SERBIAN JOURNAL OF ELECTRICAL ENGINEERING
Vol. 19, No. 2, June 2022, 207-224

UDC: 004.85::616.12-073.7-053.13 DOI: https://doi.org/10.2298/SJEE2202207J

Experimental Analysis of Filtering-Based Feature Selection Techniques for Fetal Health Classification

Immanuel Johnraja Jebadurai[1], Getzi Jeba Leelipushpam Paulraj[1], Jebaveerasingh Jebadurai[1], Salaja Silas[1]

Abstract: Machine learning techniques enable computers to acquire intelligence through learning. Trained machines can carry out various tasks, such as prediction, classification, clustering, and recommendation, within a wide variety of applications. Classification is a supervised learning technique that can be improved using feature selection techniques such as filtering, wrapping, and embedding. This paper explores the impact of filtering-based feature selection techniques on classification methods, and focuses on an analysis of correlation-based filtering techniques based on Pearson, Spearman, and Kendall rank correlation. Similarly, we explore the impacts of using statistical filtering techniques such as mutual information, chi-squared score, the ANOVA univariate test, and the univariate ROC-AUC. These filtering techniques are evaluated by implementing them with the k-nearest neighbor, support vector machine, decision tree, and Gaussian naïve Bayes classification methods. Our experiments were carried out using a fetal heart rate dataset, and the performance of each combination of methods was measured based on precision, recall, F1-score, and accuracy. An analysis of the experimental results showed that the performance metrics for the Gaussian naïve Bayes and k-nearest neighbor methods were improved by 3% through the use of the statistical feature selection technique, and a 4% improvement was observed for the decision tree and support vector machine methods using a correlation-based filtering technique. Of the statistical feature selection techniques, ANOVA and ROC-AUC were the best as they improved the accuracy by 92%; compared to the other correlation techniques, the Spearman correlation coefficient gave the best results, as it also improved the accuracy by 92%.

Keywords: Fetal ECG, Machine Learning, Feature Selection, Supervised Learning, Classification, Accuracy.

[1]Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India; E-mails: immanueljohnraja@gmail.com, getz23@gmail.com, jebaveerasingh.j@gmail.com, blessingsalaja@gmail.com

1 Introduction

Machine learning has revolutionised the fields of medicine, trade, government, and many others. With the advent of sophisticated computer hardware and software, machine learning has introduced intelligence into the applications used in daily life [1], and has incorporated intelligence into computers using supervised, unsupervised, and reinforcement learning techniques [2]. In supervised learning, a model is trained using labeled data; the expected output is fed in as input, and the system is trained on this. The trained system is then deployed to make predictions and classify incidents. In the unsupervised learning mode, unlabeled input is fed into the system, which then identifies patterns and clusters the data. Reinforcement learning involves the presence of an agent which learns the environment. Each correct action is rewarded and recorded for future responses, and a negative impact is penalised to ensure that the system will not repeat this in future predictions [3].

In this paper, we primarily focus on improving the supervised classification technique. Classification is a type of supervised learning method in which labeled data are used, and the machine is trained to group these data. In the testing phase, the machine is expected to group data without labels, based on the knowledge gained during the training phase [4]. Classification is widely used in many applications, including disease prediction, text recognition, email spam filtering, and image segmentation. The performance of a classification technique can be evaluated using various metrics such as accuracy, precision, recall, and F1-score. For a given set of data, the features are analysed, and based on the feature values, classification is carried out. However, the inclusion of irrelevant features degrades the performance of classification techniques, and it is therefore critical to analyze the feature selection technique for suitability before submitting data from a machine learning algorithm. This paper compares the results from several filtering-based feature selection techniques and analyzes their impact on the supervised classification technique. A fetal heart rate dataset is used to compare these filtering-based feature selection techniques; it consists of vital signs such as foetal movements, accelerations, and decelerations, and was collected during labor to screen fetuses for hypoxia [5].

2 Feature Selection in Machine Learning

The feature selection method selects the best features from a dataset, which helps to improve the accuracy of the machine learning algorithm. Various feature learning approaches such as filtering-based, wrapper, embedded, and hybrid methods can be combined with machine learning prediction, classification, and clustering to improve the performance in applications. This paper explores the impact of filtering-based feature selection techniques on traditional classification techniques.

2.1 Filtering-based feature selection techniques

Filtering-based feature selection techniques filter out irrelevant, constant, and duplicate features and feed only the best features to the classifier for training. A technique that relies on only a single feature is called a univariate filtering technique, whereas one that considers the entire feature space is called a multivariate feature technique.

Basic filtering-based feature selection techniques

The basic filtering technique involves eliminating the following features:

Constant features: In a dataset, columns that contain all the same features are constant features. These features are less informative to a machine learning model in terms of predicting or classifying the target. Hence, in a constant filtering-based technique, these columns are eliminated.

Quasi-constant features: In feature columns, if a given value occupies most of the rows, these are called quasi-constant features. These columns are eliminated, as they are less significant for classification.

Duplicate features: Some features may be repeated in the dataset. Using such data for training will cause inaccuracy in terms of prediction and classification of the data. Hence, such features are eliminated.

Correlation-based feature selection techniques

Correlation indicates the relationships between the features of a dataset. Highly correlated features are included for training. The various types of correlation techniques are detailed below.

<u>Pearson correlation coefficient:</u> Pearson correlation is applied to features with a linear relationship. A value of one on a relative scale indicates that the features are highly correlated, and the correlation vanishes as the value reaches −1. The formula for calculating the Pearson coefficient is:

$$P_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}},$$

where n refers to the number of samples, x_i, y_i refer to the current training and testing samples, and $\overline{x}_i$, $\overline{y}_i$ refers to the means of the training and testing samples.

<u>Spearman’s rank correlation coefficient:</u> Spearman’s rank correlation accepts features with nonlinear relationships and measures the strength of

variables over the range [−1, 1]. It is suitable for both ordinal and continuous variables. The formula used to calculate Spearman's rank correlation is:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)},$$

where n refers to the number of samples, and d_i refers to the paired rank between the features.

Kendall's rank correlation coefficient: This technique calculates a normalized score between two features that varies in the range [−1, 1]. The features are ranked based on the score. A rank of one indicates that the features are highly correlated, while a value of −1 shows that they are uncorrelated. The formula is:

$$\tau = \frac{2(n_c - n_d)}{n(n-1)},$$

where n refers to the number of samples, and n_c, n_d refer to the concordant /discordant pair [6].

Statistical and ranking-based filtering techniques

The features can be evaluated individually using a statistical test based on the target feature. The output defines whether the feature is suitable for classifying the target variable. The features are ranked, and the one with the highest ranking is used in the machine learning algorithm. The techniques in this category include mutual information, chi-squared score, analysis of variance (ANOVA) univariate test, and univariate receiver operating characteristics curve/area under curve (ROC-AUC) [7].

Mutual information: This measures the mutual dependence of two features, and indicates the extent to which a feature helps to predict or classify the target feature. The independence of two features results in a score of zero, and the entropy of the feature indicates dependence.

Chi-squared score: This is used to check the relationship between two categorical features. The observed distribution of the feature is compared with the target variable. This approach is not suitable for non-negative feature values.

ANOVA univariate test:This calculates the linear relationship between each feature in the dataset and the target. The feature with the highest correlation is selected for the machine learning algorithm.

Univariate ROC-AUC: This is used for classification problems. A decision tree (DT) is built for each single feature and the target. The output is ranked based on the ROC-AUC, and features with higher ranks are selected.

Classification techniques

Classification is a supervised machine learning technique with various applications [8 – 12]. Each item of input data belongs to one or more classes, and the classification model is expected to assign the test data from the dataset into these classes. Approaches of this type include naïve Bayes (NB), k-nearest neighbors (kNN), DT, and support vector machine (SVM).

Naive Bayes technique

In this approach, classification is performed based on Bayes' theorem. The predictor features are separated into a feature matrix, and the target feature is represented as a response vector. The features are assumed to be independent of each other. Bayes' theorem is given by (1):

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}. \quad (1)$$

To predict *A*, given the values of *B*, where *B* represents feature (f_1), feature (f_2),…, feature (f_n), (1) can be rewritten as:

$$P(A|(f_{1,}f_2\cdots\cdot f_n)=\frac{P(A|f_1)P(A|f_2)\cdots P(A|f_n)}{P(f_1)P(f_2)\cdots P(f_n)}. \quad (2)$$

Substituting (2) into (1), we get:

$$P(A|(f_{1,}f_2\cdots\cdot f_n)=\frac{P(A)\prod_{i=1}^{n}P(f_i|A)}{P(f_1)P(f_2)\cdots P(f_n)} \quad (3)$$

As the denominator is constant, (3) can be stated as follows:

$$P(A|(f_1,f_2,\ldots,f_n)=P(A)\prod_{i=1}^{n}P(f_i|A). \quad (4)$$

Now, for a given set of inputs, the probabilities of the data falling into each of the classes is calculated. The class with maximum probability is allocated to that input, as expressed in (5):

$$A=\arg\max_A P(A)\prod_{i=1}^{n}P(f_i|A). \quad (5)$$

k-nearest neighbors

kNN is a classical machine learning algorithm that is used in classification problems. For a given dataset with samples $((x_1,x_2,\ldots,x_n)$, the Euclidean distance is calculated between each sample and the others. Input x is assigned to the class with the highest probability using (6):

$$P(y=j|X=x)=\frac{1}{k}\sum_{i\in A}I(y^i=j). \quad (6)$$

When new testing data are presented for classification, the algorithm selects k neighbors in the data space, and the distances between the test data and each of the k neighbors are calculated. The test data item is allocated to the class with the minimum distance.

Decision tree

The DT technique predicts the classes for the test data by applying the decision rules accumulated at the training stage. The root node represents the entire set of data. Decision nodes are built gradually from the root node, based on the conditions imposed on the node above. The tree ends with the terminal or leaf node. Entropy (H) and information gain (IG) are the main parameters used by the DT algorithm.

Support vector machine

This algorithm plots each data point with in a feature space, and draws a hyperplane to segregate them into different classes. These approaches can be classified into linear SVM, in which the data are classified using a straight line, and nonlinear SVM, in which the data is classified using hyperplanes. SVM uses kernels to classify the data.

Need for feature selection for classification

The incorporation of feature selection into a classification model has the following advantages:

1. It improves the accuracy and reduces the execution time of classification [13];
2. It boosts the efficiency of classification techniques in terms of their precision, recall, and accuracy [9];
3. It reduces the dimensionality and noise in the data [14];
4. It identifies relevant features for classification tasks [15].

3 Dataset

In order to evaluate the performance of filtering-based feature selection techniques on a classification task, a fetal health dataset was selected [16]. This dataset consisted of 2,126 instances of cardiography (CTG) data with 22 features. Fetal cardiographs are taken throughout pregnancy to measure the urine contraction (UC) and fetal heart rate (FHR).

Catheters are placed in two locations: the higher one monitors the UC, and the lower one monitors the FHR. CTG data are utilized in order to identify early pathological effects, and can help the obstetrician to predict any future impairments [10]. Inappropriate understanding of CTG will lead to adverse effects on both the mother and fetus [17].

The dataset is visualized in Fig. 1. Its vital parameters include accelerations, fetal movements, and light and severe decelerations. These vital signs are analyzed in order to understand any abnormalities while the fetus is still in the mother's womb. An experimental analysis is carried out in the context of the need to understand the CTG accurately and to identify a suitable technique.

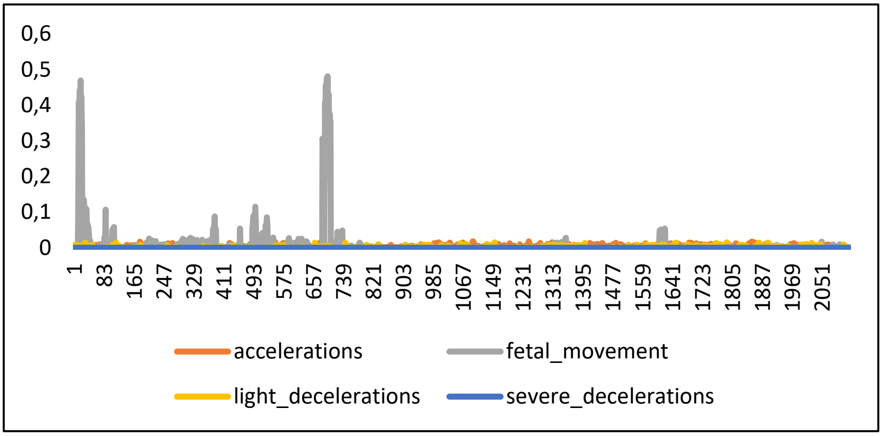

Fig. 1 – *Fetal heart rate signal.*

Performance metrics

The improvements in the results of the classification techniques due to the use of feature selection are evaluated using the following performance metrics:

Accuracy

This is the ratio between the number of correctly classified instances and the total number of instances, as shown in (7):

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}\,. \tag{7}$$

The terminologies associated with this metric are as follows:

– True positives (TP) are cases in which true cases are correctly predicted as true.

– True negative (TN) are cases where false cases are correctly predicted as false.

– False positive (FP) are cases in which false cases are wrongly predicted as true.

– False negatives (FN) are cases in which true cases are wrongly predicted as false.

Precision

Precision is a measure of the number of correctly classified cases in relation to the total number of cases, and is calculated using (8):

$$Precision = \frac{TP}{TP + FP} . \tag{8}$$

Recall

This is the ratio between the number of correctly classified true cases and the sum of the correctly classified and wrongly classified true cases. It is calculated using (9):

$$Recall = \frac{TP}{TP + FN} . \tag{9}$$

F1-score

This is the mean of the precision and recall, and is calculated as shown in (10):

$$F1\_score = \frac{2 * precision * recall}{precision + recall} . \tag{10}$$

Experiments

A set of experiments was carried out on the fetal cardiograph dataset [16]. The dataset had three classes (normal, suspect, and pathological), which were converted into numerical labels (1: normal; 2: suspect; and 3: pathological). There were 2,126 records with 22 numerical features. Implementation was carried out in Python version 3.7, and NumPy, pandas, and scikit-learn were used to import the libraries required for feature selection and classification. The values in the dataset were standardized using the MinMaxScaler function.

Various feature selection techniques were applied to the dataset of 22 features. The correlation-based feature selection technique is shown in Algorithm 1.

Algorithm 1: Correlation-based feature selection

$Corr_{features} = Correlation(data_{train}, thres)$

$function\ correlation\ (data_{train}, thres)\{$

$corr_{matrix} = data_{train}.corr(method = \text{pearson}|\text{Spearman}|\text{Kendall})$

$if\ \left(corr_{features[j]} = corr_{matrix}[i][j] > thres\right)\{\ return\ corr_{features[j]}\}$

Spearman's, Pearson's, and Kendall's methods were used to calculate the correlations between the features in this feature selection technique. If the correlation was greater than a threshold of 0.8, the feature was accepted for classification. The value of this threshold was selected after repeated experimentation.

The filter-based feature selection techniques included mutual information, chi-squared, ANOVA, and ROC-AUC. Pseudo-code for the feature selection technique is shown in Algorithm 2.

<u>Algorithm 2: Statistical-based feature selection</u>

$$filter_{feature} = SelectKBest(mutual\_info\_classif|chi2|classif, k = 10).fit(features_{train}, label_{train})$$

$$roc\_auc\_score(label_{train}, data_{train}, average = "macro", labels = labels, multi\_class"ovo")$$

$$filter_{feature} = roc\_auc\_score > 0.6$$

The SelectKBest function chooses the best features output by the filter-based feature selection technique. The value of K was set to 10 after repeated experiments. The methods mutual_info_classif, chi2, and classif were used for mutual information, chi-square, and ANOVA filter feature selection, respectively. For ROC-AUC, the function roc_auc_score was used with the "macro" average method, which calculates the unweighted mean of each feature.

The output labels used for feature selection were based on the "ovo" method, meaning one-to-one mapping. Pairwise comparisons of each label were considered in the AUC-ROC calculation. Then, each feature selection technique was applied with each classification method (NB, kNN, DT, and SVM). The experiment was repeated 10 consecutive times to obtain average values.

4 Performance Analysis

Performance analyses were carried out for both the correlation-based and statistical-based filtering techniques. Lists of features before and after filtering are given in **Table 1**. A total of 21 features were used to analyze fetal health; the Spearman correlation technique filtered out five features, leaving 16 features for fetal health analysis, while the Pearson and Kendall correlation methods each filtered out four features, leaving 17 for fetal health analysis using the various classification methods.

The 10 best features were selected from the original set of 21 using statistical-based filtering techniques. The selected features were used with the various classification techniques, and the performance metrics applied were precision, recall, F1-score, and accuracy.

Table 1a

Features before and after filtering.

Full set of features (21 features)	baseline_value, acceleration, fetal movement, uterine_contraction, light_deceleration, severe_deceleration, prolonged_decelerations, abnormal_short_term_variability, mean_value_of_short_term_variability, percentage_of_time_with_abnormal_long_term_variability, mean_value_of_long_term_variability, histogram_width, histogram_min, histogram_max, histogram_number_of_peaks, histogram_number_of_zeroes, histogram_mode, histogram_mean, histogram_median, histogram_variance, histogram_tendency
Correlation-based filtering techniques	
Spearman correlation (16 features)	baseline_value, acceleration, fetal movement, uterine_contraction, light_deceleration, severe_deceleration, prolonged_decelerations, abnormal_short_term_variability, mean_value_of_short_term_variability, mean_value_of_long_term_variability, histogram_width, histogram_min, histogram_max, histogram_mean, histogram_median, histogram_variance
Pearson correlation (17 features)	baseline_value, acceleration, fetal movement, uterine_contraction, light_deceleration, severe_deceleration, prolonged_decelerations, abnormal_short_term_variability, mean_value_of_short_term_variability, percentage_of_time_with_abnormal_long_term_variability, mean_value_of_long_term_variability, histogram_width, histogram_min, histogram_max, histogram_mean, histogram_median, histogram_variance
Kendall correlation (17 features)	baseline_value, acceleration, fetal movement, uterine_contraction, light_deceleration, severe_deceleration, prolonged_decelerations, abnormal_short_term_variability, mean_value_of_short_term_variability, percentage_of_time_with_abnormal_long_term_variability, mean_value_of_long_term_variability, histogram_width, histogram_min, histogram_max, histogram_mean, histogram_median, histogram_variance

Table 1b
Features before and after filtering.

Statistical techniques	
Mutual information (10 features)	baseline_value, acceleration, fetal movement, light_deceleration, severe_deceleration, prolonged_decelerations, abnormal_short_term_variability, mean_value_of_short_term_variability, percentage_of_time_with_abnormal_long_term_variability, mean_value_of_long_term_variability
Chi-squared (10 features)	baseline_value, acceleration, fetal movement, light_deceleration, severe_deceleration, prolonged_decelerations, abnormal_short_term_variability, mean_value_of_short_term_variability, percentage_of_time_with_abnormal_long_term_variability, mean_value_of_long_term_variability
ANOVA (10 features)	baseline_value, acceleration, fetal movement, light_deceleration, severe_deceleration, prolonged_decelerations, abnormal_short_term_variability, mean_value_of_short_term_variability, percentage_of_time_with_abnormal_long_term_variability, mean_value_of_long_term_variability
ROC-AUC (10 features)	baseline_value, acceleration, fetal movement, light_deceleration, severe_deceleration, prolonged_decelerations, abnormal_short_term_variability, mean_value_of_short_term_variability, percentage_of_time_with_abnormal_long_term_variability, mean_value_of_long_term_variability

Performance analysis using classification techniques

An experiment was carried out in which the filtered features were used to classify the fetal monitoring dataset using the kNN, SVM, DT, and Gaussian NB methods with correlation-based feature selection techniques, such as Pearson, Spearman, and Kendall correlation. The hyperparameters for the experiments were selected using the grid search algorithm, and their values are shown in **Table 2.**

Table 3 shows a comparison of the values of the precision metric for each filtering-based feature selection technique used with the Gaussian NB, DT, SVM, and kNN classification methods.

Table 2
Hyperparameters used in the experiments.

S. No	Classification technique	Hyperparameters used
1	k-nearest neighbor technique	$k = 3$, algorithm = auto, Distance = Euclidean
2	Support vector machine	kernel = poly, degree = 3, gamma = auto,
3	Decision tree	criterion = gini, splitter = random,
4	Gaussian NB	Priors = none, var_smoothing = 1e-09

Table 3
Evaluation results for precision.

	Precision (correlation techniques)				**Precision (statistical techniques)**			
Classifier (final)	**No feature selection**	**Pearson correlation**	**Spearman correlation**	**Kendall correlation**	**Mutual information**	**Chi-squared**	**ANOVA**	**ROC-AUC**
kNN	0.90	0.90	0.90	0.90	0.91	0.91	0.92	0.91
SVM	0.90	0.90	0.90	0.90	0.89	0.89	0.90	0.90
Decision tree	0.91	0.91	0.92	0.91	0.92	0.91	0.91	0.92
Gaussian NB	0.87	0.87	0.87	0.88	0.88	0.88	0.88	0.87

Fig. 2 visualizes the results in **Table 3**. From the graphs, it can be seen that the precision of the Gaussian NB method was improved when statistical feature selection techniques were applied, compared to the results from the correlation-based filtering techniques. For the DT, SVM, and kNN methods, the precision of the correlation-based and statistical-based filtering techniques remained the same.

Table 4 presents the results for the recall metric for the Gaussian NB, DT, SVM, and kNN classification methods with filter-based feature selection techniques.

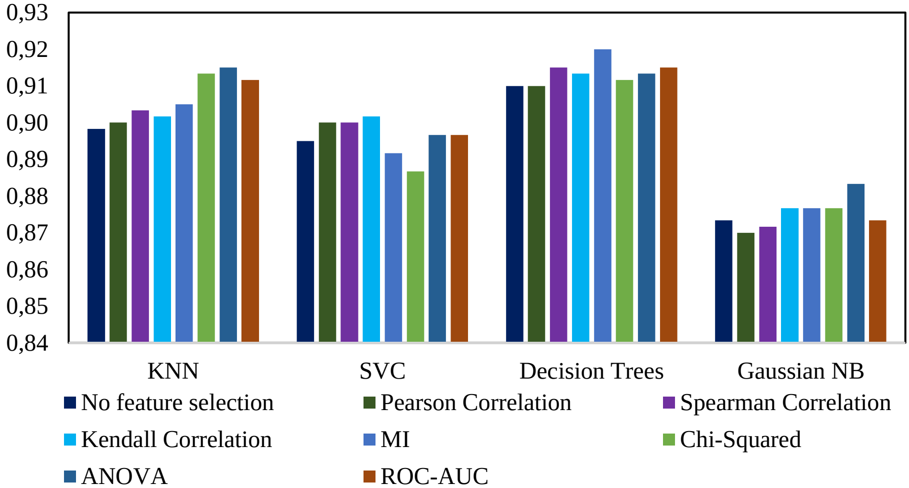

Fig. 2 – *Precision of various filtering-based feature selection techniques.*

Table 4
Evaluation results for recall.

	Recall (correlation techniques)				**Recall (statistical techniques)**			
Classifier (final)	**No feature selection**	**Pearson correlation**	**Spearman correlation**	**Kendall correlation**	**Mutual information**	**Chi-squared**	**ANOVA**	**ROC-AUC**
kNN	0.90	0.90	0.91	0.91	0.91	0.92	0.92	0.92
SVM	0.90	0.90	0.91	0.91	0.89	0.89	0.90	0.90
Decision tree	0.91	0.91	0.91	0.92	0.92	0.91	0.91	0.91
Gaussian NB	0.76	0.75	0.70	0.74	0.83	0.83	0.85	0.82

The results are represented using a bar diagram in Fig. 3. The graphs show that the recall was improved for the Gaussian NB method in statistical technique compared to that of the filtering technique. The recall for the DT method was 92% when the Kendall correlation and MI feature selection techniques were used.

For the SVM method, the recall metric showed good performance when correlation-based techniques were applied, while for kNN, the performance improved when statistical-based feature selection techniques were used.

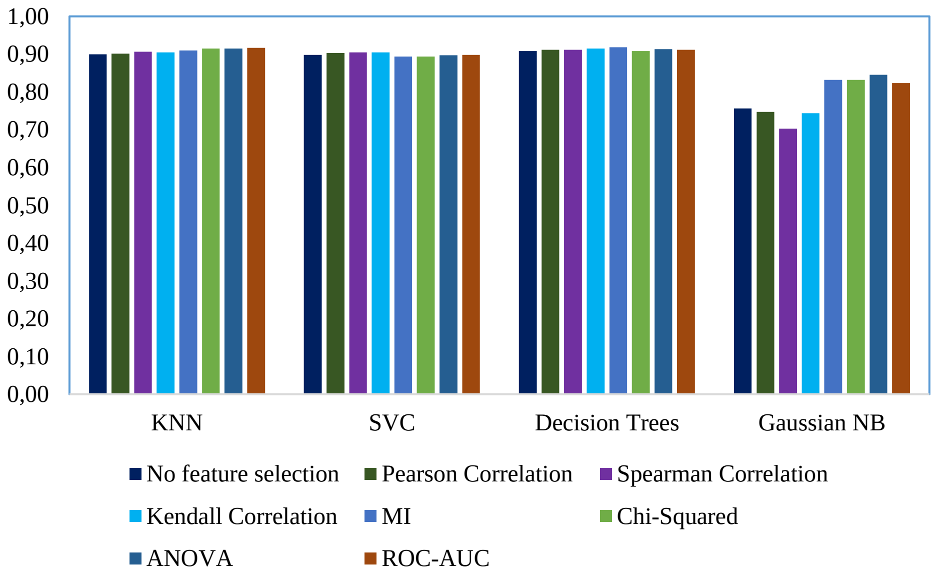

Fig. 3 – *Recall results for various filtering-based feature selection techniques.*

Table 5 shows the F1-score results for the filtering-based feature selection techniques on the classification task using the Gaussian NB, DT, SVM, and kNN methods. A comparison is also shown graphically in Fig. 4. The F1-score for the Gaussian NB method was improved when statistical techniques were used, compared to correlation-based feature selection techniques.

DT classification showed an improved F1-score with the Spearman correlation, Kendall correlation, and MI methods. The values of the F1-score were slightly lower for the MI and chi-squared feature selection techniques compared with SVM. For the kNN classification method, the F1-score was better when statistical feature selection techniques were used compared to correlation-based feature selection techniques.

Accuracy is an important performance metric when evaluating machine learning algorithms. **Table 6** summarizes the accuracy results for each classification method when correlation-based and statistical-based feature selection techniques were applied. A comparison of the experimental results is shown in Fig. 5. The accuracy of the Gaussian NB technique showed improvement when statistical techniques were used compared to correlation-based filtering techniques.

Table 5
Evaluation results for F1-score.

	F1-score (correlation techniques)				F1-score (statistical techniques)			
Classifier (final)	**No feature selection**	**Pearson correlation**	**Spearman correlation**	**Kendall correlation**	**Mutual information**	**Chi-squared**	**ANOVA**	**ROC-AUC**
kNN	0.90	0.90	0.90	0.90	0.91	0.91	0.91	0.91
SVM	0.90	0.90	0.90	0.90	0.89	0.89	0.90	0.90
Decision tree	0.91	0.91	0.92	0.92	0.92	0.91	0.91	0.91
Gaussian NB	0.78	0.78	0.74	0.78	0.85	0.84	0.86	0.84

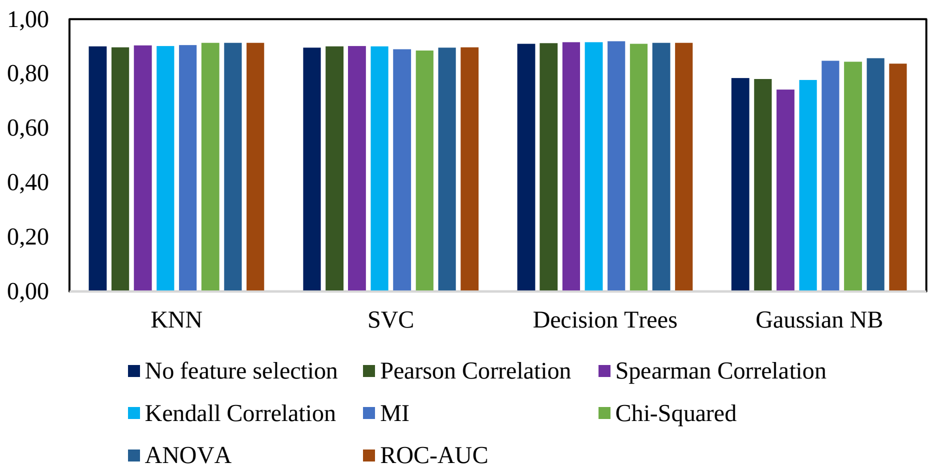

Fig. 4 – *F1-score results for various filtering-based feature selection techniques.*

The results show an accuracy of 85% for the ANOVA-based statistical technique. Similarly, the accuracy of the kNN classification method improved when statistical-based filtering techniques were applied, compared to correlation-based techniques. The accuracy reached 92% when chi-squared, ANOVA, and ROC-AUC-based feature selection filtering was used. The results for SVM showed an accuracy of 92% when the Spearman correlation method was used, which was better than the other feature selection techniques. The DT method achieved accuracy values of 91% and 92% when the filtering-based feature selection techniques were applied.

Table 6
Evaluation results for accuracy.

	Accuracy (correlation techniques)				Accuracy (statistical techniques)			
Classifier (final)	**No feature selection**	**Pearson correlation**	**Spearman correlation**	**Kendall correlation**	**Mutual information**	**Chi-squared**	**ANOVA**	**ROC-AUC**
***k*NN**	0.91	0.91	0.91	0.91	0.91	0.92	0.92	0.92
SVM	0.90	0.90	0.92	0.90	0.89	0.89	0.9	0.90
Decision tree	0.92	0.92	0.91	0.92	0.92	0.91	0.91	0.91
Gaussian NB	0.76	0.76	0.70	0.74	0.83	0.83	0.85	0.84

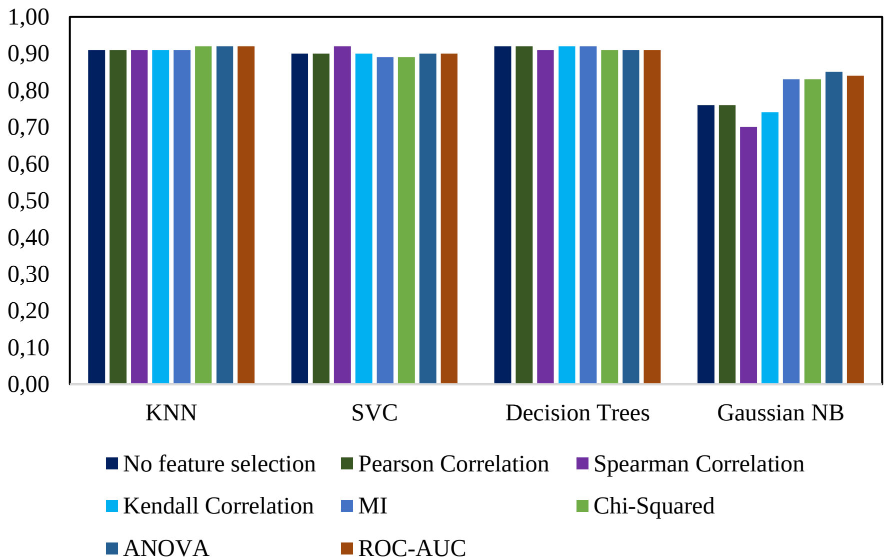

Fig. 5 – *Accuracy results for various filtering-based feature selection techniques.*

Summary

Our experimental results show that the precision of Gaussian NB classification was improved by applying statistical feature selection techniques, but did not change for correlation-based filtering techniques. In regard to the recall metric, statistical-based filtering techniques gave improved results for the

Gaussian NB and kNN methods, while correlation-based filtering techniques improved the performance of SVM and DT.

The F1-score was improved for DT, but the SVM classifier gave lower performance when a correlation-based technique was used. Gaussian NB and kNN showed improved performance with the statistical feature selection method. Correlation-based techniques improved the accuracy of DT and SVM, whereas statistical-based filtering techniques improved the accuracy of Gaussian NB and kNN.

Directions for future research

Our results show that filtering-based feature selection techniques achieved a maximum accuracy of 92% for the DT classification method. Some research directions for improving the performance of these classification techniques further include:

1. The incorporation of wrapper-based, embedded-based, and hybrid feature selection techniques;
2. Enhancing the performance by using bagging and boosting techniques for classification;
3. Adopting deep learning-based time series classification techniques for larger datasets;
4. Devising suitable optimization methods to find the best feature selection technique for the given dataset by iteration, so that the selected features can be used in machine learning classification.

5 Conclusion

In this paper, we have analyzed several filtering-based feature selection techniques using the kNN, SVM, DT, and Gaussian NB classification methods. The performance metrics of precision, recall, F1-score, and accuracy were investigated. The results of our experiments show that Gaussian NB and kNN were improved by 3% when statistical feature selection techniques were applied, and the performance of DT and SVM showed a 4% improvement when correlation-based techniques were used. The statistical techniques of ANOVA and ROC-AUC improved the accuracy by 92%. Similarly, Spearman correlation gave improved performance metrics compared to the other correlation techniques.

In future work, the performance of each classification technique could be improved using a wrapper, an embedded feature selection technique, bagging and boosting, or a deep learning-based classification method.

6 References

[1] S. B. Kotsiantis, I. D. Zaharakis, P. E. Pintelas: Machine Learning: A Review of Classification and Combining Techniques, Artificial Intelligence Review, Vol. 26, No. 3, November 2006, pp. 159 – 190.

[2] J. Qiu, Q. Wu, G. Ding, Y. Xu, S. Feng: A Survey of Machine Learning for Big Data Processing, EURASIP Journal on Advances in Signal Processing, Vol. 2016, May 2016, pp 67.

[3] M. Fatima, M. Pasha: Survey of Machine Learning Algorithms for Disease Diagnostic, Journal of Intelligent Learning Systems and Applications, Vol. 9, No. 1, February 2017, pp. 1 – 16.

[4] M. S. Mahdavinejad, M. Rezvan, M. Barekatain, P. Adibi, P. Barnaghi, A. P. Sheth: Machine Learning for Internet of Things Data Analysis: A Survey, Digital Communications and Networks, Vol. 4, No. 3, August 2018, pp. 161 – 175.

[5] S. Boudet, A. Houzé de l'Aulnoit, R. Demailly, A. Delgranche, L. Peyrodie, R. Beuscart, D. Houzé de l'Aulnoit: Fetal Heart Rate Signal Dataset for Training Morphological Analysis Methods and Evaluating them Against an Expert Consensus, Preprints, July 2019, pp. 1 – 3.

[6] Wikipedia: Concordant Pairs, Available at: https://en.wikipedia.org/wiki/Concordant_pair.

[7] P. Borah, H. A. Ahmed, D. K. Bhattacharyya: A Statistical Feature Selection Technique, Network Modeling Analysis in Health Informatics and Bioinformatics, Vol. 3, No. 1, December 2014, pp. 55.

[8] J. Neumann, C. Schnörr, G. Steidl: Combined SVM-Based Feature Selection and Classification, Machine Learning, Vol. 61, No. 1-3, November 2005, pp. 129 – 150.

[9] S. L. Shiva Darshan, C. D. Jaidhar: Performance Evaluation of Filter-Based Feature Selection Techniques in Classifying Portable Executable Files, Procedia Computer Science, Vol. 125, 2018, pp. 346 – 356.

[10] A. Subasi, B. Kadasa, E. Kremic: Classification of the Cardiotocogram Data for Anticipation of Fetal Risks Using Bagging Ensemble Classifier, Procedia Computer Science, Vol. 168, 2020, pp. 34 – 39.

[11] J. Singh, S. Bagga, R. Kaur: Software-Based Prediction of Liver Disease with Feature Selection and Classification Techniques, Procedia Computer Science, Vol. 167, 2020, pp. 1970 – 1980.

[12] N. P. N. Sreedharan, B. Ganesan, R. Raveendran, P. Sarala, B. Dennis, R. Rajakumar Boothalingam: Grey Wolf Optimisation-Based Feature Selection and Classification for Facial Emotion Recognition, IET Biometrics, Vol. 7, No. 5, September 2018, pp. 490 – 499.

[13] J. P. Li, A. Ul Haq, S. Ud Din, J. Khan, A. Khan, A. Saboor: Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare, IEEE Access, Vol. 8, June 2020, pp. 107562 – 107582.

[14] O. Cigdem, H. Demirel: Performance Analysis of Different Classification Algorithms Using Different Feature Selection Methods on Parkinson's Disease Detection, Journal of Neuroscience Methods, Vol. 309, November 2018, pp. 81 – 90.

[15] D. Lavanya, K. U. Rani: Analysis of Feature Selection with Classification: Breast Cancer Datasets, Indian Journal of Computer Science and Engineering, Vol. 2, No. 5, October 2011, pp. 756 – 763.

[16] Akshat Dubey: Fetal Cardiotocography Data, Available at: https://www.kaggle.com/akshat0007/fetalhr

[17] R. M. Grivell, Z. Alfirevic, G. M. L. Gyte, D. Devane: Antenatal Cardiotocography for Fetal Assessment, Cochrane Database of Systematic Reviews, Vol. 2015, No. 9, September 2015, p. CD007863.