

Cardiotocographic Classification using Feature Selection: A comparative Study

Septian Eko Prasetyo, Pulung Hendro Prastyo, Shindy Arti

Universitas Gadjah Mada, Yogyakarta, Indonesia

ARTICLE INFORMATION

Received: January 30th, 2021
 Revised: March 9th, 2021
 Available online: March 31st, 2021

KEYWORDS

Cardiotocography, Classification, Feature Selection, Machine Learning, Data Mining

CORRESPONDENCE

E-mail: septianepe@mail.ugm.ac.id

A B S T R A C T

Cardiotocography is a series of inspections to determine the health of the fetus in pregnancy. The inspection process is carried out by recording the baby's heart rate information whether in a healthy condition or contrarily. In addition, uterine contractions are also used to determine the health condition of the fetus. Fetal health is classified into 3 conditions namely normal, suspect, and pathological. This paper was performed to compare a classification algorithm for diagnosing the result of the cardiotocographic inspection. An experimental scheme is performed using feature selection and not using it. CFS Subset Evaluation, Info Gain, and Chi-Square are used to select the best feature which correlated to each other. The data set was obtained from the UCI Machine Learning repository available freely. To find out the performance of the classification algorithm, this study uses an evaluation matrix of precision, Recall, F-Measure, MCC, ROC, PRC, and Accuracy. The results showed that all algorithms can provide fairly good classification. However, the combination of the Random Forest algorithm and the Info Gain Feature Selection gives the best results with an accuracy of 93.74%.

INTRODUCTION

The technology in biomedical engineering has been helped to solve health and medical problem [1]. One of them applied to diagnose the condition of the fetus during pregnancy. At present, the practice of monitoring fetal health can be performed through inspection called cardiotocography [2]. The inspection records information about the fetal heart rate (FHR) and uterine contractions (UC) during pregnancy [3]. This monitoring is used by obstetricians and midwives to avoid hypoxic-ischemic conditions of the fetus during pregnancy and giving birth [4]. Hypoxic-ischemic is a brain injury due to a blockage of blood vessels in the prenatal, intrapartum, or postnatal period [5]. The results of cardiotocographic diagnosis are classified into three classes, namely normal, suspect, and pathologic [6]. A suspect condition occurs when a deficiency is at least one normal characteristic. Whereas the pathologic is a condition where the baby's health is impaired.

In previous studies, several classification algorithms have been used to classify the cardiotocographic detection including Neural Network [7], [8], Naïve Bayes Classifier [9], [10], Support Vector Machine [11], K-Means [12], Random Forest [13], K-Nearest Neighbors [14], and J48 [15]. Based on several approaches that have been performed, Neural Network gets superior accuracy compared to other classification models. In his research, Zafer CÖMERT et al [8] explained that training conducted on Artificial Neural Networks (ANN) can optimize the training process well.

Therefore, this model gets a fairly high accuracy compared to others.

Moreover, other research was also conducted to combine several classification models to find out the best performance results. In one hand, Subhan et al tested the performance of Naïve Bayes, Decision Tree, Multi-Layer Perceptron (MLP), and Radial Basis Function algorithms in classifying cardiotocography. The results show that the Decision Tree gets the best accuracy at 93.3% [16]. On the other hand, Kanika Agrawal et al tried to compare the classification results using Decision Tree, Support Vector Machine, and Naïve Bayes Classifier which were analyzed using R-Studio [17]. The results show that the classification model gets the highest accuracy in the Decision Tree algorithm as same as the research conducted by Subha et al in previous study [16].

Based on several studies and backgrounds that have been explained, in this research, a classification method will be compared using several approaches and scenarios. This study uses two experimental scenarios, the first one is a comparison of models using feature selection and the second one is the comparison of classification model without using feature selection. Thus, in this experimental result can be seen which model gets the best performance. On the other hand, combination models can be used with the best feature selection in the classification process. Thus, The combination model can improve the accuracy of results.

METHOD

This section will discuss research methods, datasets, and algorithm models used for the classification process. The research method used in this study is a comparative model in the classification of cardiocography. The dataset is used for training and testing process in the algorithm applied. While the model used in this study consists of several algorithms that will be performed for the comparison of results.

Research Methodology

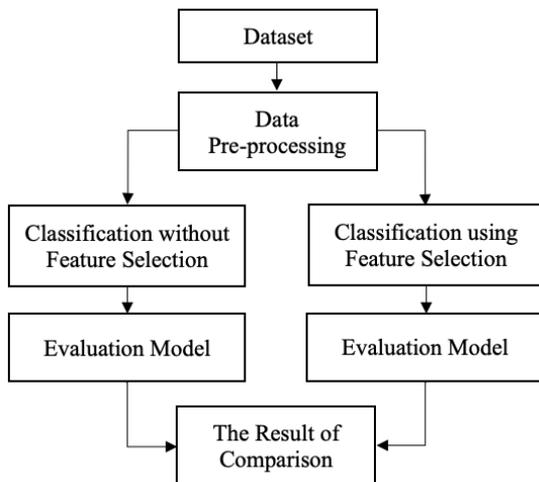


Figure 1. Research Methodology

This study compares several classification models in data mining. To obtain the desired output, there are several research processes which are applied. The first step is collecting the data required for

the classification process. In this case, the data set was obtained from a cardiocographic inspection and the expert diagnosis. Before classification, the dataset will be processed to convert raw data into higher-quality data using the pre-processing stage. The next process is classification using the models to be compared. The results of the classification will be evaluated using the evaluation model to determine the performance of each model. The research steps can be seen in Figure 1.

Dataset

This research uses a dataset provided in an open repository at the University of California Irvine (UCI) Machine Learning. The data comes from Marques de Sa et al [18] from the Biomedical Engineering Institute, Portugal. The dataset consists of 21 attributes that will be used to classify fetal classes. In this dataset, the instances have been classified into 1655 normal classes, 295 suspect classes, and 176 pathologic classes. Attributes data and analysis in the dataset can be seen in Table 1.

Data Pre-Processing

Pre-processing is needed to prepare raw data into higher quality data. The data obtained at UCI ML cannot be fully processed for classification because there are several structures that must be adjusted. Some classification algorithms require classes in the nominal data form. So, some of the class attributes for classification needs to be converted into nominal data. In the raw data, the class label uses the number 1 for the normal class, number 2 for the suspect class and number 3 for the pathological class. It is numerical data. So, it must be converted into nominal data. For normal class will be converted to letter A, suspect class to letter B and pathologic class to letter C. In addition to transforming the data, the scaling process will be performed at this pre-processing stage.

Table 1. Attribute and Dataset Analysis

No	Code	Type	Min	Max	Standard Deviation	Information
1	LB	Numeric	106	160	9.841	Fetal Heart Rate per Minute
2	AC	Numeric	0	26	3.561	Acceleration of Fetal Heart Rate per Second
3	FM	Numeric	0	564	37.125	Fetal Movement per Second
4	UC	Numeric	0	23	2.847	Number of Uterine Contractions per Second
5	DL	Numeric	0	16	2.499	Light deceleration per second
6	DS	Numeric	0	1	0.057	Severe deceleration per second
7	DP	Numeric	0	4	0.464	Prolonged deceleration per second
8	ASTV	Numeric	12	87	17.193	Percentage of time with abnormal short-term variability
9	MSTV	Nominal		57 Distinct		The average value of short-term variability
10	ALTV	Numeric	0	91	18.397	Percentage of time with abnormal long-term variability
11	MLTV	Nominal		249 Distinct		The average value of long-term variability
12	Width	Numeric	3	180	38.956	Fetal Heart Rate Histogram Width Value
13	Min	Numeric	50	159	29.56	Minimum Histogram Value of Fetal Heart Rate
14	Max	Numeric	122	238	17.944	Maximum Histogram Value of Fetal Heart Rate
15	Nmax	Numeric	0	18	2.949	Peak histogram Value of Fetal Heart Rate
16	NZeros	Numeric	0	10	0.706	Zero Fetal Heart Rate Histogram Value
17	Mode	Numeric	60	187	16.381	Histogram Mode Value
18	Mean	Numeric	73	182	15.594	Histogram Average Value
19	Median	Numeric	77	186	14.467	Median Histogram Value
20	Variance	Numeric	0	269	28.978	Histogram Variance Value
21	Tendency	Nominal		3 Distinct		Histogram Tendency Value
22	NSP	Numeric	1	3	0.614	Fetal Condition Class

As seen in Table 1, the standard deviation values for the attributes are still quite large. This condition will cause problems in several algorithms which using vector and euclidean distance. So, the scaling process is needed to reduce the range of values on a numeric feature [19].

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Formula 1 is a mathematical calculation of scaling in pre-processing. The formula gets normal data (x_{norm}) which has a vulnerability between 0-1. Where x_{min} is the minimal value and x_{max} is the maximal value.

Classification Models

This study compares the performance of several classification algorithms including Naïve Bayes, J48, Random Forest, Logistic Regression, K-Nearest Neighbors, Support Vector Machines, and Multi-Layer Perception. The following is an explanation of the classification algorithm

Naïve Bayes

Pouria Kaviani et al explain the Naïve Bayes algorithm, which is decision theory, is the simplest form of the Bayesian network model. This model gives a class of independent attributes [20]. Of all the many classification algorithms, Naïve Bayes is the simplest algorithm [21]. But with this simple structure, Naïve Bayes can provide better performance compared to other algorithms in some cases such as the classification of documents by Ting et al. [22].

J48 Algorithm

J48 algorithm is basically a development of other Decision Tree algorithms such as C4.5 and ID3 [23]. J48 is used for a single or multi-class classification model with a decision tree model. The advantage of this algorithm is able to process numerical and discrete data. Because this algorithm is a development of the C4.5 and ID3 algorithms, the accuracy of the classification result can be improved. In addition to accuracy, computing speed performance can also be reduced optimally [24].

Random Forest

Random forest is a classification algorithm which combines several branching predictions. The classification results depend on the size of the vector chosen randomly with the same data distribution for all data [25]. The Random Forest algorithm is suitable for classifying very large amounts of data. Some research study prove that random forest can provide a fairly good accuracy [26] even reaching 100% accuracy in some datasets such as research conducted by Zerina et al in the field of electrocardiogram [27].

Logistic Regression

Generally, the logistic regression is a type of regression provides associations between independent and dependent variable, which is class in classification. This is useful for analyzing observational data when it is needed to reduce the bias resulting from differences in features data [28]. The improved models of logistic

regression can increase accuracy similar to Artificial Neural Network models [29]. Some research on logistic regression is used in the medical world to know patient characteristics [30] and analyze of clinical risk scores in patients [31].

K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is a classification method with learning data closest distance to an object entity. The proximity value is determined from the K variable applied to the model [32]. This algorithm is also widely used because of some interesting features and easy implementation [33]. Several studies applied to the KNN model were conducted by Ahmad B. A Hasanat for classification in the big data field [34]. In the medical, KNN is applied to classify prostate cancer with the high results of classification and computational speed [35].

Support Vector Machine (SVM)

One of methods which widely used in supervised learning is Support Vector Machine. This model was proposed by Vapnik [36] in 1995. The basic idea of SVM is to maximize the distance between two classes. The distance is traditionally determined by the closest point [37]. Some effectiveness has been shown in study on pattern recognition [38]. Therefore, this model is able to provide better performance results than other models [39].

Multi-Layer Perceptron

Multi-Layer Perceptron is a model of a Neural Network consisting of several neurons. Another characteristic of this model is having a connecting weight between neurons [40]. The function of the neuron is processing and calculating the activation value of each unit from the input to the output layer [41]. Research on the prediction of nanofluid viscosity was conducted by Elham Heidari et al using the MLP model and produced a statistical regression analysis of 0.9999 [42].

Feature Selection

The dataset used in this study has several attributes characterize each class. In data mining, there are several attribute selection to find the best attributes having a large impact on the classification. The purpose of this feature selection is to reduce dimensions and eliminate the un-relevant feature [43]. In this research, 3 feature selection models are applied, including Correlation-based Feature Selection (CSF) Subset Evaluator, Info Gain, and Chi-Square. CFS Subset will evaluate attribute by considering the predictive ability of each individual with the level of redundancy between them [44]. CFS Subset can be calculated using Formula 2.

$$Ms = \frac{k \overline{r_{ef}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \quad (2)$$

Where M_s is a measurement of a feature subset that contains the attribute k , r_{ef} is the average of the correlation between attributes and class, while r_{ff} is the average of the correlations between attributes.

$$G(D, t) = \sum_{i=1}^m P(C_1) \log P(C_1) \quad (3)$$

$$+P(t) \sum_{i=1}^m P(C_1|t) \log P(C_1|t)$$

$$+P(\bar{t}) \sum_{i=1}^m P(C_1|\bar{t}) \log P(C_1|\bar{t})$$

Formula 3 is a mathematical model of Info Gain Feature Selection. Where C is the sum of all document samples, m is the amount of data and t is a category. If the value of G (D, t) is large, then the value will be used for the classification process in data C. Information Gain is an entropy-based feature evaluation method where this model is widely used in machine learning. The gain information is obtained and calculated using scoring and weighting. The feature will be selected if the Information Gain value of the feature is greater than or equal to the threshold value [45]. Furthermore, the last feature selection model used in this study is Chi-Square. Chi-Square will evaluate attributes which are not relevant in the classification process. The Chi-Square model uses basic statistical theory for testing the independence of features in a class [46]. To calculate the independence of these features can be written mathematically in Formula 4.

$$X^2 = \sum_{i=1}^n \frac{(O - E)^2}{E} \quad (4)$$

Where O is the observation value, n is the amount of data, and E is the expected value.

Evaluation Model

In this research, the evaluation process is carried out to provide comparative information about the classification result. The evaluation model is important to provide the performance results of the algorithm. Some of the matrices used to perform the testing are Precision, Recall, F-Measure, Matthews Correlation Coefficient (MCC), Receiver Operation Characteristic (ROC), Precision-Recall Curve (PRC) Area, and Accuracy. Each evaluation model has a function to measure the performance of the algorithm. In general, only accuracy will describe our results to show our findings. But for the health domain we need more evaluation models to get precise results. Before looking for evaluation values from the model, it needs to calculate the True Positive (TP) Rate, True Negative (TN) Rate, False Positive (FP) Rate, and False Negative (FN) Rate, where the four terms are some of the terminologies known in the Confusion Matrix. The TP rate is the amount of positive data that is detected correctly. TN rate is the amount of negative data that is detected correctly. Whereas the FP rate is the amount of positive data that is detected wrongly resulting in negative detection and the FN rate is the amount of negative data that was detected wrongly by the model. After knowing the values of that terminology, the following is the formula used to calculate the evaluation model mentioned earlier.

$$Precision = \frac{\sum_{i=1}^l TP}{\sum_{i=1}^l (FP + TP)} \quad (5)$$

$$Recall = \frac{\sum_{i=1}^l TP}{\sum_{i=1}^l (TP + FN)} \quad (6)$$

$$FMeasure = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (7)$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (8)$$

$$Accuracy = \frac{\sum_{i=1}^l (TP + TN)}{\sum_{i=1}^l (TP + TN + FP + FN)} \quad (9)$$

ROC can be obtained by measuring the threshold between False Positive Rate and True Positive Rate. Whereas the PRC Area is obtained by calculating the threshold between Recall and Precision. To find out the average error between the predicted results and the actual value, it is necessary to measure the Mean Absolute Error (MAE). MAE has a better performance compared to Root Mean Square Error [47]. MAE is mathematically defined into Formula 10.

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| \quad (10)$$

where f_i is the value of the forecast result, y_i is the result of the actual value, while n is the amount of data. In addition, KAPPA analysis will be conducted to determine the suitability of interpretation or prediction in an algorithm. In their research, Rosario et al explain why KAPPA needs to be performed in measuring the performance of classification models. They suggested that in some cases, KAPPA gave different evaluation results from MCC. The low classification results turned out to get a high score in KAPPA, but qualitatively different in MCC [48]. Therefore, the KAPPA evaluation results will be tested in the calculation of the research results. The mathematical model in KAPPA is defined in formula 11.

$$Kappa = \frac{P_o - P_e}{1 - P_e} \quad (11)$$

where P_o is the relative agreement value being predicted and P_e is the hypothesis probability of the prediction. All formulas that have been explained will be used to evaluate the results of the classification performed by each algorithm.

RESULTS AND DISCUSSION

This section will discuss the results of classification experiments with various models. The scenarios are carried out by using all the attributes and by applying the feature selection to reduce the attribute dimensions. The results of the model will be compared to determine the best performance. All experiments on each model are performed using 10 Folds cross-validation. This evaluation process will separate 9 data for the training process and 1 data for the testing process. K-Fold selection is performed to reduce the computational process during training while maintaining the accuracy of the classification results.

Result

The first experiment scenario is performed to see the performance of each algorithm for all attributes. In the dataset, there are 21 attributes that will all be used for the classification process. Each algorithm will be evaluated according to the evaluation model described in the previous chapter. Table 2 is the result of testing the algorithm by using all the attributes in the dataset without using feature selection. While Table 3 is the result of testing the algorithm by applying the CFS subset feature selection. The CFS

Subset Evaluation model applies a forward search direction with a search limit of 10 expansion nodes. Thus, the selection of features in the CFS Subset Evaluation gets the best subset with a value of 0.338. From this subset, 7 best attributes were selected to be used as classification attributes. These are AC, UC, ASTV, ALTV, DS, DP and Mean.

The next scenario is applying Info Gain Feature Selection with the 10 best attribute ranker model. The gain ratio obtained the 10 best attributes, including MSTV, ALTV, ASTV, Mean, Variance, AC, Median, Mode, Width, MLTV, and NSP. The results of the test scenario can be seen in Table 4. The final scenario is performed to evaluate the classification results using the Chi-Square feature selection. The model uses ranker with the selection of the 15 best attributes. The attribute selected in the classification process are Mean, ALTV, MSTV, Mode, Median, ASTV, DP, Variance, MLTV, Width, AC, Min, LB, UC, DL, and NSP. The test results with the Chi-Square feature selection can be seen in Table 5. Title of a table, must be put above the table. In writing the title of a table, you must capitalize each word. All tables and figures must be consecutively numbered. Figures must be made center-align and left-align for tables.

Table 2. Classification Result without Feature Selection

Algorithm	MAE	Kappa Statistic	Average Precision	Average Recall	Average F-Measure	Average MCC	Average ROC Area	Average PRC Area	Accuracy
Naïve Bayes	0.137	0.5478	0.861	0.797	0.816	0.603	0.920	0.888	79.68 %
J48	0.0705	0.7546	0.910	0.912	0.911	0.768	0.909	0.891	91.15 %
Random Forest	0.0792	0.8136	0.932	0.934	0.933	0.825	0.984	0.974	93.41 %
Logistic Regression	0.1215	0.6214	0.861	0.868	0.863	0.656	0.948	0.914	86.78 %
KNN	0.0774	0.6805	0.882	0.885	0.883	0.704	0.845	0.837	88.47 %
SVM	0.2572	0.5928	0.854	0.865	0.856	0.625	0.793	0.799	86.45 %
MLP	0.0825	0.7078	0.896	0.891	0.893	0.724	0.946	0.923	89.08 %

Table 3. Classification Result using CFS Subset Feature Selection

Algorithm	MAE	Kappa Statistic	Average Precision	Average Recall	Average F-Measure	Average MCC	Average ROC Area	Average PRC Area	Accuracy
Naïve Bayes	0.1221	0.6117	0.872	0.840	0.849	0.654	0.928	0.900	83.96 %
J48	0.0669	0.7853	0.921	0.922	0.921	0.788	0.911	0.902	92.23 %
Random Forest	0.0713	0.8126	0.932	0.933	0.932	0.817	0.979	0.972	93.32 %
Logistic Regression	0.1314	0.5877	0.849	0.858	0.852	0.626	0.936	0.901	85.84 %
KNN	0.0719	0.7047	0.891	0.892	0.892	0.724	0.863	0.852	89.22 %
SVM	0.2601	0.5545	0.842	0.854	0.843	0.591	0.775	0.786	85.37 %
MLP	0.0993	0.6755	0.883	0.886	0.884	0.687	0.940	0.927	88.57 %

Table 4. Classification Result using Info Gain Feature Selection

Algorithm	MAE	Kappa Statistic	Average Precision	Average Recall	Average F-Measure	Average MCC	Average ROC Area	Average PRC Area	Accuracy
Naïve Bayes	0.1274	0.59	0.870	0.822	0.836	0.634	0.925	0.894	82.22 %
J48	0.0679	0.7796	0.919	0.920	0.920	0.782	0.897	0.887	92.00 %
Random Forest	0.0744	0.8262	0.937	0.937	0.937	0.833	0.977	0.971	93.74 %
Logistic Regression	0.1292	0.5826	0.848	0.858	0.851	0.614	0.940	0.909	85.84 %
KNN	0.0743	0.6928	0.887	0.889	0.888	0.708	0.853	0.848	88.89 %
SVM	0.2612	0.5621	0.844	0.855	0.846	0.591	0.774	0.788	85.46 %
MLP	0.0877	0.6942	0.891	0.892	0.890	0.703	0.947	0.933	89.22 %

Table 5. Classification Result using Chi-Square Feature Selection

Algorithm	MAE	Kappa Statistic	Average Precision	Average Recall	Average F-Measure	Average MCC	Average ROC Area	Average PRC Area	Accuracy
Naïve Bayes	0.1414	0.5654	0.859	0.813	0.827	0.604	0.901	0.879	81.27 %
J48	0.0685	0.7585	0.911	0.913	0.912	0.767	0.909	0.894	91.25 %
Random Forest	0.0773	0.8162	0.933	0.935	0.933	0.827	0.982	0.972	93.46 %
Logistic Regression	0.122	0.6258	0.863	0.870	0.865	0.660	0.948	0.912	87.01 %
KNN	0.0743	0.6929	0.887	0.889	0.888	0.709	0.855	0.846	88.89 %
SVM	0.2567	0.5963	0.855	0.866	0.857	0.634	0.797	0.801	86.59 %
MLP	0.0826	0.697	0.891	0.891	0.890	0.710	0.949	0.933	89.08 %

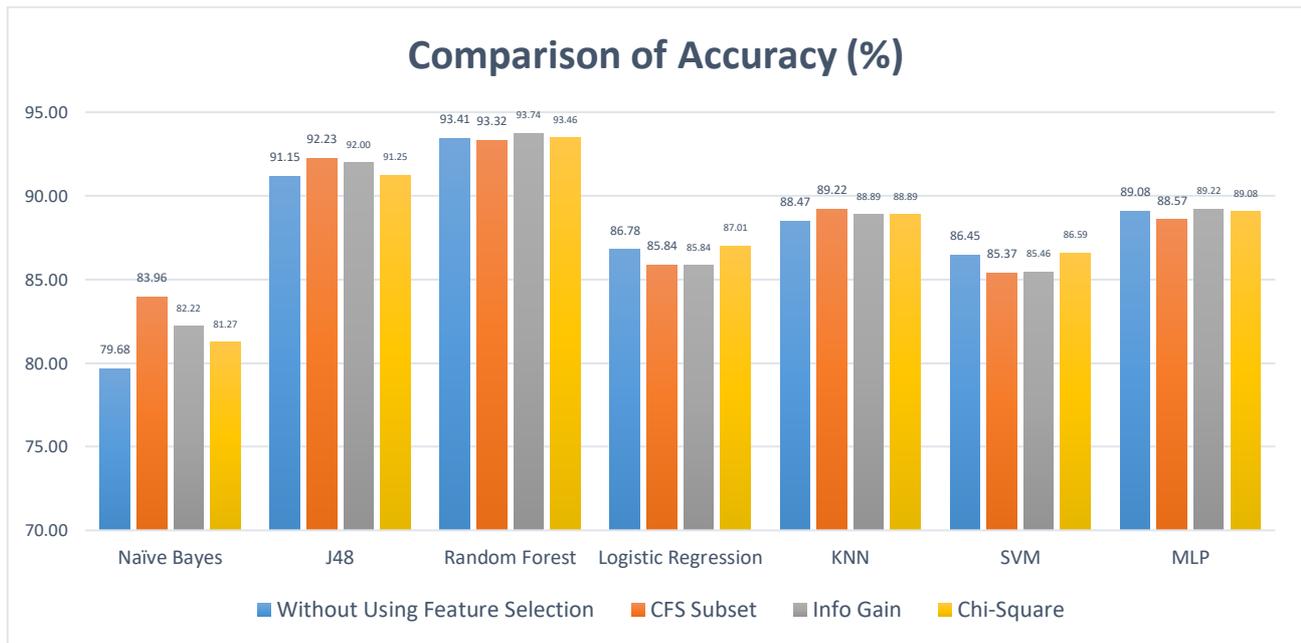


Figure 2. Comparison of Classification Result

Discussion

Based on research that has been performed, it can be seen the results of the comparison of accuracy as in Figure 2. The use of accuracy aims to simplify the result of our main findings. The testing scenarios carried out also implement several feature selections to determine the performance of each model. The results of the accuracy and evaluation of the model can be seen in Tables 2-5.

In the first experiment, the model is used to classify all features without using feature selection. Random Forest produces the highest accuracy with a percentage of 93.41%, followed by J48 with an accuracy of 91.15%. The results of the second scenario are carried out to determine the accuracy and performance of the classification by applying feature selection. In the CFS subset feature selection, Random Forest produces the highest accuracy with a percentage of 93.32% and followed by J48 with a percentage of 92.23%. In the Info Gain feature selection, the highest results are obtained by the Random Forest algorithm with a percentage of 93.74%, followed by the J48 algorithm with a percentage of 92%. Last, in the Chi-Square feature selection model, the highest accuracy is obtained in the Random Forest

model with a percentage of 93.46% followed by the J48 algorithm with an accuracy of 91.25%.

It can be concluded from this study that the combination of the Random Forest algorithm with Info Gain feature selection gets the best results with an accuracy of 93.74%. In this experiment, random forest has a good classifier compared to the other classification model in handling the high dimension, and unbalanced data. On the dataset used, we analyze that the dataset has high dimension and imbalance data. In addition, feature selection can improve accuracy in several algorithms. But in this case, some feature selection models still cannot significantly improve accuracy in the Logistic Regression, SVM, and MLP algorithms. That is because there are differences in mathematical calculations of each feature selection of the algorithm and dataset used. In other cases, the applied feature selection models can improve the accuracy of the classification models. Indirectly, algorithms and datasets affect the accuracy of the classification process.

CONCLUSIONS

Cardiotocography is a medical inspection to determine the health of the fetus in pregnancy. This result can be detected by

measuring the parameters of the heartbeat in the fetus and uterine contractions. The diagnosis process can utilize machine learning to conduct data analysis and classification. In this study, a comparative study of classification models was carried out with two main scenarios. The first scenario is performed to determine the performance of the algorithm without using feature selection. The highest accuracy obtained on the Random Forest algorithm with a percentage of 93.41%. The second scenario is carried out to determine the effect of feature selection on improving the accuracy of classification results. Based on this research, the highest accuracy of all scenarios is obtained by the Random Forest algorithm with Info Gain Feature Selection. The combination of these models produces accuracy with a percentage of 93.74%.

REFERENCES

- [1] P. U. Okorie, "The Significant of Biomedical Engineering to Medical Field in Nigeria The Significant of Biomedical Engineering to Medical Field in," *Am. J. Biomed. Sci. Eng.*, vol. 2, no. 1, 2015.
- [2] A. Illanes and M. Haritopoulos, "Fetal heart rate feature extraction from cardiotocographic recordings through autoregressive model's power spectral- and pole-based analysis," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2015-Novem, pp. 5842–5845, 2015.
- [3] C. Rotariu, H. Costin, A. Păsărică, and agoș Nemescu, "Classification of parameters extracted from cardiotocographic signals for early detection of metabolic acidemia in newborns," *Adv. Electr. Comput. Eng.*, vol. 15, no. 3, pp. 161–166, 2015.
- [4] A. Pinas and E. Chandraran, "Continuous cardiotocography during labour: Analysis, classification and management," *Best Pract. Res. Clin. Obstet. Gynaecol.*, vol. 30, pp. 33–47, 2016.
- [5] K. A. Allen and D. H. Brandon, "Hypoxic Ischemic Encephalopathy: Pathophysiology and Experimental Treatments," *Newborn Infant Nurs. Rev.*, vol. 11, no. 3, pp. 125–133, 2011.
- [6] D. Ayres-De-Campos, C. Y. Spong, and E. Chandraran, "FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography," *Int. J. Gynecol. Obstet.*, vol. 131, no. 1, pp. 13–24, 2015.
- [7] H. Tang, T. Wang, M. Li, and X. Yang, "The Design and Implementation of Cardiotocography Signals Classification Algorithm Based on Neural Network," *Comput. Math. Methods Med.*, vol. 2018, 2018.
- [8] Z. CÖMERT and A. KOCAMAZ, "A Study of Artificial Neural Network Training Algorithms for Classification of Cardiotocography Signals," *Bitlis Eren Univ. J. Sci. Technol.*, vol. 7, no. 2, pp. 93–103, 2017.
- [9] N. J. Ali Kadhim and J. Kadhim Abed, "Enhancing the Prediction Accuracy for Cardiotocography (CTG) using Firefly Algorithm and Naive Bayesian Classifier," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 745, p. 012101, 2020.
- [10] M. E. B. Menai, F. J. Mohder, and F. Al-mutairi, "Influence of Feature Selection on Naïve Bayes Classifier for Recognizing Patterns in Cardiotocograms," *J. Med. Bioeng.*, vol. 2, no. 1, pp. 66–70, 2013.
- [11] H. Ocak, "A medical decision support system based on support vector machines and the genetic algorithm for the evaluation of fetal well-being," *J. Med. Syst.*, vol. 37, no. 2, 2013.
- [12] N. Chamidah and I. Wasito, "Fetal state classification from cardiotocography based on feature extraction using hybrid K-Means and support vector machine," *ICACSIS 2015 - 2015 Int. Conf. Adv. Comput. Sci. Inf. Syst. Proc.*, pp. 37–41, 2016.
- [13] M. Arif, "Classification of cardiotocograms using random forest classifier and selection of important features from cardiotocogram signal," *Biomater. Biomech. Bioeng.*, vol. 2, no. 3, pp. 173–183, 2015.
- [14] R. Jyothi, S. Hiwale, and P. V. Bhat, "Classification of labour contractions using KNN classifier," *2016 Int. Conf. Syst. Med. Biol. ICSMB 2016*, no. January, pp. 110–115, 2017.
- [15] S. A. A. Shah, W. Aziz, M. Arif, and M. S. A. Nadeem, "Decision Trees Based Classification of Cardiotocograms Using Bagging Approach," *Proc. - 2015 13th Int. Conf. Front. Inf. Technol. FIT 2015*, pp. 12–17, 2016.
- [16] S. Velappan, D. Murugan, J. Rani, and K. Rajalakshmi, "Comparative Analysis of Classification Techniques using Cardiotocography Dataset," *IJRIT Int. J. Res. Inf. Technol.*, vol. 1, no. 12, pp. 274–280, 2013.
- [17] K. Agrawal and H. Mohan, "Cardiotocography Analysis for Fetal State Classification Using Machine Learning Algorithms," *2019 Int. Conf. Comput. Commun. Informatics, ICCCI 2019*, no. October, pp. 1–6, 2019.
- [18] M. de Sa, J. Bernardes, and A. de Campos, "Cardiotocography Data Set," *UCI - Machine Learning Repository*, 2010. <http://archive.ics.uci.edu/ml/datasets/Cardiotocography>.
- [19] S. Agarwal, *Data mining: Data mining concepts and techniques*, 3rd ed. Elsevier Inc., 2014.
- [20] P. Kaviani and S. Dhotre, "Short Survey on Naive Bayes Algorithm," *Int. J. Adv. Eng. Res. Dev.*, vol. 4, no. 11, pp. 607–611, 2017.
- [21] S. Taheri and M. Mammadov, "Learning the naive bayes classifier with optimization models," *Int. J. Appl. Math. Comput. Sci.*, vol. 23, no. 4, pp. 787–795, 2013.
- [22] S. L. Ting, W. H. Ip, and A. H. C. Tsang, "Is Naïve bayes a good classifier for document classification?," *Int. J. Softw. Eng. its Appl.*, vol. 5, no. 3, pp. 37–46, 2011.
- [23] N. Satyanarayana, Y. Ramadevi, and K. K. Chari, "High blood pressure prediction based on AAA using J48 classifier," *2018 Conf. Signal Process. Commun. Eng. Syst. SPACES 2018*, vol. 2018-Janua, pp. 121–126, 2018.
- [24] R. Patil and V. M. Barkade, "Class-Specific Features Using J48 Classifier for Text Classification," *Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018*, pp. 1–5, 2018.
- [25] Y. L. Pavlov, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
- [26] F. Da Silva, J. Desaphy, G. Bret, and D. Rognan, "iChemPIC: A Random Forest Classifier of Biological and Crystallographic Protein-Protein Interfaces," *J. Chem. Inf. Model.*, vol. 55, no. 9, pp. 2005–2014, 2015.
- [27] Z. Masetic and A. Subasi, "Congestive heart failure detection using random forest classifier," *Comput. Methods Programs Biomed.*, vol. 130, pp. 54–64, 2016.
- [28] M. P. LaValley, "Logistic regression," *Circulation*, vol.

- 117, no. 18, pp. 2395–2399, 2008.
- [29] Y. P. Lin, H. J. Chu, C. F. Wu, and P. H. Verbarg, “Predictive ability of logistic regression, auto-logistic regression and neural network models in empirical land-use change modeling - a case study,” *Int. J. Geogr. Inf. Sci.*, vol. 25, no. 1, pp. 65–87, 2011.
- [30] J. Tolles and W. J. Meurer, “Logistic regression: Relating patient characteristics to outcomes,” *J. Am. Med. Assoc.*, vol. 316, no. 5, pp. 533–534, 2016.
- [31] W. J. Meurer and J. Tolles, “Logistic Regression Diagnostics : Understanding How Well a Model Predicts Outcomes,” *J. Am. Stat. Assoc.*, vol. 317, p. 1068, 2017.
- [32] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 27, no. 1. Springer, 2006.
- [33] N. Garcia-Pedrajas, J. A. Romero Del Castillo, and G. Cerruela-Garcia, “A Proposal for Local k Values for k-Nearest Neighbor Rule,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 28, no. 2, pp. 470–475, 2017.
- [34] A. B. A. Hassanat, “Two-point-based binary search trees for accelerating big data classification using KNN,” *PLoS One*, vol. 13, no. 11, pp. 1–15, 2018.
- [35] C. Jensen, J. Carl, L. Boesen, N. C. Langkilde, and L. R. Østergaard, “Assessment of prostate cancer prognostic Gleason grade group using zonal-specific features extracted from biparametric MRI using a KNN classifier,” *J. Appl. Clin. Med. Phys.*, vol. 20, no. 2, pp. 146–153, 2019.
- [36] V. N. Vapnik, *The Nature of Statistical Learning*, 2nd ed. New-York: Springer-Verlag, 1995.
- [37] X. Shen, L. Niu, Z. Qi, and Y. Tian, “Support vector machine classifier with truncated pinball loss,” *Pattern Recognit.*, vol. 68, no. 5, pp. 199–210, 2017.
- [38] H. Byun and S. W. Lee, “Applications of support vector machines for pattern recognition: A survey,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2388, pp. 213–236, 2002.
- [39] M. W. Huang, C. W. Chen, W. C. Lin, S. W. Ke, and C. F. Tsai, “SVM and SVM ensembles in breast cancer prediction,” *PLoS One*, vol. 12, no. 1, pp. 1–15, 2017.
- [40] F. Murtagh, “Multilayer perceptrons for classification and regression,” *Neurocomputing*, vol. 2, no. 5–6, pp. 183–197, 1991.
- [41] H. Ramchoun, M. Amine, J. Idrissi, Y. Ghanou, and M. Ettaouil, “Multilayer Perceptron: Architecture Optimization and Training,” *Int. J. Interact. Multimed. Artif. Intell.*, vol. 4, no. 1, p. 26, 2016.
- [42] E. Heidari, M. A. Sobati, and S. Movahedirad, “Accurate prediction of nanofluid viscosity using a multilayer perceptron artificial neural network (MLP-ANN),” *Chemom. Intell. Lab. Syst.*, vol. 155, pp. 73–85, 2016.
- [43] J. Li *et al.*, “Feature Selection: A Data Perspective,” *ACM Comput. Surv.*, vol. 50, no. 6, 2017.
- [44] K. R. Pushpalatha and A. G. Karegowda, “CFS Based Feature Subset Selection for Enhancing Classification of Similar Looking Food Grains-A Filter Approach,” *2017 2nd Int. Conf. Emerg. Comput. Inf. Technol. ICECIT 2017*, pp. 1–6, 2018.
- [45] S. Lei, “A feature selection method based on information gain and genetic algorithm,” *Proc. - 2012 Int. Conf. Comput. Sci. Electron. Eng. ICCSEE 2012*, vol. 2, pp. 355–358, 2012.
- [46] X. Jin, A. Xu, R. Bie, and P. Guo, “Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3916 LNBI, pp. 106–115, 2006.
- [47] C. J. Willmott and K. Matsuura, “Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance,” *Clim. Res.*, vol. 30, no. 1, pp. 79–82, 2005.
- [48] R. Delgado and X. A. Tibau, “Why Cohen’s Kappa should be avoided as performance measure in classification,” *PLoS One*, vol. 14, no. 9, pp. 1–26, 2019.