably effective at classifying instances, it's not particularly strong in any specific area (precision, recall, accuracy). The consistency of the metrics between training and testing suggests that the model is not overfitting much.

E. *Support Vector Machines (SVM)*

By varying the regularization parameter, C, and applying hyperparameter tuning to SVM, the machine learning model was trained, and the results presented in Fig. 12, Fig. 13, Fig. 14, and TABLE VII.

Fig. 12 SVM Accuracy vs C

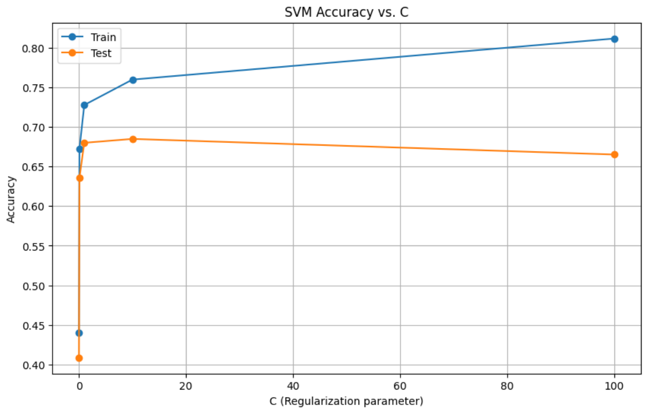


Fig. 13 SVM Confusion Matrix

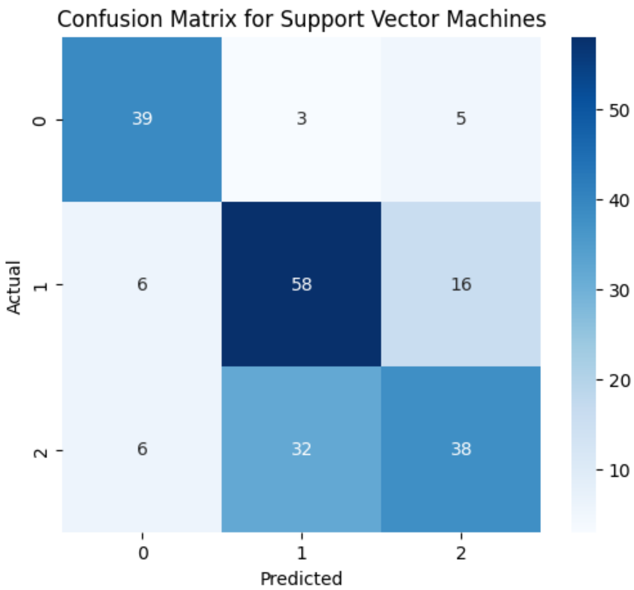


Fig. 14 Original Data vs. Predicted Data for SVM

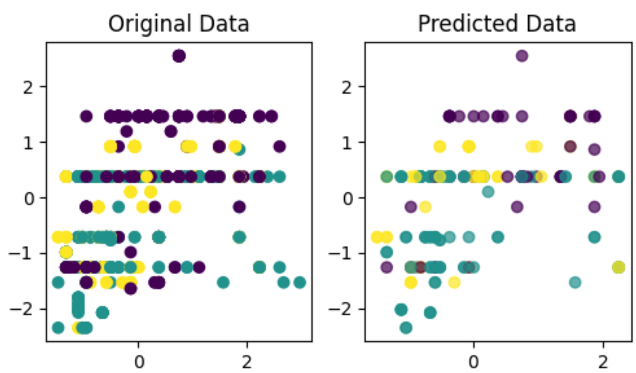


TABLE VII. SVM PERFORMANCE METRICS

Algorithm	Set	Accuracy	Precision	Recall	F1-Score
K-Nearest Neighbor	Train	0.6896	0.7051	0.7090	0.7038
	Test	0.6896	0.6959	0.6896	0.6799

The results show that the hyperparameter tuning of the SVM model shows that the tuning improved the model's fit to the training data but may not have adequately addressed the model's ability to generalize.

VI. CONCLUSION

Five (5) machine learning algorithms, namely: Decision Tree, Random Forest, K-Nearest Neighbor, Support Vector Machines, and Naïve Bayes, were applied to the dataset introduced in this paper, and hyperparameter tuning was subsequently applied to the models, except Naïve Bayes as it had the lowest performance across all metrics, in an initial model training and prediction run, indicating it may not be the best fit for the dataset.

Algorithm	DT	RF	KNN	SVM
Train Accuracy (Baseline)	0.935882	0.935882	0.762022	0.727497
Train Accuracy (Tuned)	0.935882	0.934649	0.709001	0.811344
Test Accuracy (Baseline)	0.817734	0.812808	0.635468	0.679803
Test Accuracy (Tuned)	0.827586	0.817734	0.689655	0.665025
Max Test Accuracy	0.827586	0.817734	0.807882	0.684729
Max Test Accuracy At	max_depth = 22	max_depth = 13	n_neighbours = 1	regularization param, c = 10.0

Hyperparameter tuning had varied effects on different algorithms. Based on results obtained, the best model to use for maternal health risk classification is Decision Tree, as it has the highest test accuracy (with risk of overfitting) for the dataset that had three classes in the target variable.

KNN shows less overfitting and significant improvement with tuning, making it a potentially more reliable choice in some scenarios.

VII. FUTURE RESEARCH

While the simple, loop structure, tuning improved various model's fit to the training data, it may not have adequately addressed the models' ability to generalize. A more exhaustive technique like gridsearch, combined with additional tuning, or data preprocessing and class balancing can lead to more generalization capabilities.

VIII. CONTRIBUTIONS

Ademola Adenike Deborah (StudentID:14446986) contributed to dataset procurement, data preparation and project report proof reading.

Sunday Moses Benjamin (Student ID: 14203482) contributed to dataset preprocessing, application of machine learning algorithms, and compilation of project report.

IX. APPENDIX

The Maternal Health Risk Dataset and python programming file is available at the following link: https://github.com/mosesbenjamin/classification-of-maternal-healthrisk-features.git

X. REFERENCES

(2023). Retrieved from UCI Machine Learning Repository: https://archive.ics.uci.edu/dataset/863/maternal+health+risk

Aida Ali, S. M. (3013). Classification with class imbalance problem: a review. *Int. J. Advance Soft Compu. Appl*, 31.

Ashfaq Ahmed K, S. A. (2013). Comparative Prediction Performance with Support Vector Machine and Random Forest Classification Techniques. *International Journal of Computer Applications*, 5.

Kedar Potdar, T. S. (2017). A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers. *International Journal of Computer Applications*, 4.

Mohamed, A. E. (2017). Comparative Study of Four Supervised Machine Learning Techniques for Classification . *International Journal of Applied Science and Technology* , 14.

Schroeder, R. (2009). *Social, Ethical, and Legal Issues in Presence Research and Applications.*

Shujun Huang, N. C. (2018, 1 15). *Cancer Genomics & Proteomics.* Retrieved from https://cgp.iiarjournals.org/content/15/1/41.short

Subhash Mondal, A. N. (2023). Machine learning-based maternal health risk prediction model for IoMT framework. *International Journal of Experimental Research and Review (IJERR)*, 15.