JOIN (Jurnal Online Informatika)
Volume 5 No. 2 | December 2020: 153-160
DOI: 10.15575/join.v5i2.656
p-ISSN: 2528-1682
e-ISSN: 2527-9165

Genetic Algorithm to Optimize k-Nearest Neighbor Parameter for Benchmarked Medical Datasets Classification

Rizki Tri Prasetio[1]
[1]Department of Information Technology, Universitas Adhirajasa Reswara Sanjaya, Indonesia

Article Info

Article history:

Received October 24, 2020
Revised November 23, 2020
Accepted December 03, 2020
Published December 30, 2020

Keywords:

Feature Selection
Genetic Algorithm
k-Nearest Neighbor
Parameter Optimization

ABSTRACT

Computer assisted medical diagnosis is a major machine learning problem being researched recently. General classifiers learn from the data itself through training process, due to the inexperience of an expert in determining parameters. This research proposes a methodology based on machine learning paradigm. Integrates the search heuristic that is inspired by natural evolution called genetic algorithm with the simplest and the most used learning algorithm, k-nearest Neighbor. The genetic algorithm were used for feature selection and parameter optimization while k-nearest Neighbor were used as a classifier. The proposed method is experimented on five benchmarked medical datasets from University California Irvine Machine Learning Repository and compared with original k-NN and other feature selection algorithm i.e., forward selection, backward elimination and greedy feature selection. Experiment results show that the proposed method is able to achieve good performance with significant improvement with p value of t-Test is 0.0011.

Corresponding Author:

Rizki Tri Prasetio,
Department of Information Technology,
Universitas Adhirajasa Reswara Sanjaya,
Jl. Sekolah Internasional No. 1-2 Antapani Bandung, Indonesia
Email: rizki@ars.ac.id

1. INTRODUCTION

Recently, the application of machine learning in medical purposes is a major in demand for medical applications. Mostly, diagnosis methods in medical field are based on data classification approaches and systematized [1]. Use of Computer-Aided Diagnosis (CAD) systems can assist doctors to diagnose patient illnesses [2], Classification is the most commonly performed CAD system among the various tasks that can be performed [3]. In an effort to ensure accurate diagnostic assistance, the main problems with the classification of medical datasets can be categorized as complex optimization problems [1].

Research that seeks to optimize both data and algorithms with the aim of increasing the accuracy of data classification to identify potential patients has been carried out by various researchers [4]. In the recent studies, metaheuristic algorithms such as particle swarm optimizations [1] [5] [6] or genetic algorithms [7] [8] [9] and also data mining techniques such as neural networks [10] [11] [12] or *k*-nearest Neighbor [13] [14] [15] were applied to perform classification of medical data and obtained with very satisfy results.

k-Nearest Neighbor (*k*-NN) algorithm is a method that uses a supervised Algorithm (Wu, et al., 2008). Which is a classification method that is easy to understand and implement [16] and simplest amongst of all machine learning algorithms (Gorunescu, 2011). The closest object (*k*) around a classified object is a representation of the k-NN algorithm [17]. A dataset that has multimodal classes is very suitable for implementing the k-NN algorithm [18] as well as applications where many class labels on single object [16].

There are several major problems that affect the performance of the k-NN algorithm. One of them is the selection of the *k* parameter [16]. The result can be sensitive to noise points, if *k* is too small, that may lead the algorithm toward overfitting [19]. On the other hand, the neighborhood may include too many points from other classes if *k* is too large, that may lead to low accuracy [20]. Selection of *k* based on data is the best choice [21].

e-ISSN 2528-1682
p-ISSN 2527-9165

Another key issue is the presence of noise or irrelevant features in medical datasets which can greatly degrade the accuracy of the k-NN algorithm [22], or inconsistency is found between the scale of the features and its importance [21].

High dimensional data commonly involved in medical dataset [23]. Classification complexity will increase and reduce the effect of the model when using high-dimensional data [24], efficiency of most machine learning algorithms will deal with a serious obstacle. “Curse of dimensionality”, is a term to this obstacle [25]. While retaining important information, the data dimension needs to be reduced. The major keys of dimensional reduction are feature extraction [26] and feature selection [27]. High computational costs are required during the data mining process to handle large data sets. Effectively reduce time and memory [28] and cut computing costs when reducing dimensions [25].

Reducing the number of features while maintaining acceptable classification accuracy is the main goal of dimension reduction [7]. The effectiveness of the resulting classification algorithm is very influenced by the feature selection [29]. In some cases, the accuracy of future classification can be improved based on the result of feature selection [25].

Given optimized a subset and a set of candidate features that performs the best under classification system is the problem of feature selection [29]. To perform optimization, genetic algorithms are often used. Genetic algorithms are sophisticated optimization [30] and have less of a tendency to become stuck in local minima [21]. In machine learning, to evaluate the fitness of other algorithms, genetic algorithms may be used [22]. Genetic algorithms are easily parallelizable and have been used for classification as well as other optimization problems and a wide range of optimization depending on the objective function (fitness) [31].

In this research, we integrates genetic algorithm for features selection and parameters optimized *k*-NN applies to classify five benchmarked medical datasets, namely, Wisconsin breast cancer diagnostic and prognostic [32], diabetic retinopathy Debrecen [33], cardiotocography [34] and SPECTF image of heart disease [35]. There are several reasons to choose the Genetic Algorithm: The 'universal optimizer' as a capability of the genetic algorithm can be used to optimized problems in various fields, genetic algorithms can set parameters correctly through the right balance between exploration and exploitation. And the main feature is the simplest and easiest to implement. Main objectives of this research are to improve accuracy of five benchmarked medical datasets classification by applying genetic algorithm as feature selection and to improve performance of *k*-NN classifier algorithm by optimizing *k* value using genetic algorithm.

2. METHODS

This research proposes a methodology based on data mining paradigm. This paradigm integrates the search heuristic that is inspired by natural evolution called genetic algorithm with the simplest and the most used learning algorithm, *k*-nearest Neighbor

2.1. Datasets

This research is experimented on five medical datasets obtained from University California Irvine (UCI) Machine Learning (https://archive.ics.uci.edu/ml/datasets.html). The details of these medical datasets is listed in Table 1 that contains number of instances, features and classes. The training and testing datasets are randomly generated.

1. Wisconsin Breast Cancer (Diagnostic), the dataset is available at the University of Wisconsin. It contains 569 instances with 32 features which are used to predict benign or malignant growths [32].
2. Wisconsin Breast Cancer (Prognostic), the dataset is obtained from University of Wisconsin. There are 198 instances with 20 features which are used to predict recurrent and nonrecurrent [32].
3. Diabetic Retinopathy, this dataset was collected from University of Debrecen and contains about 1151 instances with 20 features which are used to predict whether it is contain diabetic retinopathy or not [33].
4. Cardiotocography (CTGs), this dataset was created by Diogo Ayres-de-campos at the University of Porto. It contains 2126 instances with 23 features which are used to predict fetal state [34].
5. Heart Disease (SPECTF), this dataset is based on data from University of Colorado. It contains 45 features with 267 instances which are used to identify whether patients are normal or not [35].

Table 1. Description of datasets

Dataset	Number of instances	Number of features	Number of classes
Wisconsin Breast Cancer (Diagnostic)	569	32	2
Wisconsin Breast Cancer (Prognostic)	198	34	2
Diabetic Retinopathy Debrecen	1151	20	2
Cardiotocography (CTGs)	2126	23	3
Heart Disease (SPECTF)	267	44	2

Algorithm 1. Basic Genetic Algorithm

Begin
INITIALIZE random candidate solutions within population;
EVALUATE each candidate;
WHILE (stop condition is satisfied) *DO*
SELECT chromosome;
RECOMBINE pairs of chromosome;
MUTATE offspring;
EVALUATE new candidates (offspring);
SELECT individual candidate for next generation;
End

2.2. Genetic Algorithm

Genetic algorithm (GA) is a heuristic, parallel and stochastic, parallel search algorithm inspired by Charles Darwin that introduced principle of natural selection [36]. Holland were the first researcher that propose GA for the very first time [37]. Mimicking a computational process whereby natural selection through biological processes is the basis of the concept of genetic algorithms. Through this process the stronger individual is more likely to become the winner in a competitive environment [38] and implement them to solve research and business problems [39].

Mate selection, reproduction, mutation and cross-cutting of genetic information are factors that have inspired genetic algorithm frameworks. The following three factors are used by genetic algorithms: [21]

1. *Selection*
   Choosing which chromosome to reproduce is the selection operator's reference. Each chromosome (candidate solution) will be evaluated using a suitability function, Possibility of being selected to reproduce depends on the quality of the chromosomes, the better the chromosomes, the more likely the candidate will be selected.
2. *Crossover*
   Creates two new offspring by selecting a random locus and exchanging the order left and right of that locus between the two chromosomes selected during selection. this process is called recombination which is done by the crossover operator. For example, in a binary representation, two strings 00000000 and 11111111 can be crossed at the sixth locus of the string respectively to produce two new offspring 00000111 and 11111000.
3. *Mutation*
   The bits or digits at a particular locus on a chromosome are randomly changed by mutation operators. For example, after crossing each other, a binary string can mutate at locus two to 10111000 from the original binary string 11111000. This stage of mutation avoids the incorporation of prematurely into the local optimum and provides new information on the genetic pool.

2.3. *k*-Nearest Neighbor

k-Nearest Neighbor (*k*-NN) algorithm is a method that uses a supervised Algorithm (Wu, et al., 2008). Which is simplest [21], can be used for prediction and estimation, but is most often used for classification. k-NN algorithm is instance-based learning, where unclassified data can be found by simply comparing it with the most similar data in the training set [39].

k-NN classifies objects based on the nearest *k* number of objects around them [21] then determining labels based on the dominance of certain classes in its environment [16]. Parameter *k* comes from most similar pieces of data from our known dataset[17]. To conclude, the *k*-NN is the simplest among all machine learning algorithms, simple majority vote of its neighbor is used to classifying an object [21] from the k most similar pieces of data [17].

Distance metric use to defined “Closeness” between object with its neighborhood, such as Euclidean distance or Manhattan distance [22]. To construct the algorithm, we need the following items (algorithm input):

e-ISSN 2528-1682
p-ISSN 2527-9165

Algorithm 2. Basic *k*-Nearest Neighbor Algorithm

Begin
INITIALIZE *D*, training set; *T*, test object; *k*, neighborhood;
FOREACH *d* IN *D* DO
COMPUTE distance between *T* and *d*;
SORT the computed distance ascending;
SELECT *k* nearest object corresponding to *k* nearest distance;
EVALUATE most frequent class label among *k* nearest object;
End

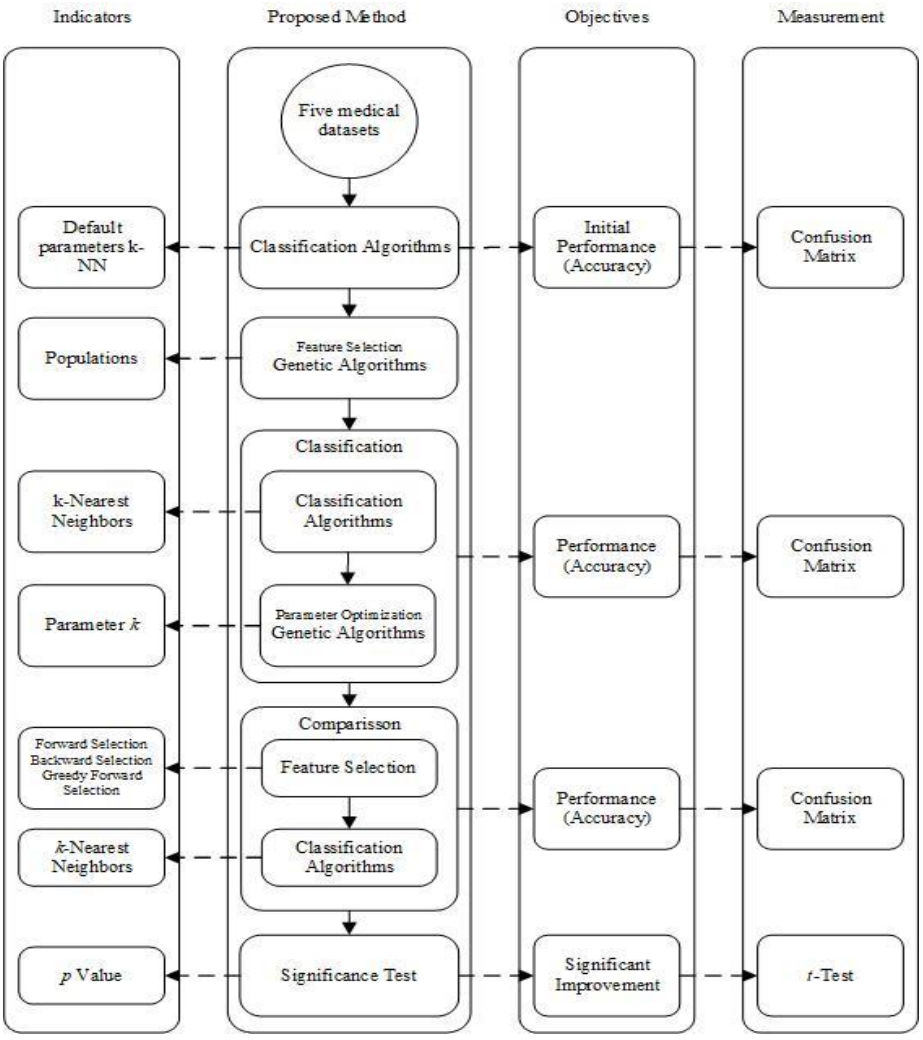

Figure 1. Proposed Method

1. Training set consist of labeled data [21] used for evaluating an unlabeled data[16];
2. Distance metric to calculate similarity between objects [21] used to compute the closeness of objects [16];
3. Value of *k*, the number of neighborhood [16] belonging to the training set, based on which we will achieve the classification of a new object [21];
4. A Method to determine the class of the target object based on the classes and distances of the *k* nearest neighbor [16].

By using the following three requirement of the algorithm, a new (unlabeled data) object will be classified: [21]

1. Between new object and every records in training set, compute distance (similarity) using distance metrics;
2. Sort the distance from every records in training set based on shortest distance, identify *k* nearest object.
3. Find label that appears most frequently among *k* nearest object, assign the label using majority voting.

2.4. Proposed Method

The proposed method integrates genetic algorithm for features selection and parameters optimized *k*-NN applies to classify five benchmarked medical dataset that explained in Table 1. The proposed method can be seen in Figure 1. An early data processing begins by dividing five datasets into training and testing data using split validation, respectively. *k*-NN with default parameters is applied for each training data to results initial performance.

Genetic algorithms are applied for each training data for features selection. Features selection is used to find the features that best represents the class on that dataset. Parameters optimized *k*-NN then applied to training data that has been feature selected. After that, validate the models which are produced by *k*-NN, calculate how much accuracy generated by the model tested in testing data. Repeat the process of selecting features using genetic algorithms, until optimal features are obtained.

The results obtained from the performance of the features selection by genetic algorithm are then compared with other algorithms that can be used for feature selection i.e., backward elimination [40] [41] [42], forward selection [43] [44] [29] and greedy feature selection [45] [46] [47].

Table 2. Experiment Results of Proposed Method and Basic *k*-NN

Datasets	Proposed Method	Basic *k*-NN
breast-cancer (D)	99.2.%	94.15%
breast-cancer (P)	86.44%	78.75%
diabetic-retinopathy	71.69%	61.16%
cardiotocography	98.59%	90.91%
heart (SPECTF)	87.5%	77.5%

This comparison is to determine whether performance of genetic algorithms is better than any other algorithms in performing feature selection. The results obtained from proposed method then tested with results obtained from *k*-NN with default parameters to determine whether the proposed method performance results improved the accuracy of the five datasets significantly using a t-test [13] [48] [31] significance test.

3. RESULT AND DISCUSSION

This research conducted several experiments, experiments using the *k*-NN algorithm with unoptimized parameters of the five unselected features datasets, experiments using the *k*-NN algorithm with optimized parameters of the five datasets in Table 1 that have not been selected features dan experiments using the *k*-NN algorithm with optimized parameters of five datasets in Table 1 which have been selected feature using genetic algorithm, backward elimination, forward selection dan greedy feature selection.

All experiments use split validation to split the datasets randomly. The experiment using default parameters configuration for genetic algorithm, backward elimination, forward selection and greedy forward selection. The experimental results set forth in Table 2 stated that the proposed method can improve the accuracy of the five benchmarked datasets with a 5% - 10% increase in comparison with the *k*-NN algorithm without optimization and features selection.

Highest improved performance was obtained from the classification of the Diabetic Retinopathy dataset with an increase of 10.53% of 61.16% with the most optimal *k* is 86. Meanwhile, the lowest improved performance was obtained from the classification of Breast Cancer Diagnostic dataset with only 5.05% increase from 94.15% with the optimal k is 8.

Improved performance on Breast Cancer Prognostic dataset is 7.69% from 78.75% with optimal *k* is 57, Cardiotocography datasets increased by 7.68% from the original 90.91% with optimal *k* is 23 dan SPECTF Heart dataset increased by 10% from 77.5% with the most optimal k is 23. Based on experiment results in this research, t-Test were used to determine proposed method can improve classification significantly. *t*-Test Paired Two Sample for Means were used in results between before and after using proposed method.

The test results of *t*-Test generate that the proposed method can improve the performance of *k*-NN in terms of accuracy significantly in all datasets marked with *p* value of *t*-Test < 0.05. *t*-Test results can be seen in Table 4. The results of the experiments described in Table 3 stated that the proposed method is superior when compared to other feature selection algorithms across all benchmarked datasets. The results on backward elimination and forward selection were slightly lower 0.37% - 5.96% when compared to genetic algorithm, and the lowest results obtained by greedy feature selection. Based on experiment results, to determine whether feature selection can improve performance in the classification of medical datasets significantly. *t*-Test Paired Two Sample for Means were used in results obtained from all features selection algorithms.

The test results of *t*-Test generate that the feature selection can improve the performance of *k*-NN in terms of accuracy significantly in all datasets except greedy feature selection marked with *p* value of *t*-Test < 0.05. *t*-Test results for significance of using features selection can be seen in Table 5.

e-ISSN 2528-1682
p-ISSN 2527-9165

Table 3: Experiment Results of Featured Selection *k*-NN

Datasets	Proposed Method	Forward Selection	Backward Selection	Greedy Selection
breast-cancer (D)	99.2%	98.83%	97.08%	92.4%
breast-cancer (P)	86.44%	84.75%	83.05%	79.66%
diabetic-retinopathy	71.69%	68.99%	69.28%	68.12%
cardio-tocography	98.59%	91.22%	92.63%	79.78%
heart (SPECTF)	87.5%	86.25%	85%	82.5%

Table 4: *t*-Test Results of Proposed Method compared with *k*-NN

	Proposed Method	Normal
Mean	88.684	80.494
Variance	125.98713	170.19713
Observations	5	5
Pearson Correlation	0.995007081	
df	4	
t Stat	8.376046049	
P(T<=t) one-tail	0.000555628	
t Critical one-tail	2.131846786	
P(T<=t) two-tail	0.001111256	
t Critical two-tail	2.776445105	

k-nearest Neighbor algorithm is easy to implement [21] and high accuracy [17] for a variety of applications. Compared to any other complex algorithms like neural network and support vector machine, k-NN results still remarkable. From the results of this research, it can be concluded that parameter optimized *k*-NN combine with genetic algorithms as feature selection is superior when compared to other feature selection algorithms on five benchmarked medical datasets.

Table 5: *t*-Test Results of Featured Selection Classifier compared with *k*-NN

Algorithms	*p* Value of *t*-Test	Results
Genetic Algorithm	0.0011	Significant ($p<0.05$)
Forward Selection	0.02	Significant ($p<0.05$)
Backward Elimination	0.01	Significant ($p<0.05$)
Greedy Feature Selection	0.99	Not Significant ($p>0.05$)

4. CONCLUSIONS

Genetic algorithms are applied to select features and optimizing *k* parameter for *k*-nearest Neighbor to improve accuracy of five benchmarked medical datasets. Proposed method is proven effective to be able improve accuracy, and furthermore the different test results among five datasets produce significant difference.

Comparison of the feature selection algorithms are proposed to compare the accuracy of the results among genetic algorithms, forward selection, backward elimination and greedy feature selection. Genetic algorithms are proven to have the highest accuracy compared with any others feature selection algorithms.

In this research, in general, genetic algorithms applied to select features and optimizing parameters to improve accuracy of five benchmarked medical datasets. In further research, some things can be applied to enhance the research, which uses other algorithms for parameter optimizing or other methods to reduce dimensionality of medical datasets.

ACKNOWLEDGEMENTS

This study is supported by The Ministries of Research, Technology, And Higher Education of Republic Indonesia.

REFERENCES

[1] C. V. Subbulakhsmi and S. N. Deepa, "Medical Dataset Classification: A Machine Learning Paradigm Integrating Particle Swarm Optimization with Extreme Learning Machine Classifier," *The Scientific World Journal*, vol. 2015, pp. 1-12, 2015.

[2] Y. Unal and E. Kocer, "Diagnosis of Pathology on the Vertebral Column with Backpropagation and Naive Bayes Classifier," Turkey, 2013.

[3] R. T. Prasetio and E. Ripandi, "Optimasi Klasifikasi Jenis Hutan Menggunakan Deep Learning Berbasis Optimize Selection," *Jurnal Informatika*, pp. 100-106, 2019.

[4] G. S. Babu and S. Suresh, "Meta-cognitive RBF network and its projection based learning algorithm for classification problems," *Applied Soft Computing Journal,* vol. 13, no. 1, pp. 654-666, 2013.

[5] H. H. Inbarani, A. T. Azar and G. Jothi, "Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis," *Computer Methods and Programs in Biomedicine,* vol. 113, no. 1, pp. 175-185, 2014.

[6] P.-C. Chang, J.-J. Lin and C.-H. Liu, "An attribute weight assignment and particle swarm optimization algorithm for medical database classifications," *Computer Methods and Programs in Biomedicine,* vol. 107, no. 3, pp. 382-392, 2012.

[7] M. L. Raymer, W. F. Punch, E. D. Goodman, L. A. Kuhn and A. K. Jain, "Dimensionality reduction using genetic algorithms," *IEEE Transactions on Evolutionary Computation,* vol. 4, no. 2, pp. 164-171, 2000.

[8] J. Yang and V. Honavar, "Feature Subset Selection Using a Genetic Algorithm," *Feature Extraction, Construction and Selection,* pp. 117-136, 1998.

[9] S. Shah and A. Kusiak, "Cancer gene search with data-mining and genetic algorithms," *Computers in Biology and Medicine,* vol. 37, no. 2, pp. 251-261, 2007.

[10] M. A. Mazurowski, P. A. Habas, J. M. Zurada, J. Y. Lo, J. A. Baker and G. D. Tourassi, "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance," *Neural Networks,* vol. 21, no. 2, pp. 427-436, 2008.

[11] M. Brameier and W. Banzhaf, "A comparison of linear genetic programming and neural networks in medical data mining," *IEEE Transactions on Evolutionary Computation ,* vol. 5, no. 1, pp. 17-26, 2001.

[12] F. Amato, A. Lopez, E. M. Pena-Mendez, P. Vanhara, A. Hampi and J. Havel, "Artificial neural networks in medical diagnosis," *Journal of Applied Biomedicine,* vol. 11, no. 2, pp. 47-58, 2013.

[13] R. T. Prasetio and Pratiwi, "Penerapan Teknik Bagging pada Algoritma Klasifikasi untuk Mengatasi Ketidakseimbangan Kelas pada Dataset Medis," *Informatika,* vol. 2, no. 2, pp. 395-403, 2015.

[14] N. Suguna and K. Thanushkodi, "An Improved k-Nearest Neigbor Classification Using Genetic Algorithm," *IJCSI International Journal of Computer Science,* vol. 7, no. 2, pp. 18-44, 2010.

[15] M. A. Jabbar, B. L. Deekshatulu and P. Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm," *Procedia Technology,* vol. 10, pp. 85-94, 2013.

[16] X. Wu and V. Kumar, The Top Ten Algorithms in Data Mining, Boca Raton: Taylor & Francis Group, LLC, 2009.

[17] P. Harrington, Machine Learning in Action, New York: Manning Publication, 2012.

[18] R. T. Prasetio, A. A. Rismayadi and I. F. Anshori, "Implementasi Algoritma Genetika pada k-nearest neighbours untuk Klasifikasi Kerusakan Tulang Belakang," *Jurnal Informatika,* pp. 186-194, 2018.

[19] D. T. Larose, Discovering Knowledge in Data: An Introduction to Data Mining, New Jersey: John Wiley & Sons, Inc., 2005.

[20] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh and Q. Yang, "Top 10 Algorithms in Data Mining," Springer-Verlag, London, 2008.

[21] F. Gorunescu, Data Mining: Concepts, Models, and Techniques, Verlag Berlin Heidelberg: Springer, 2011.

[22] J. Han, M. Kamber and J. Pei, Data Mining Concepts and Techniques, San Fransisco: Morgan Kauffman, 2012.

[23] R. T. Prasetio and S. Susanti, "Prediksi Harapan Hidup Pasien Kanker Paru Pasca Operasi Bedah Toraks Menggunakan Boosted k-Nearest Neighbor," *JURNAL RESPONSIF: Riset Sains & Informatika,* pp. 64-69, 2019.

[24] K. K. Bharti and P. K. Singh, "A three-stage unsupervised dimension reduction method for text clustering," *Journal of Computational Science,* vol. 5, no. 2, pp. 156-169, 2014.

[25] O. Maimon and L. Rokach, Data Mining and Knowledge Discovery Handbook, Second Edition ed., New York: Springer, 2010.

[26] Z. Liu, T. Chai and J. Tang, "Multi-frequency signal modeling using empirical mode decomposition and PCA with application to mill load estimation," *Neurocomputing,* vol. 169, pp. 392-402, 2015.

[27] T. Jirapech-Umpai and S. Aitken, "Feature selection and classification for microarray data analysis: evolutionary methods for identifying predictive genes," *BMC Bioinformatics,* vol. 6, p. 148, 2005.

[28] S. Shilaskar and A. Ghatol, "Dimensionality Reduction Techniques for Improved Diagnosis of Heart Disease," *International Journal of Computer Applications* , vol. 61, no. 5, pp. 1-8, 2013.

[29] A. Jain and D. Zongker, "Feature Selection: Evaluation, Application and Small Sample Performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 19, no. 2, pp. 153-158, 1997.

[30] I. H. Witten, E. Frank and M. A. Hall, Data Mining: Practical Machine Learning Tools and Technique, Third Edition ed., Amsterdam: Elsevier Inc., 2011.

[31] R. T. Prasetio and D. Riana, "A Comparison of Classification Methods in Vertebral Column Disorder with the Application of Genetic Algorithm and Bagging," Bandung, 2015.

[32] O. L. Mangasarian, W. N. Street and W. H. Wolberg, "Breast cancer diagnosis and prognosis via linear programming," *Operations Research,* vol. 43, no. 4, pp. 570-577, 1995.

[33] B. Antal and A. Hajdu, "An ensemble-based system for automatic screening of diabetic retinopathy," *Knowledge-Based Systems,* vol. 60, pp. 20-27, 2014.

[34] D. Ayres-de-campos, J. Bernardes, A. Garrido, J. Marques-de-Sa and L. Pereira-Leite, "SisPorto 2.0: A Program for Automated Analysis of Cardiotocograms," *The Journal of Maternal-Fetal Medicine,* vol. 9, pp. 311-318, 2000.

[35] L. A. Kurgan, K. J. Cios, R. Tadeusiewicz, M. Ogiela and L. S. Goodenday, "Knowledge discovery approach to automated Cardiac SPECT Diagnosis," *Artificial Intelligence in Medicine,* vol. 23, pp. 149-169, 2001.

[36] A. Nowe, Genetic Algorithms, Encyclopedia of Astrobiology ed., Berlin: Springer, 2014.

[37] J. H. Holland, Adaption in Natural and Artificial Systems, Cambridge: MIT Press, 1975.

[38] K. F. Man, K. S. Tang and S. Kwong, "Genetic Algorithms: Concepts and Applications," *IEEE Transactions on Industrial Electronics,* vol. 43, no. 5, pp. 519-534, 1996.

[39] D. T. Larose, Data Mining Methods and Models, New Jersey: John Wiley & Sons, Inc., 2006.

[40] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research,* vol. 3, pp. 1157-1182, 2003.

[41] S. Abe, "Modified Backward Feature Selection by Cross Validation," Bruges, 2005.

[42] S. Derksen and H. J. Keselman, "Backward, Forward and Stepwise Automated Subset Selection Algorithms," *British Journal of Mathematical and Statistical Psychology,* vol. 45, pp. 265-282, 1992.

[43] F. G. Blanchet, P. Legendre and D. Borcard, "Forward Selection of Explanatory Variables," *Ecology,* vol. 89, no. 9, pp. 2623-2632, 2008.

[44] S. Abe, Support Vector Machine for Pattern Classification, Second Edition ed., New York: Springer London, 2010.

[45] E. L. Dyer, A. C. Sankaranarayanan and R. G. Baraniuk, "Greedy Feature Selection for Subspace Clustering," *Journal of Machine Learning Research,* vol. 14, pp. 2487-2517, 2013.

[46] H. Vafaie and I. F. Imam, "Feature Selection Method: Genetic Algorithms vs Greedy-like Search," Louisville, 1994.

[47] A. K. Farahat, A. Ghodsi and M. S. Kamel, "Efficient Greedy Feature Selection for Unsupervised Learning," *Knowledge Information System,* vol. 35, pp. 285-310, 2013.

[48] T. Setiyorini and R. S. Wahono, "Penerapan Metode Bagging untuk Mengurangi Data Noise pada Neural Network untuk Estimasi Kuat Tekan Beton," *Journal of Intelligent Systems,* vol. 1, no. 1, pp. 37-42, 2015.