Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2025. Vol. 4. No. 2; e-ISSN: 2808-4519

Classification of Fetal Health Using the K-Nearest Neighbor Method and the Relieff Feature Selection Method

Anita[1]*, Asrul Abdullah [2], Syarifah Putri Agustini Alkadri [3]

[1,2,3]Fakultas Teknik Universitas Muhammadiyah Pontianak
anitaant18@gmail.com[1],asrul.abdullah@unmuhpnk.ac.id[2], agustini.putri@unmuhpnk.ac.id[3]*

Abstract

Understanding fetal health early can reduce risks to the pregnancy and the womb. Identifying correlations among factors influencing fetal well-being helps medical professionals clarify key impacts. Quantified relationships between features and labels also guide future research. This study focuses on three aspects: evaluating KNN model performance with and without ReliefF feature selection, analyzing the impact of feature removal, and assessing ReliefF's ability to identify critical features for fetal health classification.The research begins by framing fetal health classification as a supervised machine learning task using labeled datasets. A cardiotocographic dataset from the UCI Machine Learning Repository supports data collection. Initial analysis identifies data types and detects outliers, followed by preprocessing, feature selection, and KNN model training. Model testing uses metrics such as accuracy and recall. Results show the KNN model with ReliefF features achieves an accuracy of 0.896. Testing a pruned model by removing high-importance features slightly reduces accuracy to 0.866. These findings confirm ReliefF's effectiveness in identifying essential features and optimizing model efficiency without compromising quality. This study underscores ReliefF's role in improving KNN performance for fetal health classification.

***Keywords**: Fetal Health Classification, Feature Selection ReliefF, Machine learning, Train-test split, K-Nearest Neighbors*

1. Introduction

Fetal health monitoring is a crucial aspect of prenatal care to ensure a healthy pregnancy and the safety of both mother and baby. Understanding fetal development and early detection of potential health issues can help reduce perinatal mortality. Medical technologies like CTG are used to collect detailed data on fetal health, consisting of 21 features and 3 labels: normal, suspect, and pathologic. Identifying critical features within this data can be achieved using feature selection methods such as ReliefF, which has proven to enhance model effectiveness. Computerized CTG has been shown to significantly reduce perinatal mortality compared to traditional CTG [1].

Fetal health classification is essential for monitoring development and identifying potential issues early. With proper classification, healthcare providers can offer more effective treatments and reduce pregnancy risks [2]. Since ReliefF relies on nearest neighbor calculations like KNN, KNN is the machine learning model used in this research. The K-NN method has been validated in previous studies for fetal health classification [3].

When data contains numerous attributes, feature selection becomes critical. ReliefF assists in identifying the most relevant attributes, as supported by prior studies demonstrating improved model performance [4]. ReliefF explains feature-label relationships, prioritizes important features, and enhances model training efficiency. This study, titled “Fetal Health Classification Using K-Nearest Neighbor and ReliefF Feature Selection Methods,” aims to develop a classification model using KNN, evaluate the impact of ReliefF on model performance, and determine the optimal *k* value with the Elbow method. The research contributes to improving classification accuracy, reducing pregnancy risks, and supporting better medical decision-making.

2. Research Method

The research steps consist of dataset collection, dataset analysis, data pre-processing, feature selection, determining the K value, classification with conventional KNN, classification with KNN and ReliefF, and classification testing 4o

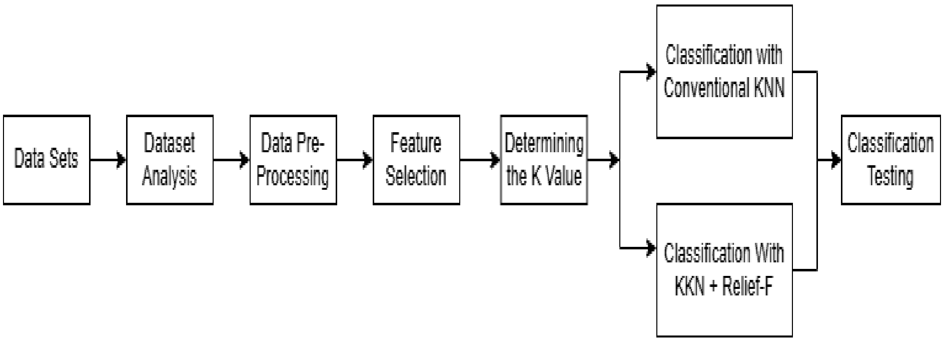

Fig. 1: Research Methods

2.1. Identification of Problems

Fetal health classification falls under supervised learning in machine learning, as the dataset used contains labels. This study aims to examine the impact of feature selection on model performance with a dataset containing numerous features.

2.2. Data Collection

This stage involves collecting the dataset for fetal health classification. Data collection is a systematic process to extract patterns using machine learning, statistics, and databases [5]. The structured dataset, with a target variable, falls under a supervised learning classification task. The data is sourced from the UCI Machine Learning Repository and is based on the cardiotocography dataset.

2.3. Implementation Model

The best-performing model from the testing phase will be deployed on a web interface using Streamlit and Python after being saved with the Pickle module.

3. Result and discussion

3.1. Data Collection Results

The dataset used in this study is a public dataset from UCI Machine Learning, called 'Cardiotocography,' donated on 09/06/2010. It contains 2,126 records with 22 features, including fetal heart rate (FHR) and uterine contraction (UC) measurements classified by obstetricians [6]. The dataset has multi-class labels: 'Normal,' 'Suspect,' and 'Pathologic.' Detailed descriptions of the features and labels are provided

3.2. Data Pre-Processing Results

The data preprocessing results include the handling of outliers in each feature, based on previous outlier analysis. The outlier handling results are shown in Figures 1 and 2 for the "percentage of time" feature.

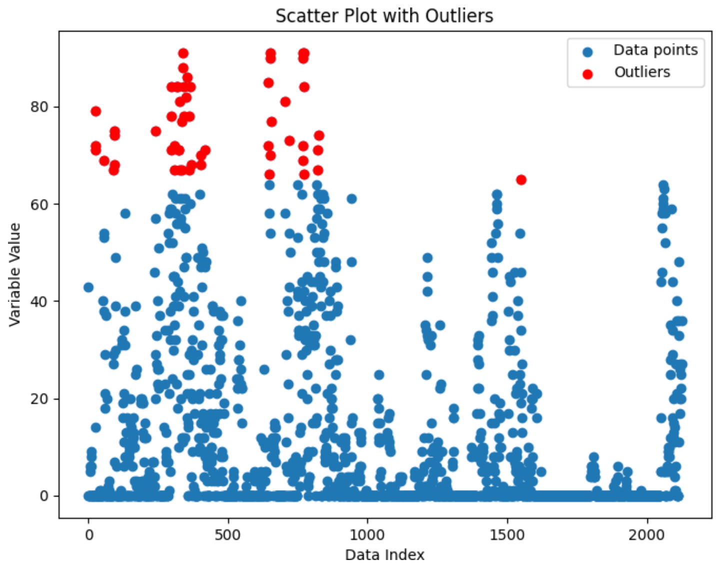

Fig. 2: Outliers in Data after Handling Outliers

Not many outlier data points are handled because what is handled is only outlier data points with the class label 'Normal/1.0'.

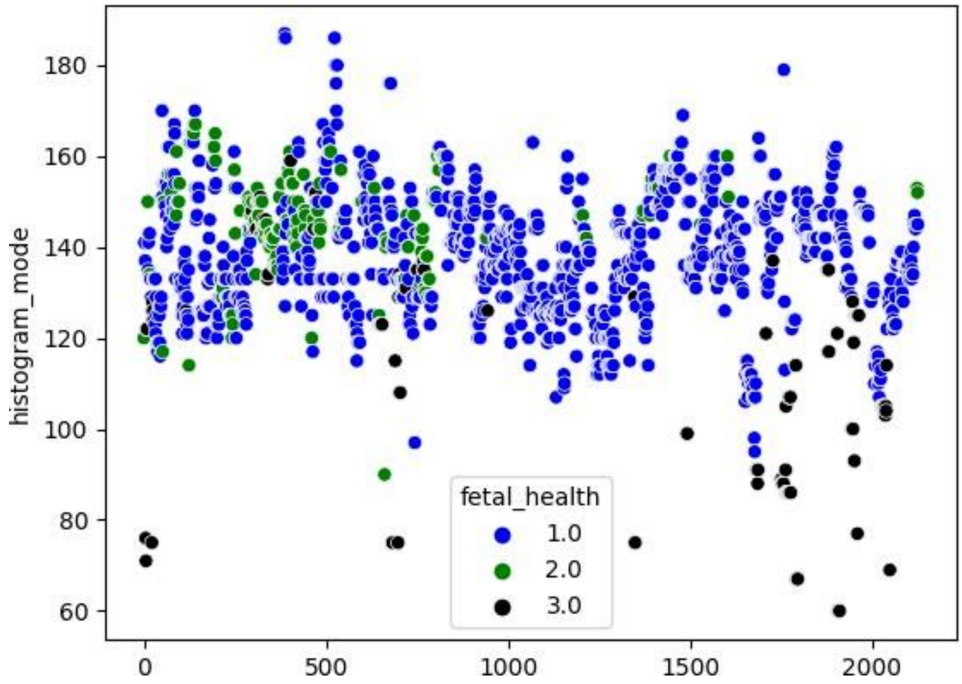

Fig. 3: Scatter Plot Histogram Mode after Handling Outlier

Features with outliers show a similar pattern, where most outliers are labeled as 'pathologic/3.0.' Since information on 'pathologic' cases is crucial, outlier handling is limited to data points labeled as 'Normal' to maintain the relevance of critical information.

3.3. Model Training Results

The K-Nearest Neighbor (KNN) model is used to classify fetal health, with the K value being a key hyperparameter. The optimal K value is determined using the Elbow method, as shown in Figure 4

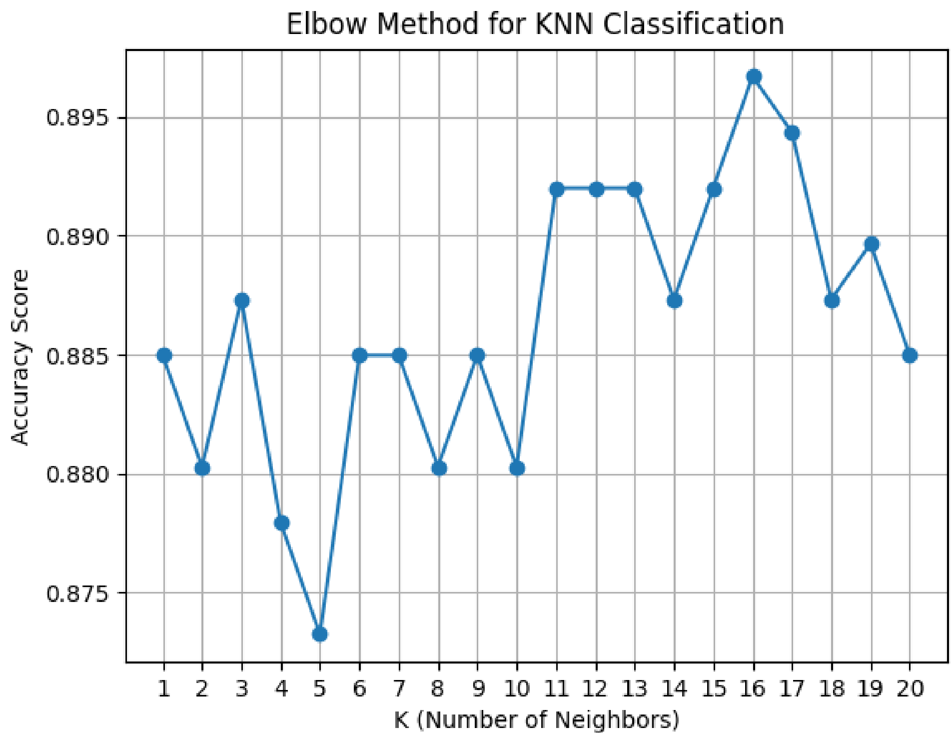

Fig. 4: Determining the K Value using the Elbow Method

Based on the elbow plot in Figure 5.11, the optimal K value for the KNN model is K = 16. Accuracy results with and without the optimal K are shown in Table 1

Tabel 1: Results of Determining the K-Value in the Model

Proses	Akurasi	Recall
Sebelum Penentuan Nilai-K	0.873239436	0.873239436
Setelah Penentuan Nilai-K dengan K=16	0.896713615	0.896713615

4. Conclusion

The KNN model successfully predicts fetal health, and the ReliefF feature selection method enhances its performance effectively. KNN and ReliefF were evaluated using cross-validation and train-test split, comparing datasets with and without feature selection. Metrics such as accuracy and recall were used, supported by confusion matrices to analyze model behavior. The KNN model achieved an accuracy of 0.835 with cross-validation and 0.896 with train-test split using Relieff.

The top feature, "percentage of time with abnormal long-term variability," positively influenced the model, as removing it reduced accuracy from 0.896 to 0.866. The removal of seven features using ReliefF had minimal impact, reducing cross-validation accuracy by only 0.001, while improving the model's training and prediction speed.

Acknowledgement

The author would like to thank the Faculty of Information and Computer Engineering, Muhammadiyah University of Pontianak for supporting this work.

References

[1] N. Rahmayanti, H. Pradani, M. Pahlawan, and R. Vinarti, “Comparison of machine learning algorithms to classify fetal health using cardiotocogram data,” *Procedia Comput. Sci.*, vol. 197, no. 2021, pp. 162–171, 2021, doi: 10.1016/j.procs.2021.12.130

[2] D. I. Annisa, R. Ariyanto, and A. T. R. Hayati Ririd, “Klasifikasi Kehamilan Beresiko Dengan Menggunakan Metode K-Nearest Neighbor (Studi Kasus Dinas Kesehatan Kabupaten Malang),” *J. Inform. Polinema*, vol. 3, no. 1, p. 34, 2016, doi: 10.33795/jip.v3i1.20

[3] M. T. Alam *et al.*, “Comparative Analysis of Different Efficient Machine Learning Methods for Fetal Health Classification,” *Appl. Bionics Biomech.*, vol. 2022, 2022, doi: 10.1155/2022/6321884.

[4] R. N. Yusra, O. S. Sitompul, and Sawaluddin, “Kombinasi K-Nearest Neighbor (KNN) dan ReliefF Untuk Meningkatkan Akurasi Pada Klasifikasi Data,” *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 1, pp. 0–5,2021.

[5] Alexander Zubchenko, “Data Collection For Machine Learning: The Compelete Guide,” *WAVERLEY*, 2021.

[6] D. Campos and J. Bernardes, “Cardiotocography.” 2010. [Online]. Available: https://doi.org/10.24432/C51