Proceedings of the 2nd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/74/20240456

Prediction of maternal health risk levels based on multiple machine learning algorithms

Xu Yang

School of Information Engineering, Chongqing Jianzhu College, Chongqing, 400072, China

yangxu012@126.com

Abstract. With the development and popularisation of the Internet of Things (IoT) technology, more and more medical devices have begun to connect with the Internet, forming IoT medical devices. Among them, for pregnant women's health risk monitoring, IoT devices can achieve remote monitoring, data transmission, intelligent analysis and other functions to provide more comprehensive health management services for pregnant women. To evaluate the model's effectiveness in predicting the health risk level of pregnant women, we use four metrics: precision, accuracy, recall, and F1 score. Where precision indicates the proportion of correctly predicted samples to the total number of samples; accuracy indicates the proportion of true cases to the number of all samples predicted to be positive; recall indicates the proportion of true cases to the number of all samples that are actually positive; and F1 score is the reconciled average of precision and recall. By using these metrics, we can more fully assess the predictive effectiveness of the model. In this paper, we use three models, decision tree, random forest and logistic regression, for training and testing. All three models are common machine learning algorithms, but they each have different strengths and weaknesses. By training and testing the three models, we found that the random forest model performed the best, with a prediction accuracy of 80%. In conclusion, this paper describes an IoT-based health risk monitoring system for pregnant women, trained and tested using three models: decision tree, random forest and logistic regression, it can provide more comprehensive health management services for pregnant women.

Keywords: Maternal health, Risk levels, Decision tree, Machine learning algorithms.

1. Introduction

With the development and popularisation of the Internet of Things (IoT) technology, more and more medical devices have begun to connect with the Internet, forming IoT medical devices [1]. Among them, for pregnant women's health risk monitoring, IoT devices can achieve remote monitoring, data transmission, intelligent analysis and other functions to provide more comprehensive health management services for pregnant women [2]. Meanwhile, machine learning technology also has an important role in this regard [3].

The use of IoT devices for health risk monitoring of pregnant women is of great significance. On the one hand, pregnant women need to undergo various tests and observations during pregnancy to ensure the health of the foetus [4]. However, under the traditional healthcare model, pregnant women need to travel frequently to hospitals for checkups, which is not only time-consuming and labour-intensive but

© 2024 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/74/20240456

also prone to traffic congestion and gathering of people [5]. In contrast, remote monitoring through IoT devices allows pregnant women to obtain professional health management services at home [6]. On the other hand, IoT devices can monitor pregnant women's physiological indicators, exercise, diet and other information in real time, and transmit these data to the cloud for analysis and processing [7]. This is very beneficial for both doctors and pregnant women themselves to better understand the health status of pregnant women, detect abnormalities in time and take appropriate measures [8,9].

Machine learning techniques also have an important role in health risk monitoring of pregnant women by IoT devices. Machine learning is an artificial intelligence technology that can identify relationships between data and predict future trends and outcomes by analysing and learning from large amounts of data [10]. In the context of pregnant women's health risk monitoring, machine learning techniques can help doctors better understand the health status of pregnant women and provide more accurate diagnosis and treatment plans.

In conclusion, health risk monitoring of pregnant women using IoT devices is of great importance. Meanwhile, machine learning techniques have an irreplaceable role in this regard. In this paper, we use various machine learning algorithms to predict the health status of pregnant women based on the physiological state information of pregnant women obtained from IoT devices, and compare the prediction effect of various machine learning algorithms to provide a certain foundation for subsequent research.

2. Data set introduction

Data were collected from different hospitals, community clinics, and maternal health facilities through an IoT-based risk monitoring system. The data included age, systolic blood pressure, diastolic blood pressure, BS, heart rate, and risk level of the pregnant women, with age referring to the age of the woman at which she became pregnant; systolic blood pressure refers to the upper limit of the blood pressure in mmHg; diastolic blood pressure refers to the lower limit of the blood pressure value in mmHg; BS refers to the level of glucose, expressed as a molar concentration; and heart rate refers to the normal resting heart rate, expressed in number of heartbeats per minute. The target variable was the health risk level of the pregnant women, and the data for the section are shown in Table 1.

Table 1. Selected datasets.

Age	Systolic BP	Diastolic BP	BS	Body Temp	Heart Rate	Risk Level
25	130	80	15	98	86	high risk
35	140	90	13	98	70	high risk
29	90	70	8	100	80	high risk
30	140	85	7	98	70	high risk
35	120	60	6.1	98	76	low risk

The target variable (health risk level of pregnant women) was classified into three categories: low, medium and high risk, which were statistically analysed in histograms and bar charts, and the results are shown in Figures 1 and 2.

DOI: 10.54254/2755-2721/74/20240456

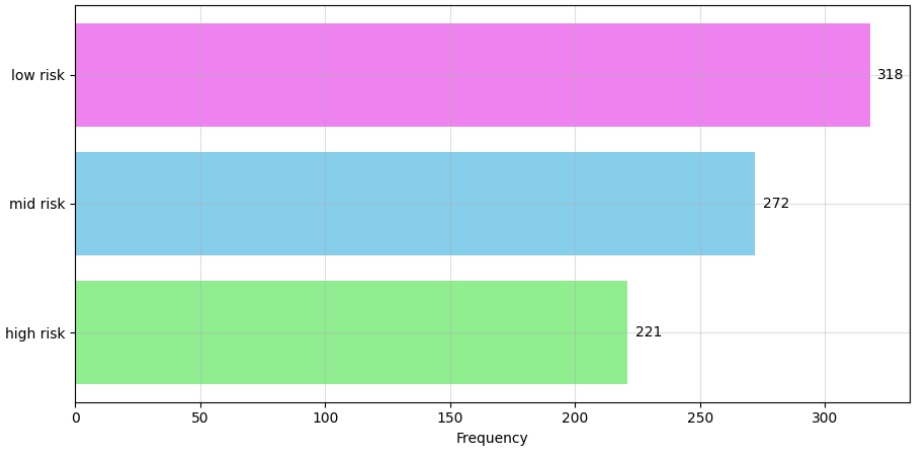

Figure 1. Bar chart.
(Photo credit: Original)

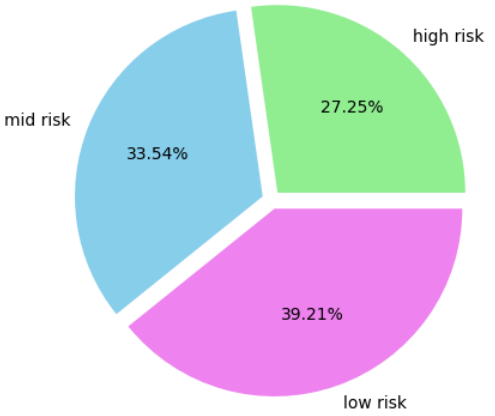

Figure 2. Pie chart.
(Photo credit: Original)

As can be seen from the figure, the highest number of health risk levels for pregnant women in the dataset is low risk, followed by medium risk, with the least amount of data for high risk.

3. Machine Learning Algorithms

3.1. Decision Tree Model

A decision tree is a model for decision making based on a tree structure. It is formed by dividing the dataset into multiple subsets, each of which corresponds to a node, eventually forming a tree structure. In the process of decision tree construction, it is necessary to choose appropriate attributes as the basis for node division, so that the samples within each subset are as similar as possible, and the differences between different subsets are as large as possible. This division can be evaluated and selected by indicators such as information gain and Gini coefficient. In the specific implementation, the root node needs to be determined first, and then the optimal feature is found as the first node to be divided according to the selected division index. Next, divide the dataset according to the value of this feature and create a new node for each subset. For each new node, the above steps are repeated until certain stopping conditions are met. Decision tree models have the advantages of easy interpretation,

DOI: 10.54254/2755-2721/74/20240456

visualisation and fast prediction. However, they are prone to overfitting problems when dealing with high-dimensional data and complex relationships, and are sensitive to outliers.

3.2. Random forest

Random Forest is an integrated learning method based on decision trees. It creates multiple decision trees by randomly selecting samples and features and combines them to make predictions. Random forest can effectively reduce the overfitting problem and has high accuracy and stability. In random forest, each decision tree is constructed independently, so it can be computed in parallel to improve efficiency. At the same time, selecting a part of the features to be divided at each node can also effectively reduce the dimensionality and improve the generalisation ability of the model. Random forest can help to understand the data and model by feature importance assessment.

3.3. Logistic regression

Logistic regression is a linear model widely used in classification problems. It multiplies the input features with weights and adds a bias term, and then maps the result to between [0,1] by a sigmoid function indicating the probability that the sample belongs to a positive class. During the training process, the loss function (e.g. cross entropy) needs to be minimised, and the weights and bias terms are updated by optimisation algorithms such as gradient descent. Logistic regression has the advantages of simplicity, computational speed and interpretability, and has been widely used in binary classification problems. Logistic regression controls the complexity of the model by adding regularisation terms and supports features such as online learning and incremental learning, making it perform well in large-scale data processing and real-time prediction.

4. Experiments and results

The training set, validation set and test set were divided according to the ratio of 6:2:2, and the three models of decision tree, random forest and logistic regression were used for training respectively, the validation set was used to validate the training process, and the test set was used for testing the model at the end of the training process, and the model's effectiveness in predicting the health risk class of pregnant women was evaluated using the precision, the accuracy, the recall, and the F1 scores, and the results are shown in Table 2 and Fig. 3. Shown.

Table 2. Modelling evaluation.

	Decision Tree	Random Forest	Logistic Regression
Accuracy	0.78	0.80	0.66
Precision	0.78	0.81	0.65
Recall	0.78	0.80	0.66
F1 Score	0.78	0.80	0.65

DOI: 10.54254/2755-2721/74/20240456

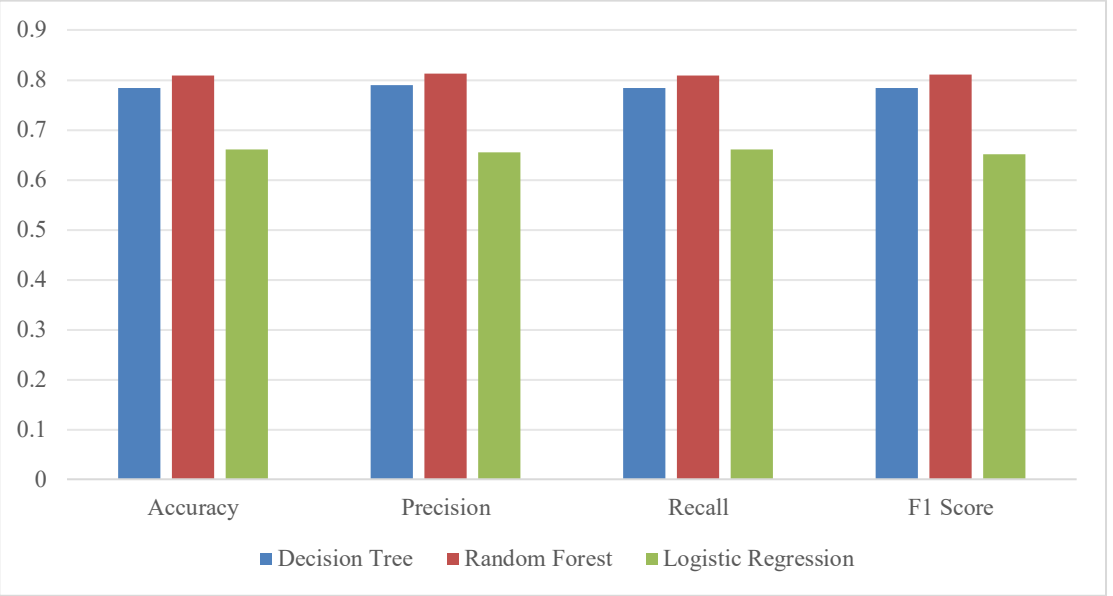

Figure 3. Modelling evaluation.
(Photo credit: Original)

From the results, it can be seen that the best performing machine learning algorithm model in predicting the health risk level of pregnant women based on IoT devices is the Random Forest model with 80% accuracy of prediction, also the Decision Tree algorithm is better with 78%, and the logistic regression model is worse with 66%.

5. Conclusion

In this paper, we have used three models, Decision Tree, Random Forest and Logistic Regression, for training and testing. All three models are common machine learning algorithms, but they each have different advantages and disadvantages. The decision tree model has the advantages of easy interpretation, visualisation and fast prediction, but is prone to overfitting problems when dealing with high-dimensional data and complex relationships, and is sensitive to outliers. The random forest model is an integrated learning method based on decision trees, which can effectively reduce the overfitting problem and has high accuracy and stability. The logistic regression model has the advantages of simple computation, easy interpretation and high interpretability, and it is also faster in dealing with large-scale data, but it is weak in fitting non-linear relationships.

By training and testing the three models, we found that the random forest model performed the best, with a prediction accuracy of 80%. This is related to the fact that the random forest model can effectively reduce the overfitting problem and has high accuracy and stability. In addition, the decision tree model also had a better prediction effect of 78%, which is related to the advantages of the decision tree model such as easy interpretation, visualisation and fast prediction. The logistic regression model, on the other hand, has a poorer prediction effect of 66%, which may be due to the weak fitting ability of the logistic regression model for non-linear relationships.

In conclusion, this paper describes an IoT-based health risk monitoring system for pregnant women, trained and tested using three models: decision tree, random forest and logistic regression. By evaluating the metrics of precision, accuracy, recall and F1 score, we found that the random forest model performs the best and can provide more comprehensive health management services for pregnant women.

In conclusion, health risk monitoring of pregnant women using IoT devices is of great importance. Meanwhile, machine learning techniques have an irreplaceable role in this regard. In this paper, we use various machine learning algorithms to predict the health status of pregnant women based on the physiological state information of pregnant women obtained from IoT devices, and compare the

DOI: 10.54254/2755-2721/74/20240456

prediction effect of various machine learning algorithms to provide a certain foundation for subsequent research.

References

[1] J. K M ,K. A B ,E. C W , et al.Exploring the accuracy of self-reported maternal and newborn care in select studies from low and middle-income country settings: do respondent and facility characteristics affect measurement?[J].BMC Pregnancy and Childbirth,2023,23(1):448-448.

[2] Kokou R D ,Hubert A ,Ishaque F S , et al.Increased investment in Universal Health Coverage in Sub-Saharan Africa is crucial to attain the Sustainable Development Goal 3 targets on maternal and child health.[J].Archives of public health = Archives belges de sante publique,2023,81(1):34-34.

[3] R R W ,Sushant J ,V. G S .Health Monitoring Devices System and Technology[J].Electrochemical Society Transactions,2022,107(1):

[4] Fatemeh S ,Iman A ,Sina L , et al.Long-Term IoT-Based Maternal Monitoring: System Design and Evaluation[J].Sensors,2021,21(7):2281-2281.

[5] Fatemeh S ,Iman A ,Sina L , et al.Long-Term IoT-Based Maternal Monitoring: System Design and Evaluation[J].Sensors,2021,21(7):2281-2281.

[6] J P S ,Ismael M ,Estephania M , et al.Indicators for monitoring maternal and neonatal quality care: a systematic review.[J].BMC pregnancy and childbirth,2019,19(1):25.

[7] Li J ,Bai X ,Xu Q , et al.Identification of Customer Churn Considering Difficult Case Mining[J].Systems,2023,11(7):

[8] Lakshmi B ,Indumathi T ,Ravi N .An Hybrid Approach for Prediction Based Health Monitoring in Pregnant Women[J].Procedia Technology,2016,24(C):1635-1642.

[9] Islam R M .Safe Motherhood Promotion in Bangladesh: Evidence from a NGO's Local Level Health Monitoring and Advocacy Project[J].Journal of Family Medicine and Community Health,2015,2(2):

[10] Brinkel J ,Krämer A ,Krumkamp R , et al.Mobile Phone-Based mHealth Approaches for Public Health Surveillance in Sub-Saharan Africa: A Systematic Review[J].International Journal of Environmental Research and Public Health,2014,11(11):11559-11582.