Available online at www.garmian.edu.krd

Passer Journal

Passer 3 (2021) 174- 179

http://passer.garmian.edu.krd/

A Comparison Study of Data Mining Algorithms for blood Cancer Prediction

Noor Bahjat Tayfor [1*], Snoor Jamal Mohammed [2]

[1] *Department of Information Technology, Kurdistan Technical Institute, Sulaimani 46001, Kurdistan Region, Iraq*
[2] *Directorate of Education Rozhawa, Sulaimani 46001, Kurdistan Region, Iraq*

Received 15 February 2021; revised 17 February 2021;
accepted 24 March 2021; available online 26 July 2021

doi:10.24271/psr.29

ABSTRACT

Cancer is a common disease that threats the life of one of every three people. This dangerous disease urgently requires early detection and diagnosis. The recent progress in data mining methods, such as classification, has proven the need for machine learning algorithms to apply to large datasets. This paper mainly aims to utilise data mining techniques to classify cancer data sets into blood cancer and non-blood cancer based on pre-defined information and post-defined information obtained after blood tests and CT scan tests. This research conducted using the WEKA data mining tool with 10-fold cross-validation to evaluate and compare different classification algorithms, extract meaningful information from the dataset and accurately identify the most suitable and predictive model. This paper depicted that the most suitable classifier with the best ability to predict the cancerous dataset is Multilayer perceptron with an accuracy of 99.3967%.

Keywords: Classification Algorithms, Prediction, Data Mining Techniques, Blood Cancer, Hematology, Weka.

1. Introduction

Cancer disease is a malignant tumour that attacks an organ or tissue cells. This dangerous disease needs early diagnosis and decision making in treatment as soon possible; otherwise, the patient's life will be under the threat if any lateness occurs. Medical data can help predict diseases because they might contain useful information that can reduce the mortality rate and enhance the patient's quality of life [1]. Meanwhile, data mining techniques have widely used in the healthcare domain, for instance: fraud detection, misuse of health insurance, medical discovery, effective treatment and the best medical practice [2].

Data mining is extensively used in several domains, such as market analysis, stock market, economy prediction, credit estimation, fraud and intrusion detection, hazard prediction, predicting consumer conduct, education system assessment, public relations and weather forecast.

Data mining has different association rule mining, classification, prediction, clustering, time series analysis and outlier analysis. The essential task in data mining is classification as it reduces medical cost and improves early disease diagnosis.

The objective of this paper is to use different data mining techniques such as Naive Bayes, Logistic Regression, Support Vector Machine and Multilayer Perceptron to classify cancerous datasets into blood cancer and non-blood cancer and assist the users in extracting vital information and determine the best algorithm for the most accurate predictive model. There have many works that aimed to classify cancer dataset into their precise type such as Leukemia. However, no prior work intends to group all kinds of blood cancer into one target label, for instance, a blood cancer.

Our datasets have been collected from Hiwa Hospital, a Governmental Cancer Hospital located in Sulaimani City-Kurdistan region of Iraq- Iraq. Our collected data belong to three different departments at Hiwa Hospital for Cancer (Oncology, Hematology and Pediatric). To achieve the goals of our research, we firstly apply classification algorithms to classify cancer dataset. Then, compare them in terms of the number of correctly classified instances, the number of incorrectly classified instances, Root Mean Squared Error value and time taken to build a predictive model. Whereas, evaluation has used to assess each classifier's effectiveness and decide which classifier performs the best effect. Those metrics are accuracy, precision, recall, f-measure and ROC curve. We have used the Weka tool to complete our comparison and evaluation processes.

This paper organises as follows: Section 2 includes Related Work, which covers some previous research conducted to predict cancer using data mining techniques. The Materials and Methods presented in details in section 3. The findings of this work covered in Results and Discussion section 4. Finally, this work

* Corresponding author
E-mail address: noor.tayfor@kti.edu.krd (Instructor).
Peer-reviewed under the responsibility of the University of Garmian.

has concluded in paragraph 5 and the future work stated in Conclusion and Future Work.

2. Related Work

Researchers, to get new knowledge, depend mainly on data mining techniques—the reason for the massive numbers of data in medical domains is widely available nowadays. This part describes the past research that deals with the problem of classifying and predicting blood cancer disease:

Data mining techniques have been applied to classify Complete Blood Count (CBC) sample of a blood disease patient. CBC sample has classified as either routine Hematology or blood cancer disease. In this study, three data mining methods have been used association rules to discover relations among variables, rule induction to find out patterns associated with blood diseases and deep learning which uses a hieratical level to train data. The best accuracy has given to deep learning classifiers with 79.45% [3].

Another study has been conducted on 13 Hematological parameters to predict the abnormality in a blood smear. The dataset of 1362 students at the age of (17-19) has been collected from the automated blood cell counter. J48 algorithm obtained the most accurate model to predict RBC morphology using Hematological parameters of four datasets (MCV, MCH, Hct and RBC) [4].

Furthermore, a comparative study has been performed using Weka tool to develop a mobile application to identify the best algorithm for users working on discovering Hematological data comments. The researchers have used three algorithms J48, Multilayer Perceptron and Naïve Bayes. The most accurate algorithm was J48 with an accuracy of 97.16% while the worst algorithm was Naïve Bayes with an accuracy of 70.28% [5].

Feature Extraction methodology from microarray genes is also considered to impact classification and clustering methods significantly because the gene takes as input. This research used gene expression data to discriminate two types of nearly similar cancers Acute Myeloid Leukemia (AML) and Acute Lymphoblastic Leukemia (ALL). The best classification results had achieved when feature selection methods used [6].

Additionally, Hemogram blood data has been tested using data mining techniques to diagnose leukaemia, inflammatory, bacterial or viral infections, HIV infection and anaemia diseases. The research has been developed a new algorithm called weight base k-means clustering algorithm to identify the mentioned diseases. The researchers discovered that the clustering algorithm weight base k-means performed better than k-means and fuzzy c-means [7].

Furthermore, Blood Cell Counter (CBC) has been tested to predict leukaemia by finding out the correlations between blood properties and leukaemia in terms of age, gender and patients status using data mining algorithms. These classification algorithms are KNN, Decision Tree (DT) and Support Vector Machine (SVM) where DT classifier outperformed the other techniques with accuracy 77.30% [8].

Likewise, the cancer type has been predicted depending on the gene expression data. In this study, fourteen classification algorithms have been evaluated by using three different cancer Microarray Gene Expression data such as Breast Cancer, Lymphoma and Leukaemia. The comparative analysis indicated that none of the applied classifiers outperformed the others in terms of accuracy [9].

Besides, a survey study has been conducted to review several works that used data mining techniques to classify and diagnose Myeloid Dysplastic Syndrome (MDS) and Acute Myeloid Leukemia (AML). Those techniques include clustering, regression, classification and prediction. This research firstly stated the importance of data mining approach. Then, providing the list of previous papers. Finally, comparing the actual accuracy of the mentioned works [10].

A new methodology has also been considered to make the use of Weka software easier, which would be utilised later in medical bioinformatics. The features used primarily are 49 data preprocessing tools, eight clustering algorithms, 15 attribute evaluations, 76 classification algorithms, three association rules and ten algorithms for feature selections. It has been concluded that Weka could be used to diagnose leukaemia, as demonstrated in the medical bioinformatics investigations [11].

However, all the mentioned research did not state or even try to group all the blood cancer types into one class or label.

3. Materials and Methods

Our proposed study includes classifying several types of cancer into blood cancer and non-cancer disease. Thus, our proposed flowchart has been designed, as illustrated in Figure 1:

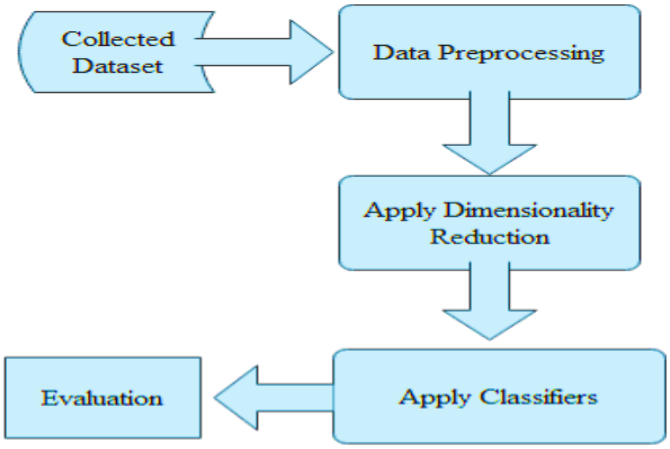

Figure 1: Data Flowchart

3. 1. Collected Dataset

Our dataset has collected from Hiwa Public Hospital for Cancer Patients, which locates in Sulaimani City/ Kurdistan Region of Iraq- Iraq. The dataset for different kinds of Cancer Disease shown in Table 1:

Table 1: Description of attributes used for predicting cancer type

Column Name/ Attributes	Meaning	Values/ Instances	Attribute Role
Gender	Patient Gender/ Sex	Male; Female	Regular
Blood Type	Patient Blood Type	A+; B+; O+; etc.	Regular
Stage	Cancer Stage	None; I; II; III; etc.	Regular
Sub Type	Cancer cell characteristics	Chronic; Plasma Cell Myeloma; etc.	Regular
Primary Site	The place where cancer starts growing up	Bone Marrow; Cervical Lymph Nodes; etc.	Regular
System	The infected system in the body	Blood, Myeloid, etc.	Regular
Topographical Code	Indicates the site of organ system of a tumour where it arose	C42.1; C77.4; etc.	Regular
Morphological Code	Describes the cell type (or histology) of the tumour	9866/3; 9673/3; etc.	Regular
Diagnosis Type	The kind of cancer	Multiple Myeloma; Breast Cancer; Lung Cancer; Chronic Lymphocytic Leukemia; etc.	Regular
Cancer Type	Blood Cancer refers to blood cancer types such as **Acute Lymphocytic Leukemia,** Multiple Myeloma; Chronic Lymphocytic Leukemia; etc. Non-Blood Cancer indicates other cancer types like Lung Cancer, Breast Cancer, Brain Cancer, etc.	Blood Cancer or Non-Blood Cancer (both labeled by us)	Label

Our dataset file is saved in (.xls) format. Although, Weka cannot read Excel file. The mentioned file has been converted to a Comma-Separated Values (.csv) format to be converted later into Attribute-Relation File Format (.arff) format easily read by Weka.

Our dataset, as shown in Table 1, includes 10-attributes and 11171 instances.

3. 2 Data Preprocessing

3. 2.1 Data Cleaning

Before starting with any experimental test, we should prepare our data and check them out carefully. We removed outliers and some columns that contain inconsistent data or unnecessary information. Also, we determined a default value (O+) for those patients with empty blood type value. Furthermore, we have removed some columns that could identify the patient, such as phone number and columns written in Kurdish language and contain numeric values.

3. 2. 2 Data Imbalance

Our collected dataset is somewhat imbalanced where the number of instances for the Non-Blood Cancer class is 9093 while Blood Cancer class comprises 2078 only. This issue leads to overfitting issue. We will utilise the Synthetic Minority Oversampling Technique method (SMOTE) to [12]. This method generates a new synthetic instance by calculating the variation between the instance's feature vector and its nearest neighbour. It will then be multiplied by a number selected randomly in a range of 0 and 1 and eventually will be added to the instance. The number of cases will expand to 17405 after employing two iterations where 9093 for Non-Blood Cancer class and the remaining part belongs to the Blood Cancer class, which comprises 8312 instances.

3. 3 Data Reduction

The random projection has been utilised as a tool for dimensionality reduction [13]. The data dimensionality will be reducted using a random matrix where the data will be projected onto lower-dimensional space. Therefore, the number of attributes in the data will be shrunk while much of its variation will be preserved like Principle Component Analysis (PCA), particularly the class attribute. First of all, the NominalToBinary filter will be applied, which converts all the attributes to numeric values, then the dimensionality will be reduced. Random Projection is much less computationally expensive than PCA.

3. 4 Implemented Algorithms

Weka is a data mining tool which has been used to carry out implementations and experimentations. Weka stands for (Waikato Environment for Knowledge Analysis) and is written in Java Programming Language at Waikato. We used Weka in order to predict our dataset using several classification algorithms, for instance, Bayes and Functions as they have been categorised in Weka. Furthermore, we have measured out the tendency of each

classifier by reporting its accuracy value and some evaluation metrics.

Weka data mining tool has four applications Explorer, Experimenter, Knowledge flow and Simple CLI. In this paper, we are interested only in two techniques Explorer and Experimenter:

3. 4. 1 Explorer Interface:

This application has several panels such as Preprocess, Classify, Cluster, Associate, Select attributes and Visualise. However, our main aim in this interface focuses only on the Classify panel. Now, we are going to explain our used classifiers, as stated below:

3. 4. 1. 1 Naïve Bayes

Naïve Bayes classifier is a probabilistic, practical and straightforward supervised learning algorithm. It assumes that each feature is statistically independent and connects equally to the target class [14]. Moreover, Naïve Bayes' idea is based on applying the same conditional probability rules of Bayesian Network [15].

3. 4. 1. 2 Logistics Regression

Logistic Regression classifier is used for the likelihood terms also frequently used in binary classification [16]. It works to fit the logistic model during the training phase. During the test phase, it transforms its target output using the logistic sigmoid function (i.e., creates "S" shaped curve when the graph plotted) to obtain a probability value into a range of (0, 1) [17].

3. 4. 1. 3 Support Vector Machine (SVM)

Sequential Minimal Optimisation (SMO) uses polynomial or RBF kernels to train the support vector machine (SVM) classifier. SVM has been extensively used for classification, regression and density estimation [18]. During the training stage, it converts all nominal attributes into binary values and replaces missing values. SVM aims at maximising the margin between classes by finding the optimal separating hyperplane [19].

3. 4. 1. 4 Artificial Neural Network (ANN)

Artificial neural network (ANN) is a mathematical structure inspired by the organisation and practical feature of biological neural networks [20]. The Multi-Layer Perceptron (MLP) model is a feed-forward neural network and is the most common neural network model [21]. MLP model consists of multiple layers of nodes in a directed graph: the input layer, one or more hidden layers and the output layer. Each layer fully connected to the subsequent one, except for the input layer. MLP uses a supervised learning technique called backpropagation gradient descent to train the network, which minimises the variation between the network output and the desired output. Two steps could acquire this: computing gradient of the loss/error function, then updating current parameters in response to the gradient. These steps are repeated until the loss function reaches its minimum value.

3. 4. 2 Experimenter

This application provides facility for comparison of different classification algorithms. Each algorithm executes ten times with cross-validation ten folds; then the accuracy value is gained.

4. Results and Discussion

In this section, we will discuss our experiments by using Weka 3.8. We will apply all the classification algorithms addressed in section 3.4. To evaluate each classification algorithm and achieve better accuracy, we will use 10-fold-cross validation which subsets our dataset randomly into ten folds (i.e., nine folds for the training set and one fold for the test set). Moreover, our classifiers' results will be trained out based on the following:

Correctly Classified Accuracy: refers to the percentage of correctly classified features.

Incorrectly Classified Accuracy: refers to the percentage of incorrectly classified features.

Root Mean Square Error (RMSE): determines the differences between predicted value by the classification model and the estimated/ observed one.

Time: shows the possible required time (in seconds) to build the classification model.

Table 2: Performance parameters of the algorithms

	Correctly Classified Instances	Incorrectly Classified Instances	Root Mean Squared Error (RMSE)	Time (in seconds)
Naïve Bayes	16887 (97.0238%)	518 (3.4946%)	0.1561	0.01 s
Logistic Regression	16988 (97.6041%)	417 (2.3959%)	0.1319	0.34 s
Support Vector Machine	17043 (97.9201%)	362 (2.0799%)	0.1442	0.38 s
MLP (default hidden layer)	17256 (99.1439%)	149 (0.8561%)	0.087	8.97 s
MLP (2 hidden layers)	17194 (98.7877%)	211 (1.2123%)	0.108	4.75 s
MLP (4 hidden layers)	17239 (99.0463%)	166 (0.9537%)	0.0915	7.55 s
MLP (6 hidden layers)	17273 (99.2416%)	132 (0.7584%)	0.082	10.29 s
MLP (8 hidden layers)	17300 (99.3967%)	105 (0.6033%)	0.0742	13 s
MLP (10 hidden layers)	17289 (99.3335%)	116 (0.6665%)	0.0777	20.05 s

As it can be noticed from Table 2, Naïve Bayes classifier seems to be the fastest classifier. Although the percentage of correctly classified instances is 97.0238% and the incorrectly classified instances are 3.4946%. RMSE is 0.1561.

Logistic Regression and Support Vector Machine classifiers got roughly the same results. The proportion of correctly classified instances are 97.6041% and 97.9201%, respectively. Additionally, the percentage of incorrectly classified instances is 2.3959% and 2.0799%, respectively, and the value of RMSE is 0.1319 and 0.1442, respectively. Nonetheless, Logistic Regression took only 0.34 seconds to build its model while SVM built its classification model in 0.38 seconds which looks a bit longer.

Multilayer Perceptron appears to be quite complicated. It demonstrates to be more accurate when the number of hidden layer increases but takes a much longer time to build the model. Furthermore, the option of default hidden layers value manifests more accurately than setting out the hidden layers to 2 or 4 and takes longer time 8.97 seconds to construct the classification model.

Therefore, we recommended eight hidden layers as the optimal number of hidden layers. The percentage of correctly classified instances and incorrectly classified instances are 99.3967% and 0.6033% respectively, which looks more accurate than selecting two hidden layers, 4 or 6 or even keeping the default number. The MLP model took precisely 13 seconds to be constructed. The RMSE is 0. 0742, which is the least value of all the mentioned classifiers.

To evaluate the performance of our classifiers, we have used the following parameters:

Confusion Matrix (or Contingency Matrix): gives us a summary of the performance of a classifier. The following table 3 shows us the typical confusion matrix.

Table 3: 2×2 Confusion Matrix for our binary classifier

		Predicted Class	
		Blood Cancer	**Non-Blood Cancer**
Actual Class	**Blood Cancer**	TP	FN
	Non- Blood Cancer	FP	TN

TP (i.e., True Positive): Number of correctly classified Blood Cancer instances.
FN (i.e., False Negative): Number of incorrectly classified Non-Blood Cancer instances.
FP (i.e., False Positive): Number of incorrectly classified Blood Cancer instances.
TN (i.e., True Negative): Number of correctly classified Non-Blood Cancer instances.

After that, the Precision, Recall and F1 Score will be calculated:

Accuracy: The proportion of correctly predicted observations over the total number of instances.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Precision: the proportion of TP observations with respect to the total predicted Blood Cancer observations.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall: the proportion of TP observations with respect to all actually predicted Blood Cancer observations.

$$\text{Recall} = \frac{TP}{TP+FN}$$

F1 Score: the harmonic average of precision and recall. It could be calculated as the following equation:

$$\text{F1 Score} = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

ROC Area (Receiver Operating Characteristic Curve): shows the comparison between two classification models; the True Positive Rate (TPR) and False Positive Rate (FPR).

Table 4: Weighted average performance parameters of the algorithms based on the confusion matrix

Algorithm	Accuracy	Precision	Recall	F-Measure	ROC Curve
Naïve Bayes	97.0238%	0.970	0.970	0.970	0.994
Logistic Regression	97.6041%	0.976	0.976	0.976	0.997
Support Vector Machine	97.9201%	0.979	0.979	0.979	0.980
MLP (default hidden layer)	99.1439%	0.991	0. 991	0. 991	0.999
MLP (2 hidden layers)	98.7877%	0.988	0.988	0.988	0.993
MLP (4 hidden layers)	99.0463%	0.990	0. 990	0. 990	0.999
MLP (6 hidden layers)	99.2416%	0.992	0.992	0.992	0.999
MLP (8 hidden layers)	99.3967%	0.994	0. 994	0. 994	0.999
MLP (10 hidden layers)	99.3335%	0.993	0.993	0.993	0.999

As it is literally obvious, the most accurate classifier is MLP with 8 hidden layers as it got the best values in all evaluators: precision, recall, f-measure and ROC curve.

Table 5: Experimenter Result

	Best Accuracy (V)	Worse Accuracy ( )	Worst accuracy (*)
Naïve Bayes		93.91	
Logistic Regression	97.65		
SVM	97.73		
MLP (8 hidden layers)	99.17		

Table 5 shows the best algorithms in terms of accuracy (i.e., percent correct instances) Logistic Regression, SVM and MLP that significantly better than Naïve Bayes as they are followed by (V). Naïve Bayes, which got accuracy 93.91%, is considerably

worse than the other classifiers and is followed by nothing ( ). To be noted, the significance level is 5% or 0.05.

5. Conclusion and Future Work

Effective medical diagnosis needs knowledge discovery from the medical database. Extracting knowledge from information stored in the database is data mining's goal as generating a comprehensible description of patterns. This paper discussed the notion of classifying several kinds of cancer disease to blood cancer or non-blood cancer using open-source WEKA data mining tool. Weka has four interfaces. We have only used two interfaces: Explorer and Experimenter. In this research, we have used four classification algorithms: Naïve Bayes, Logistic Regression, Support Vector Machine and Multilayer Perceptron for our experimentation. All these mentioned algorithms were implemented using Weka to find algorithm accuracy. The accuracy of each classifier compared in terms of correctly classified instances, incorrectly classified instances, root mean squared error and time taken to build the model.

Additionally, to evaluate each classifier's performance, we have used the following evaluators: accuracy, precision, recall, f-measure and ROC curve. As proved in this paper, the maximum accuracy went to MLP classifier. It got the best results in all the mentioned assessors regardless of the time, which was almost longer than other classifiers that did not get pretty satisfying results during the evaluations.

In the future, we will get a dataset of another case. Similarly, we will test data mining techniques' performance and determine which algorithm will outperform the others.

Conflict of interests

None.

References

1. M. Durairaj and V. Ranjani, "Data Mining Applications In Healthcare Sector: A Study," *Int. J. Sci. Technol. Res.*, vol. 2, no. 10, pp. 29–35, 2013.
2. H. C. Koh and G. Tan, "Data mining applications in healthcare.," *J. Healthc. Inf. Manag.*, vol. 19, no. 2, pp. 64–72, 2005, doi: 10.4314/ijonas.v5i1.49926.
3. A. M. El-Halees and A. H. Shurrab, "Blood Tumor Prediction Using Data Mining Techniques," *Heal. Informatics - An Int. J.*, vol. 6, no. 2, pp. 23–30, 2017, doi: 10.5121/hiij.2017.6202.
4. S. Saichanma, S. Chulsomlee, N. Thangrua, P. Pongsuchart, and D. Sanmun, "The observation report of red blood cell morphology in Thailand teenager by using data mining technique," *Adv. Hematol.*, vol. 2014, pp. 4–9, 2014, doi: 10.1155/2014/493706.
5. M. N. Amin and A. Habib, "Comparison of Different Classification Techniques Using WEKA for Hematological Data," *Am. J. Eng. Res.*, no. 43, pp. 2320–847, 2015, [Online]. Available: www.ajer.org.
6. K. Li, M. Yang, G. Sablok, J. Fan, and F. Zhou, "Screening features to improve the class prediction of acute myeloid leukemia and myelodysplastic syndrome," *Gene*, vol. 512, no. 2, pp. 348–354, 2013, doi: 10.1016/j.gene.2012.09.123.
7. S. Vijayarani and S. Sudha, "An efficient clustering algorithm for predicting diseases from hemogram blood test samples," *Indian J. Sci. Technol.*, vol. 8, no. 17, 2015, doi: 10.17485/ijst/2015/v8i17/52123.
8. K. A. S. A. Daqqa, A. Y. A. Maghari, and W. F. M. Al Sarraj, "Prediction and diagnosis of leukemia using classification algorithms," *ICIT 2017 - 8th Int. Conf. Inf. Technol. Proc.*, no. October, pp. 638–643, 2017, doi: 10.1109/ICITECH.2017.8079919.
9. G. Krishna, B. Kumar, N. Orsu, and S. B., "Performance Analysis and Evaluation of Different Data Mining Algorithms used for Cancer Classification," *Int. J. Adv. Res. Artif. Intell.*, vol. 2, no. 5, pp. 49–55, 2013, doi: 10.14569/ijarai.2013.020508.
10. M. Durairaj and R. Deepika, "Prediction of Acute Myeloid Leukemia Cancer Using Dataming - A Survey," *Int. J. Emerg. Technol. Innov. Eng.*, vol. 1, no. 2, pp. 94–98, 2015, doi: ISSN: 2394-6598.
11. S. David, A. Saeb, and K. Al Rubeaan, "Comparative Analysis of Data Mining Tools and Classification Techniques using WEKA in Medical Bioinformatics," *Comput. Eng. Intell. ...*, vol. 4, no. 13, pp. 28–39, 2013, [Online]. Available: http://iiste.org/Journals/index.php/CEIS/article/view/9348.
12. A. Fern and S. Garc, "SMOTE for Learning from Imbalanced Data : Progress and Challenges , Marking the 15-year Anniversary," vol. 61, pp. 863–905, 2018.
13. A. Mylavarapu, Sachin; Kaban, "Random projections versus random selection of features for classification of high dimensional data," in *Computational Intelligence (UKCI)*, 2013, pp. 305–312.
14. S. Misra, H. Li, and J. He, *Machine Learning for Subsurface Characterisation*, 1st Editio. Gulf Professional Publishing, 2020.
15. M. Rathi and A. K. Singh, "Breast Cancer Prediction using Naïve Bayes Classifier Breast Cancer Prediction using Naïve Bayes Classifier," vol. 1, no. 2, pp. 77–80, 2012.
16. T. Ayer, J. Chhatwal, O. Alagoz, C. E. Kahn, R. W. Woods, and E. S. Burnside, "Informatics in radiology: Comparison of logistic regression and artificial neural network models in breast cancer risk estimation," *Radiographics*, vol. 30, no. 1, pp. 13–22, 2010, doi: 10.1148/rg.301095057.
17. J. Mandák and J. Hančlová, "Use of logistic regression for understanding and prediction of customer churn in telecommunications," *Statistika*, vol. 99, no. 2, pp. 129–141, 2019.
18. J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, no. xxxx, 2020, doi: 10.1016/j.neucom.2019.10.118.
19. S. H. Hsieh, Z. Wang, P. H. Cheng, I. S. Lee, S. L. Hsieh, and F. Lai, "Leukemia cancer classification based on support vector machine," *IEEE Int. Conf. Ind. Informatics*, pp. 819–824, 2010, doi: 10.1109/INDIN.2010.5549638.
20. I. M. Nasser and S. S. Abu-naser, "Predicting Tumor Category Using Artificial Neural Networks," vol. 3, no. 2, pp. 1–7, 2019.
21. S. Agrawal and J. Agrawal, "Neural network techniques for cancer prediction: A survey," *Procedia Comput. Sci.*, vol. 60, no. 1, pp. 769–774, 2015, doi: 10.1016/j.procs.2015.08.234.