Performance Evaluation of Different Supervised Machine Learning Algorithms in Predicting Linear Accelerator Multileaf Collimator Positioning's Accuracy Problem

Hamed S. El-Ghety[1], Ismail Emam[2], AbdelMagid M. Ali[3]

Electrical Engineering Department, Faculty of Engineering, Aswan University, Aswan, Egypt[1, 3]
Oncology department, Faculty of Medicine, Ain Shams University, Cairo, Egypt[2]

Abstract—Radiation Oncology is one of the businesses that employs Machine Learning to automate quality assurance tests so that errors and defects can be reduced, avoided, or eliminated as much as possible during tumor therapy using a Linear Accelerator with MultiLeaf Collimator (Linac MLC). The majority of Machine Learning applications have used supervised learning algorithms rather than unsupervised learning algorithms. However, in most cases, there is a clear bias in deciding which supervised machine learning algorithm to use. And prediction findings may be less accurate as a result of this bias. As a result, in this study, an evidence is presented for a novel application of Logistic Regression technique to predict Linac MLC positioning accuracy, which achieved 98.68 percent prediction accuracy with robust and consistent performance across several sets of Linac data. this evidence was obtained by comparing the performance of various supervised machine learning algorithms (i.e. Logistic Regression, Decision Tree, Support Vector Machine, Random Forest, Naive Bayes, and K-Nearest Neighbor) in the prediction of Linac MLC's positioning accuracy problem using leaves' positioning displacement datasets with labelled results as training and test datasets. For each method, two parameters were used to evaluate performance: prediction accuracy and the receiver operating characteristics curve. Based on that evaluation, the right selection sequence was proposed for supervised Machine Learning algorithms in order to achieve near-optimal prediction performance for Linac MLC's leaf positioning accuracy problem. As a result, the selection bias, as well as the negative side effects (i.e. ineffective preventive maintenance plan for Linac MLC to avoid and solve causes of inaccurate leaf displacement such as motor fatigue and stuck problems) could have occurred were successfully avoided.

Keywords—Linear accelerators; logistic regression; performance evaluation; prediction methods; supervised learning

I. INTRODUCTION

Machine learning applications have been utilized in different industries including Radiation Oncology [6]. In Radiation Therapy, some researches summarizes potential various clinical applications such as head, neck, lung, and prostate cancer as well as radiation toxicity [1][2][3]. Other researches states that differences between planned and actual displacements of multi-leaf collimators (MLCs) are source of errors in dose distributions during radiotherapy [4]. However, Radiation Therapy is still considered niche area with big crude data that needs extensive use of machine learning applications. And since the precision medicine in radiation oncology, radiation toxicity and complication factors are inevitable conditions for oncology patients after radiotherapy [1][4][5] and since most of time the use of popular supervised learning algorithms (e.g. Support Vector Machine and Decision Tree) are supported by previous prediction accuracies in other industries regardless of differences in nature of the data itself which is considered a selection bias that may produce less accurate prediction results. So this paper focuses on performance evaluation of different popular supervised learning algorithms in the prediction of leaf displacement accuracy problem utilizing Linear Accelerator with Multi-Leaf Collimator (Linac MLC) by comparing two Criteria factors: the prediction accuracy of the algorithm and the corresponding receiver operating characteristics curve.

This work will help researchers tackling similar Linac MLC prediction problems with the same nature of displacement data to use logistic regression technique confidently to get near-optimal prediction. At the same time, this work will guide researchers in other business areas as well to follow the same evaluation process practice that is undertaken in this paper, prior to using a typical supervised learning algorithm with a typical data of certain nature, by this way, they can properly select the most suitable supervised learning algorithms that gives near-optimal prediction.

As follows, this paper will have seven remaining sections: Methods for supervised learning; Using supervised learning in Linac MLC; Methodology; Implementation; Results and discussion; Conclusion; Acknowledgment; References.

II. METHODS FOR SUPERVISED LEARNING

This section gives a brief background on different supervised machine learning algorithms.

A. Decision Tree (DT)

In machine learning, DT is one of the most useful and reliable classifiers. The decision tree has a hierarchical design that employs the divide-and-conquer strategy [7]. As a result, it can be used for classification. And reduced to a series of simple if-then statements [14].

B. Support Vector Machine (SVM)

SVM is a supervised learning algorithm. SVM is used for classification in many applications. Using the margin concept, the Linear Support Vector Classifier determines an optimal separating hyper-plane. This represents the distance between the hyper-plane and the nearest points to it on either side, and can be maximized for better generalization [8].

C. Random Forest (RF)

RF classifier is composed of several DTs, like how many trees build up a forest. Deep Decision trees frequently overfit the training data, which means that any minor change in the given data will produce a large variance in classification results. In other words, nature of training data makes them more likely to give wrong predictions with the test dataset. Random forest's decision trees needs to be trained using different portions of training dataset [9]. To classify a new sample, the sample's input vector must be passed down through the forest with each DT. Following that, each DT considers a different section of the input vector to determine the classification conclusion. The forest then decides whether to use the classification with the most 'votes' (for discrete classification outcomes, such as the MLC case study used in this research) or the average of all trees in the forest (for numeric classification outcome). Because the RF algorithm considers the results of multiple DTs, it can reduce the variance caused by considering only one DT for the same dataset [9].

D. Logistic Regression (LR)

LR is a normal type of regression where two state variables can be modelled easily. Thus, it helps to determine the likelihood that a new sample is associated with a typical class. And if it's used to classify binary samples, an input sample with a probability value greater than 0.50 is classified as 'class A'; otherwise, it is classified as 'class B [10].

E. Naïve Bayes (NB)

The NB classifier is a categorization strategy that computes the likelihood of an event based on prior knowledge of the event's conditions. Despite the fact that features in a class may be interdependent, so it considers that an item in that class is not directly associated to any other items [11].

F. K-Nearest Neighbor (KNN)

KNN classifier involves using a database to classify unknown cases. The observations are displayed in a three-dimensional space, with the number of qualities or features that each observation possesses indicated. Based on its similarity to other data points in the model, a new point is classified using some similarity measures [15]. KNN determines the new point's class by selecting the K closest points to the new example and voting for the most frequent class among them to be the new point's class, and so on, where K is the number of neighbors [8]. Fig. 1 illustrates the KNN method with k=1.
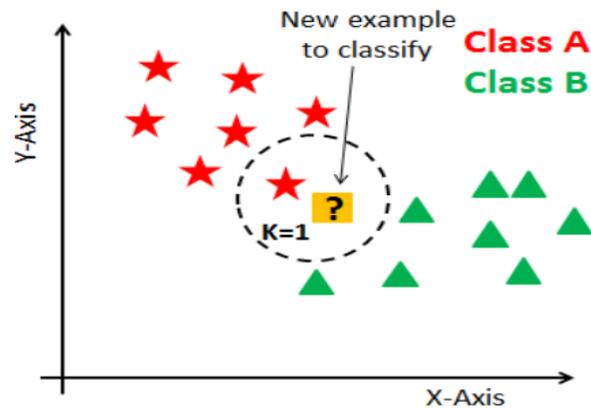

Fig. 1. Illustaration of K-Nearest Neighbor Method with k =1.

III. USING SUPERVISED LEARNING IN LINAC MLC

Because Modern radiotherapy procedures necessitate the use of high-precision beam shaping devices due to the reliance on administered dosage modulation. Random errors should be eliminated by paying close attention to the accuracy and performance of the MLC. Systematic errors must be identified and reduced [4][5].

Using Supervised Learning Algorithms to predict the problem of leaf displacement accuracy in a multi-leaf Collimator mounted in a Linear Accelerator Head will result in accurate positioning based on the shape of the tumor being treated while protecting other nearby body organs, thereby contributing to accurate radiation dose delivery to oncology Patients. Fig. 2 shows a photo of a Multileaf Collimator [12].
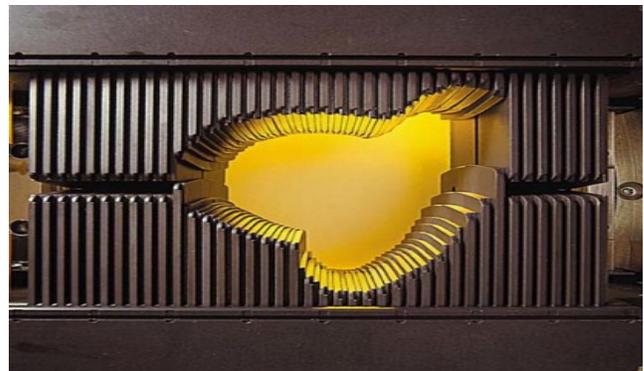

Fig. 2. Photo of Multi-leaf Collimator MLCi2 While A-Bank Leaves (Upper Leaves) and B-Bank Leaves (Lower Leaves) are positioned to take the Precise Shape of the Tumor so that the Treatment Beam can get through that Shape only [12].

IV. METHODOLOGY

As shown in Fig. 3, the application-related Data (i.e. Linac MLC's Leaves' displacement Dataset) will be used for Training while developing the Learning model provided by the supervised Machine Learning algorithm, as well as for testing to evaluate the developed learning model. As a result, we could finally compute accuracy, draw the receiver operating characteristic curve (ROC), and evaluate the various algorithms.
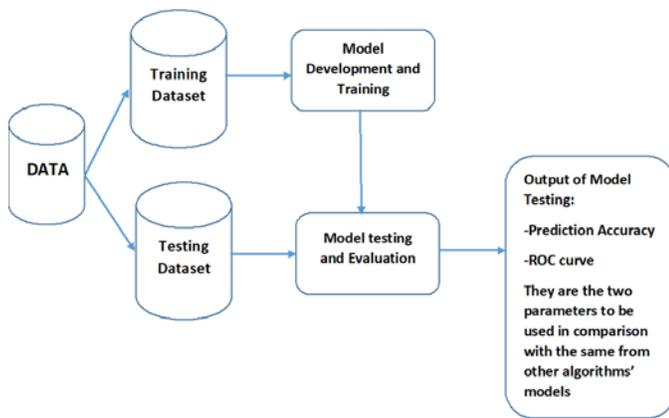
Fig. 3.   Steps followed to Assess the Performance of SVM, Decision Tree, RF, Logistic Regression, NB and KNN Algorithms.

A. Confusion Matrix

The confusion matrix shown in Table I is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known [8].

TABLE I.   CONFUSION MATRIX FOR TWO CLASSES

	PREDICTED CLASS	
	POSITIVE	*NEGATIVE*
ACTUAL CLASS	TRUE POSITIVE (*tp*)	FALSE NEGATIVE (*fn*)
	FALSE POSITIVE (*fp*)	TRUE NEGATIVE (*tn*)

The above confusion matrix taught us that there are two possible predicted classes: "True" and "False" [8]. In this paper, for example, we predict the absence of a leaf positioning accuracy problem for a typical leaf out of 40 pair leaves in Linac MLC, as a result, we can extract and define the following terms:

True positives (tp): These are cases where the prediction was correct and the leaves did not have positioning accuracy problem.

True negatives (tn): These are cases where the prediction was correct and the leaves have positioning accuracy problem.

False positives (fp): These are cases where the prediction was incorrect and the leaves actually did not have positioning accuracy problem.

False negatives (fn) are cases where the prediction was incorrect and the leaves actually have positioning accuracy problem. Table I can be used to extract the following rules [3]:

$$Accuracy = (tn + tp)/(tn + fn + fp + tp) \qquad (1)$$

$$Error = (fn + fp)/(tn + fn + fp + tp) \qquad (2)$$

$$Sensitivity = recall = tp/(tp + fn) \qquad (3)$$

$$Specifity = tn/(fp + tn) \qquad (4)$$

B. Receiver Operating Characteristics (ROC)

The receiver operating characteristics (ROC) curve and confusion matrix are frequently used to evaluate the diagnostic ability of supervised machine learning algorithms [7]. The Y axis of receiver operating characteristic curves has the tp rate and the X axis has the fp rate. That is, the "ideal" point is at the top left corner of the plot, where fp rate is zero and tp rate is one. However, this isn't very realistic, but it does imply that a larger area under curve (auc) is usually preferable. Fig. 4 depicts an example of evaluation for various ROC curves. Whereas the blue curve has the lowest auc, indicating poor prediction performance, the red curve has the highest auc, indicating excellent prediction performance [13].
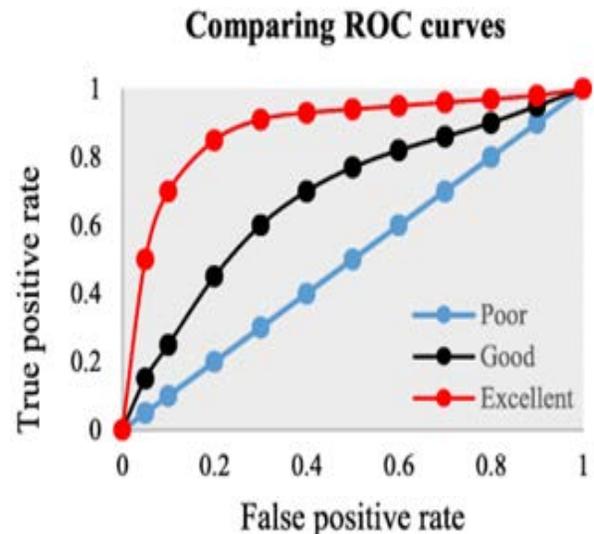

Fig. 4.   Illustrates Example Evaluation for different ROC Curves; where Blue Curve is Poor because Area under Vurve is the Lowest and Red Curve is Excellent because Area under Curve is Highest.

C. Data Collection and Processing

Good test and training Dataset was collected over a working year (252 days) for an MLCi2 Multi-leaf Collimator (MLC) mounted in an Elekta Synergy Linear Accelerator. The MLC has 40 leaf-pairs Linac (80-leaves). The 40 leaf-pairs are numbered into two banks "A&B" as: (A1, A2,.….,A40 and B1,.…,B40). The tolerance of the leaf positioning accuracy is 2 mm, while the action level is 3 mm. Table II shows a sample of collected data with labeled input features and its associated labeled result (i.e. answer to question is there isn't positioning accuracy problem?).

As shown in Fig. 5, 70% of MLC leaves displacement dataset was used for training the supervised machine learning models (e.g. x_train represents training features and y_train is the labelled result of training dataset) whereas 30% of the dataset was used to test and evaluate the trained models (e.g. x_test represents test features and y_test is the labelled result of test dataset).

Data is then processed using Python 3.8 using PyCharm IDE. Python package (i.e. Scikit-learn 0.23.2) was used to implement DT, SVM, RF, LR, NB and KNN Classifiers.

TABLE II. SAMPLE CASES FOR DISPLACEMENT DATA CASES FOR 80 LEAVES MULTI-LEAF COLLIMATOR (TRUE:T /FALSE:F)

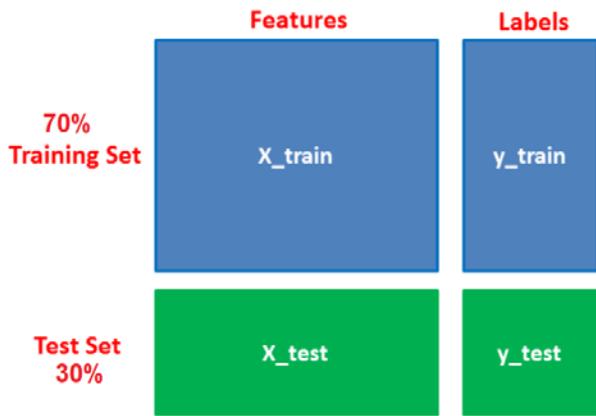| LINEAR ACCELERATOR MULTI-LEAF COLLIMATOR MLC LEAVES' DISPLACEMENT (A1,A2,…A40,B1,B2,….B40)										NO POSITIONING ACCURACY PROBLEM?	
A1	A2	A3	A4	A5	...	B37	B38	B39	B40	*TRUE T/FALSE F*	
										A BANK LEAVES	*B BANK LEAVES*
0.24	0.99	0.42	0.38	1.17	…	0.24	1.28	0.62	1.04	T	T
0.38	1.1	0.46	0.34	0.94	…	0.55	0.36	1.1	0.61	A	B
										T	T
0.42	0.84	0.84	0.92	0.24	…	1.09	0.96	0.75	0.77	A	B
										T	T
0.55	0.74	0.38	1.15	0.95	…	0.75	0.59	0.46	1.06	A	B
										T	T
1.13	2.12	0.98	0.96	1.29	…	1.32	1.35	1.89	0.74	A	B
										F	F

Fig. 5. Illustration of how Linear Accelerator Multileaf Collimator MLC Leaves' Displacement Dataset is Prepared for use by Machine Learning Algorithms.

V. IMPLEMENTATION

A. Prediction for MLC's Positioning Accuracy

Table III shows Measured Accuracy achieved by different Classifiers (i.e. DT, SVM, RF, LR, NB and KNN classifiers) in prediction of 40-pairs MLC's positioning accuracy problem.

TABLE III. EVALUATION OF SUPERVISED MACHINE LEARNING ALGORITHMS IN PREDICTION OF LINEAR ACCELERATOR MULTI-LEAF COLLIMATOR POSITIONING ACCURACY PROBLEM

		A-Bank Leaves	B-Bank Leaves
Decision Tree	Accuracy	94.74 %	93.42 %
Support Vector Machine	Accuracy	99.98%	97.37 %
Random Forest	Accuracy	97.36 %	98.68 %
Logistic Regression	Accuracy	98.68 %	98.68 %
Naïve Bayes	Accuracy	97.37 %	94.74 %
K-Nearest Neighbor	Accuracy	96.05 %	98.68 %

It shows that DT, SVM, RF, LR, NB and KNN classifiers were able to predict leaf positioning accuracy problem for Linac MLC successfully with average accuracy of 94.08%, 98.68%, 98.02%, 98.68%, 96.05%, and 97.37%, respectively. While Fig. 6 shows different ROC curves, one per each classifiers based on their prediction performance, so that we can compare between algorithms based on steepness of the curve and area under curve (auc). Where it's clear that red curve representing logistic regression algorithm has the highest ROC area under curve of 0.992, while green curve representing Decision Tree classifier has the lowest ROC area under curve.
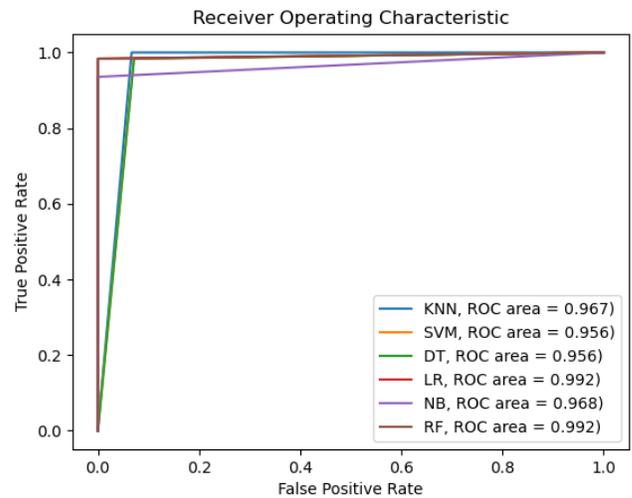

Fig. 6. Receiver Operating Characteristic Curve for Prediction of Linear Accelerator Multi-leaf Collimator's Positioning Accuracy per each Algorithm.

VI. RESULT AND DISCUSSION

DT, SVM, RF, LR, NB and K-Nearest Neighbor classifiers were able to predict leaf positioning accuracy problem for Linac MLC successfully with average accuracy of 94.08%, 98.68%, 98.02%, 98.68%, 96.05%, and 97.37%, respectively.

Although DT has reasonable average accuracy of 94.08% but it's the lowest one amongst other algorithms. And the ROC curve for DT performance has the lowest steepness and

ROC area under curve as well. So it's not recommended to select DT as first choice to use on such MLC's leaves' displacement dataset nature. On the other side, SVM performed well with average accuracy of 98.68% but it showed lower steepness in ROC Curve and area under curve.

Other classifiers RF, LR, NB and KNN showed better area under curve, and their average prediction accuracies were RF with 98.02%, LR with 98.68%, NB with 96.02% and finally KNN with 97.37%.

However, it's important to note that, Logistic Regression Classifier has the highest ROC area under curve of 0.992 and it showed exceptional performance by having the same classification prediction accuracy of 98.68% over two different datasets of same structure and nature but different values (i.e. MLC's A Bank Leaves' displacements and MLC's B Bank Leaves' displacement) which indicates more performance stability than other classifiers even SVM itself. In this paper, bias of algorithm selection have been successfully avoided and the recommended selection should consider the relevancy order based on classifiers evaluation as shown in Table IV, in context of application area related to MLC's Leaves' displacement Data. And accordingly we do recommend using LR as first choice because it has highest average prediction accuracy and the most stable performance across different MLC's datasets of the same nature.

TABLE IV. PROPOSED SELECTION ORDER FOR CLASSIFIERS TO USE IN PREDICTION OF MLC POSITIONING ACCURACY PROBLEM

Classifiers	Average Accuracy	Selection Order
Logistic Regression	98.68%	1
Support Vector Machine	98.68%	2
Random Forest	98.02%	3
K-Nearest Neighbor	97.37%	4
Naïve Bayes	96.05 %	5
Decision Tree	94.08%	6

On the other hand and according to application area perspective, the high accuracy of the prediction for MLC's Leaves' positioning problem would enable the physicist in oncology center to design customized service/preventive maintenance plans for each individual Linac MLC treatment machine particularly. And that could help to avoid MLC movement failure during radiation therapy sessions.

VII. CONCLUSION

This work is undertaken to avoid lower performance in prediction process of Linac MLC's positioning accuracy problem. In this paper, performance of DT, SVM, RF, LR, NB and KNN Classifiers is examined by measuring their prediction accuracies utilizing the same two sets of training and testing data for Linac MLC's leaves' positioning displacement data as well as receiver operating characteristic curves for the predicted outcomes per each algorithm. Findings in this study show that Logistic regression Classifier

has exceptional performance by producing the same classification prediction accuracy of 98.68% over two different datasets of same structure and nature but different values (i.e. MLC's A Bank Leaves' displacements and MLC's B Bank Leaves' displacement) which indicates more performance stability than other classifiers even SVM itself. The findings show that values and structure of data affect the prediction accuracy of supervised learning algorithm applications across different industries and not necessarily the same performance. In order to increase the prediction accuracy in the same time, further research work is needed on more training and test datasets over longer periods (e.g. five years), and a multi-institutional study (e.g. different healthcare providers which uses the same model of Linac MLC).

REFERENCES

[1] Suk Lee, Kwang Hyeon Kim, Choi Suk Woo, Jang Bo Shim, Yuan Jie Cao, Kyung Hwan Chang and Chul Yong Kim "Predictive Solution for Radiation Toxicity Based on Big Data", May 2017,ch.9,p175-191 Radiotherapy.

[2] El Naqa I, Bradley JD, PE L, Hope AJ, Deasy JO. Predicting radiotherapy outcomes using statistical learning techniques. Phys Med Biol. 2009;54(18):S9.

[3] Dean JA, Wong KH, Welsh LC, Jones AB, Schick U, Newbold KL, Bhide SA, Harrington KJ, Nutting CM, Gulliford SL. Normal tissue complication probability (NTCP) modelling using spatial dose metrics and machine learning methods for severe acute oral mucositis resulting from head and neck radiotherapy. Radiother Oncol. 2016 Jul;120(1):21–7.

[4] Joel N K Carlson, Jong Min Park, So-Yeon Park, Jong In Park, Yunseok Choi and Sung-Joon Ye,"A machine learning approach to the accurate prediction of multi-leaf collimator positional errors",Journal of Physics in Medicine & Biology,March 2016, Vol 61, Number 6,2514.

[5] El Naqa I, Li R, Murphy MJ. Machine learning in radiation oncology: theory and applications. Switzerland, Springer; 2015. 39.

[6] The R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, Machine learning: an artificial intelligence approach. Springer Science & Business Media, 2013.

[7] Fawcett T. An introduction to ROC analysis. Pattern Recogn Lett. 2006;27(8):861–74.

[8] Amr E. Mohamed, "Comparative Study of Four Supervised Machine Learning Techniques for Classification," International Journal of Applied Science and Technology Vol. 7, No. 2, June 2017, pp 5-17.

[9] Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.

[10] Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression. Wiley; 2013.

[11] I. Rish, "An empirical study of the naive Bayes classifier," in IJCAI 2001 workshop on empirical methods in artificial intelligence, 2001, vol. 3, 22, pp. 41–46: IBM New York.

[12] Elekta user manuals for MLCi2, Elekta© Crawley, UK.

[13] Jasmina Dj. Novakovic´,Alempije Veljovic´, Sinisa S. Ilic´,Zeljko Papic´,Milica Tomovic´. "Evaluation of Classification Models in Machine Learning" Theory and applications of Mathematics and Computer Science Vol 7 No.1, 2017,pp 39-46.

[14] Oded Maimon • Lior Rokach, Data Mining and Knowledge Discovery Handbook, NY: Springer. 2010, pp. 149-152.

[15] Jiawei Han and Micheline Kamber. Data Mining: Concepts and Techniques, NY: Elsevier. 2010, pp. 291-307.