Evidence-based Clinical Engineering:
Machine learning algorithms for prediction of defibrillator performance

Almir Badnjević[1,2], Lejla Gurbeta[1,2], Mehrija Hasičić[1], Lejla Bandić[1], Zerina Mašetić[1], Živorad Kovačević[1], Jasmin Kevrić[1], Leandro Pecchia[3]

[1] International Burch University, Sarajevo, Bosnia and Herzegovina

[2] Medical Device Inspection Laboratory Verlab Ltd., Sarajevo, Bosnia and Herzegovina

[3] School of Engineering, University of Warwick, United Kingdom

almir@verlab.ba, lejla@verlab.ba, mehrija.hasicic@ibu.edu.ba, lejla.bandic@ibu.edu.ba, zerina.masetic@ibu.edu.ba, zivorad.kovacevic@ibu.edu.ba, jasmin.kevric@ibu.edu.ba, l.pecchia@warwick.ac.uk

ABSTRACT

Poorly regulated and insufficiently supervised medical devices (MDs) carry high risk of performance accuracy and safety deviations effecting the clinical accuracy and efficiency of patient diagnosis and treatments. Even with the increase of technological sophistication of devices, incidents involving defibrillator malfunction are unfortunately not rare.

To address this, we have developed an automated system based on machine learning algorithms that can predict performance of defibrillators and possible performance failures of the device which can affect performance. To develop an automated system, with high accuracy, overall dataset containing safety and performance measurements data was acquired from periodical safety and performance inspections of 1221 defibrillator. These inspections were carried out in period 2015-2017 in private and public healthcare institutions in Bosnia and Herzegovina by ISO 17020 accredited laboratory. Out of overall number of samples, 974 of them were used during system development and 247 samples were used for subsequent validation of system performance. During system development, 5 different machine learning algorithms were used, and resulting systems were compared by obtained performance.

The results of this study demonstrate that clinical engineering and health technology management benefit from application of machine learning in terms of cost optimization and medical device management. Automated systems, based on machine learning algorithms, can predict defibrillator performance with high accuracy. Systems based on Random Forest classifier with Genetic Algorithm feature selection yielded highest accuracy among other machine learning systems. Adoption of such systems will help in overcoming challenges of adapting maintenance and medical device supervision mechanism protocols to rapid technological development of these devices. Due to increased complexity of healthcare institution environment and increased technological complexity of medical devices, performing maintenance strategies in traditional manner is causing a lot of difficulties.

The innovation *in the paper is concept of applying machine learning techniques in medical device (defibrillator) management. System like this are first step in introducing machine learning methods and intelligent systems into optimization of maintenance management of medical devices.*

Keywords— Automated System, Machine learning, Medical Device, maintenance, management, prediction, performance, inspection, evidence-based.

1 Introduction

Medical staff are nowadays more confident while performing diagnosis and treatment due to sophistication of Medical Devices (MDs) which allows them better data analysis and control over diagnosis or therapy. Despite of the existence of international standards, directives and regulations for medical devices [1-5] that regulate all aspect of life-cycle and are obligatory for the manufacturers, unfortunately, cases of MDs malfunction are not rare. Various incidents involving patient injuries and incidents with death outcomes caused by medical devices are reported every year by users, healthcare professional or manufacturers. One of the world's most prominent databases containing this data is the FDA Manufacturer's Facility Device Experience (MAUDE) database [6] available for the USA market, and European Database on Medical Devices (EUDAMED) database, for European area [7] and other national vigilance tools [8,9].

The number of these incidents is alarming and suggests that medical device post-market surveillance, supervision mechanisms and maintenance strategies are not implemented efficiently to ensure patient safety and quality of healthcare. Despite the existence of self-test protocols that are usually built in the MDs software, often medical professionals cannot recognize performance malfunction which directly affects patient diagnosis and/or treatment. Such malfunctions are seen as huge deviations of MDs patient related output parameters. For instance, in case of blood pressure devices, if device functions properly but has inaccurate measurements it will result in patient taking therapy for either higher or lower blood pressure than actual blood pressure level, so inaccurate diagnosis results in inappropriate treatment. Similarly, for a defibrillator, this means that critically ill patient will be either treated with higher or lower energy level than needed which results in failed resuscitation or burned patients. Events like this can be efficiently prevented by adequate supervision mechanisms, where medical devices that are already used in healthcare institutions would be periodically tested for safety and performance characteristics. Even though, post-market surveillance is defined and introduced into legislation, it is practiced differently across the world [10]. For instance, in Bosnia and Herzegovina these safety and performance inspections are defined through Legal Metrology Framework for Medical Devices. [11-14] According to this framework, defibrillators are periodically tested by legally appointed, ISO 17020 accredited laboratory. All safety and performance measurements and defibrillator information, such as serial number, type, manufacturer, location is stored in developed database [15].

Experience has showed that the power of data is in its analysis, so in this study, we aimed to investigate, how collected data can be used to predict device performance and future failures in order to optimize current medical device management strategies. According to Taghipour [16] annual medical device maintenance and management cost in healthcare institution is approximately 1% of the total budget. Healthcare institutions, unfortunately cut these costs so usage of medical devices as a consequence results in higher rate of incidents with serious injuries or deaths of patients. Also, they state that numerous optimization models for medical device maintenance have been developed, but healthcare institutions still do not benefit from these method as other industries do. Due to increased complexity of healthcare institution environment and increased technological complexity of medical devices, performing maintenance strategies in traditional manner is causing a lot of difficulties. Traditional medical device management is based on software programs [17,18] that provide continuous updates, increase inventory accuracy, documents maintenance history, and data analysis/reports. By introducing machine learning into healthcare, in terms of medical device management strategies, presented challenges can be resolved since raw data can be transformed into useful information that improve outcomes and reduce the burden on the healthcare system. Machine learning algorithms can learn from experience with respect to some task and some performance measure. [19-21].

The purpose of this study was to develop automated systems based on machine learning algorithms for prediction of performance of defibrillators based on data measured during periodical safety and performance inspections. Developed automated systems were evaluated and compared based on their performance. Our hypothesis was that an accurate automated system based on machine learning could successfully predict defibrillator performance and potential failures. To build such system, data from 1221 defibrillator inspections performed by appointed, ISO 17020 accredited laboratory in private and public healthcare institutions in Bosnia and Herzegovina was used.

2 Methods

The methodology of development of system for prediction of defibrillator performance is presented in Fig. 1. For such purpose, 5 different machine learning algorithms were used, that are: (1) Decision Tree, (2) Random Forest, (3) K-nearest Neighbor, (4) Support Vector Machines and (5) Naïve Bayes algorithm. The reason for choosing these algorithms is that they are general representatives of each group of classifiers, which were tested in various research studies [22-28]. Characteristics of these algorithms are given in text below.

For development of automated system for classification of defibrillator performance two approaches were used. In first approach, 5 different machine learning algorithms was used on original dataset to develop the systems, and in the second approach, dataset was firstly analyzed using feature selection algorithms. Feature selection was done using (1) InfoGain algorithm, (2) Decision Tree (DT) algorithm and (3) Wrapper algorithm to determine which parameters in the dataset have significant impact to defibrillator performance status, and which parameters have less impact. In this way optimization of developed system in terms of number of input parameters is done.

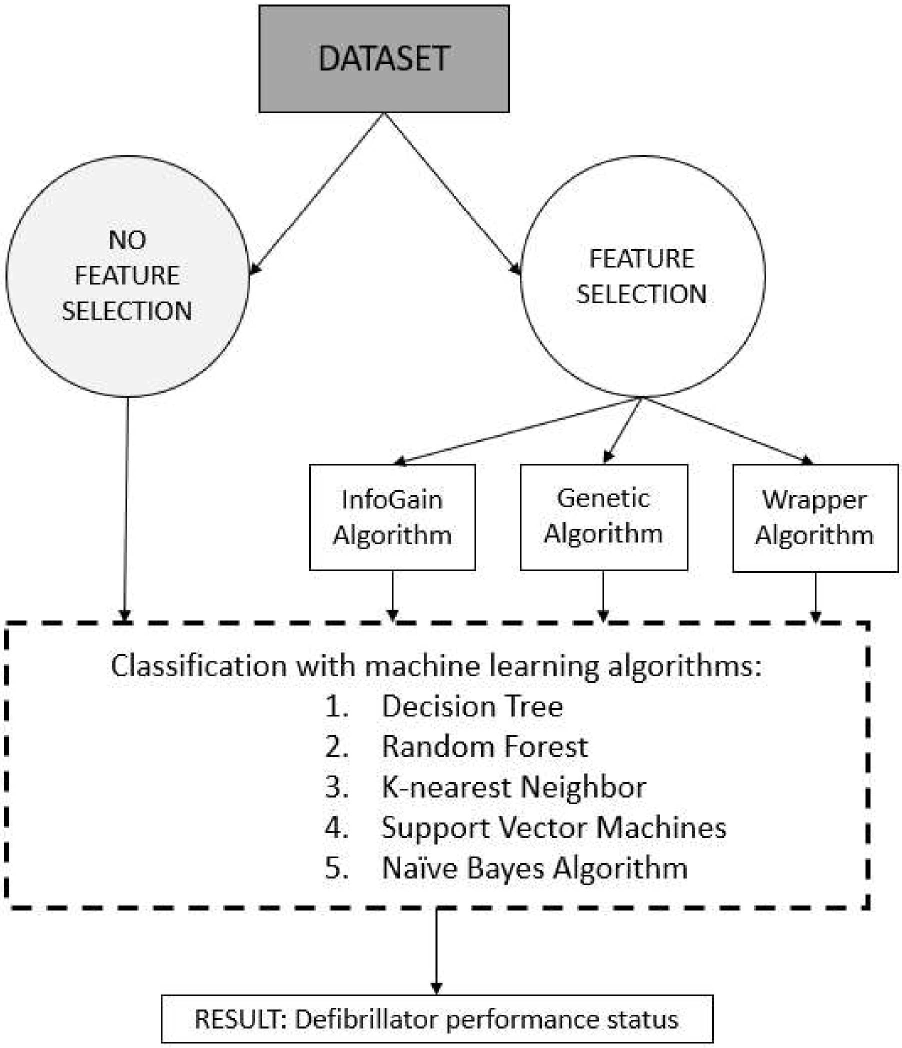

Figure 1. Block diagram of the classification system

2.1 Dataset

For the development of system for prediction of defibrillator performance, a dataset of 1221 samples was used. These samples were acquired during annual periodical inspections (2015 – 2017) of defibrillators in healthcare institutions in Bosnia and Herzegovina according to legal metrology framework for medical devices. [30-37] So, this dataset stored in the developed database [29], consists of measurements taken from the same defibrillator three years in a row (2015-2016-2017), measurements taken from the same defibrillator two years in a row (2016-2017) and measurement taken on a single defibrillator once (either 2015 or 2016 or 2017). Performance measurements were taken by Fluke Biomedical Impulse 7000 DP [30] and electrical safety measurements were taken by Fluke Biomedical ESA 620 [31].

Each sample consists of seven groups of features that are: (1) defibrillator output energy measurements - result of performance inspection, (2) measurement that are result of safety inspection, (3) information about device age, (4) manufacturer, (5) type and (6) information about preventive/corrective maintenances and (7) inspection decision. Overall, each sample consists of 38 features.

Electrical safety inspection was performed according to IEC 60601 Medical electrical equipment [3] electrical safety standard. Measurements of following parameters were taken: (1) mains voltage (live to neutral, neutral to earth, live to earth), (2) protective earth resistance, (3) insulation resistance (normal condition, mains to protective earth), (4) earth leakage current (applied parts and normal condition, open neutral, normal condition - reversed mains, open neutral – reversed mains), (5) enclosure leakage current (applied parts, normal condition, open neutral, normal condition – reversed mains, normal condition – reversed mains, open earth – reversed mains), and (6) patient applied parts leakage current.

During performance inspection, each defibrillator was tested in 8 measuring points equally distributed in the working range of the device. The definition of those points depended on defibrillator type. For monophasic defibrillators they were distributed in range [2 - 360] (J), and for biphasic defibrillators in range [2 – 230] (J) or [2 – 270] (J).

Inspection decision is formed based on safety and performance inspection. [32-39] It can be either positive (device passed inspection) or negative (inspection fail – faulty device) [32]. Accurate device means that defibrillator based on safety and performance inspection conforms to all technical and metrological requirements defined in national legislation, international standards and medical device directives/regulations. These devices are safe to use and probability for their failure in next period is lower. In case that some or multiple parameters of defibrillator has performance or safety error that is not in allowed limits the device is marked as faulty meaning that its potentially hazardous for usage on patients and should be removed from usage and should be subject of corrective maintenance.

Four groups of data were made for the purpose of development of automated system based on machine learning techniques. The first subset was original dataset containing all features, and three more as a result of application of feature selection algorithms, Figure 1 as follows: features selected with Info gain, features selected with Genetic algorithm and features selected with Wrapper algorithm.

All groups of data were then divided into training and testing subset. Equal representation of the ratio of positive and negative inspection outcomes was ensured and splitting ratio is 80-20 (%) was used. This splitting ratio is common for application of machine learning algorithms [40]. Training subset consisted of 947 samples, and testing subset consisted of 274 samples.

2.2 Machine learning algorithms

2.2.1. Feature selection algorithms

To enable creating the accurate prediction model while removing irrelevant and redundant attributes, reducing the complexity of the model and increasing the model performance [40] feature selection algorithms were applied. Three different feature selection methods that belong to three general classes of feature selection algorithms [40] were applied to the whole dataset. Those are: (1) InfoGain algorithm, (2) Decision Tree algorithm and (3) Wrapper algorithm [40], with characteristics shown in Table 1. Such approach has been used in previous studies as well. [41-46]

Table 1. Characteristics of applied feature selectors

Algorithm	InfoGain Algorithm [41,42]	Genetic Algorithm [43]	Wrapper Algorithm [44-46]
Group	Filter method	Stochastic general search method	Wrapper method
Ranking	Features are ranked by a score.	Inspired by procedures of natural evolution.	Selects a set of features considering them as a search problem.
Purpose	Computes how much information with respect to the classification target the attribute gives.	Operates on a population of individuals by selecting individuals according to their level of fitness in the domain.	Actual classification algorithm that builds a model with a subset of the attributes and evaluates the performance of this model.

2.2.2. Classification algorithms

Following the practice of usage variety of machine learning techniques, for the development of predictive models in biomedicine [47-50] and for medical device management [51-57] this paper presents investigation of application of different machine learning algorithms for prediction of performance of defibrillators for the purpose of optimizing medical device management strategies in healthcare institutions.

a) Decision Tree (DT) algorithm

DT has a tree like structure, where leaves represent outputs and branches represents connection between those outputs and inputs that lead to them. A binary DT separates the data (parent node) into two new nodes (child nodes) using chosen split criterion that determines calculation of the best split features [58]. We applied C4.5 decision tree which uses Gain Ratio as splitting criterion, calculated as:

$$GainRation(A) = \frac{Gain(A)}{SplitInfo(A)} \quad [1]$$

$$Gain(A) = Info(D) - Info_A(D) \quad [2]$$

$$Info(D) = -\sum_{i=1}^{m} p_i log_2(p_i) \quad [3]$$

$$Info_A(D) = \sum_{j=1}^{n} \frac{|D_j|}{|D|} \times Info(D_j) \quad [4]$$

$$SplitInfo(A) = \sum_{j=1}^{n} \frac{|D_j|}{|D|} \times log_2\left(\frac{|D_j|}{|D|}\right) \quad [5]$$

where p_i is the probability that some tuple in partition D belongs to the class C_i. Log function with the base 2 is used because the information is encoded in bits.

Considering the ability of DT to work with the datasets that contain both categorical and numerical data types values and it is good in handling missing values [58], it is the suitable method to be applied in this research with the data of both data types. DT created for the classification is also pruned, with confidence factor 0.25.

Pruning let trees be smaller, less complex and avoids overfitting (i.e., improves correct classification on the test data). Pruning also emphasizes the important features within the database, Figure 2. Pruning of the subtree is done by removing its branches and adding a leaf, maximizing a pruning index [58]. In Figure 2, rectangles (leaves) represent classes. Class marked with YES denotes predicted positive inspection outcome and NO denotes predicted negative inspection outcome. Ellipses represent attributes from the dataset, which serve as splitting features for which Gain Ratio is calculated.

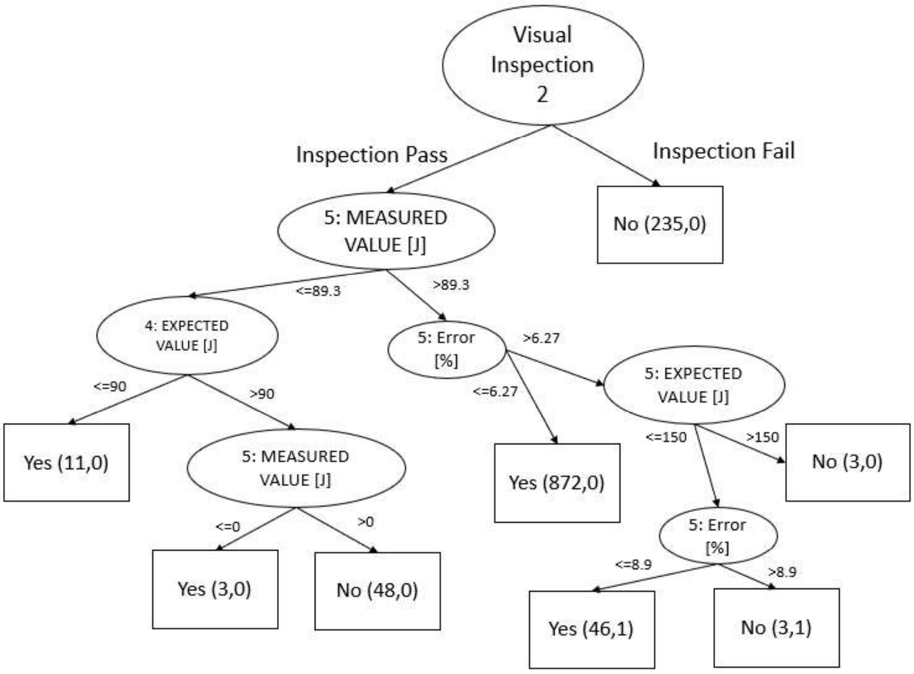

Figure 2. Decision tree model for defibrillator dataset

b) Random Forest (RF) algorithm

Random Forest (RF) is a classification method that combines multiple trees in a forest, so each tree depends of the value of the vector θ_k, randomly selected and distributed among all trees [59,60].

The basic idea of a random forest is to iteratively partition data into boxes using simple rules that minimize the error at each split (node). Each node is split using the best split among all variables, in standard trees. Instead of searching for the best feature while splitting a node, random forest searches for the best feature among a random subset of features. This process creates a wide diversity, which generally results in a better model [59,60].

In random forest algorithm, generalization error is given by [60]:

$$PE^* = P_{X,Y}(mg(X,Y) < 0) \quad [6]$$

$$mg(X,Y) = av_K I(h_K(X) = Y) - max_{j \neq Y} av_K I(h_K(X) = j) \quad [7]$$

Parameter mg in equations 6 and 7 represent margin function which measures the difference between the average number of votes at random vectors and the average vote for any other output. It has only two parameters: the number of variables in the random subset and the number of trees.

The forest consisted of 20 trees and 6 variables in the random subset, is shown in the Figure 3. Different values of trees and random variables were applied, considering the classification accuracy as the fitting function. Each tree yields classification performance within its subsets and using majority vote, the best performance among all trees is selected to be a representative

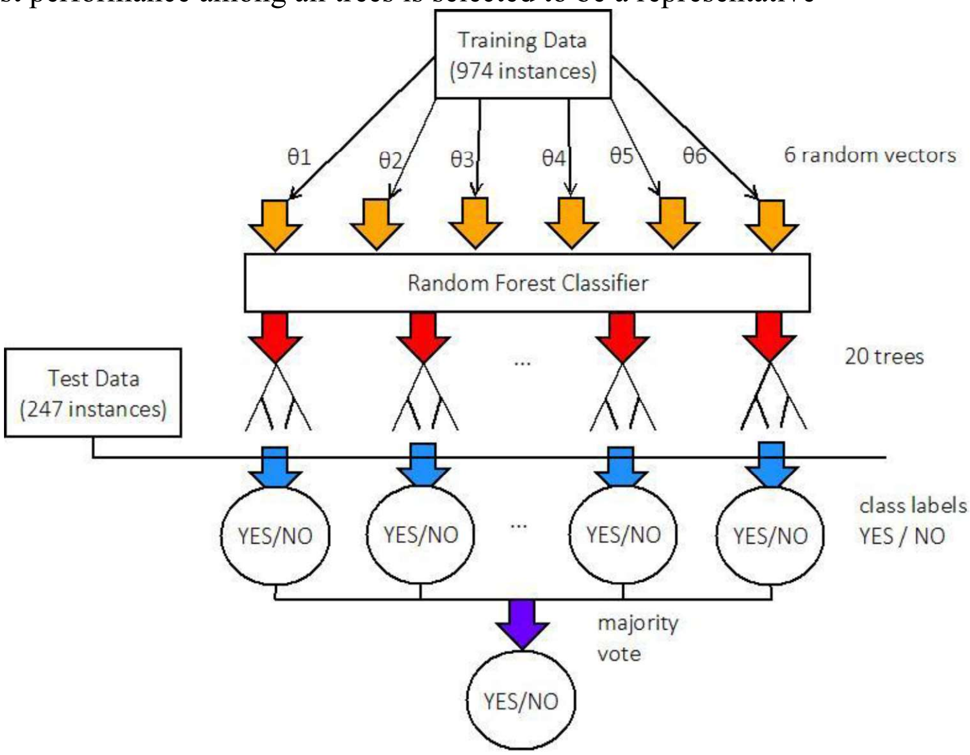

Figure 3. The classification schema of Random Forest classifier

c) k-Nearest Neighbor (k-NN) algorithm

k – Nearest Neighbor (k-NN) is pattern-based classification method that searches for a "nearest" training set T, described by the number of attributes n. The closeness of the sets is described by Euclidian distance, defined as [49]:

$$d(X_1, X_2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2} \quad [8]$$

Role of factor k is to define the border in nearest neighbor area, greater the value is smoother will be the border between classes. Different values of k were applied to the dataset such as k = 3, 5, 7, 9, 11, 13 considering the accuracy as fitting function.

Since the calculation of the distance between the test instances and training instances is computed, k-NN is computationally very intensive method, as no learning is involved. However, k-NN performs very well for datasets with binary class problems. Therefore, k-NN is suitable for the dataset considered in this research, as there are no many instances and it is binary class problem.

d) Support Vector Machine (SVM)

Large-margin separation and kernel functions are the two key concepts in classification with SVM [50]. Firstly, the SVM maps the inputs into a high-dimensional feature space and searches for a separating hyperplane that maximizes the margin between two classes in this space, as shown in the Figure 4. It uses dot product functions, called kernels, to find the optimal hyperplane in the feature space. The resulting solution of the optimal hyperplane can be represented as a combination of a few input points that are called support vectors [61].

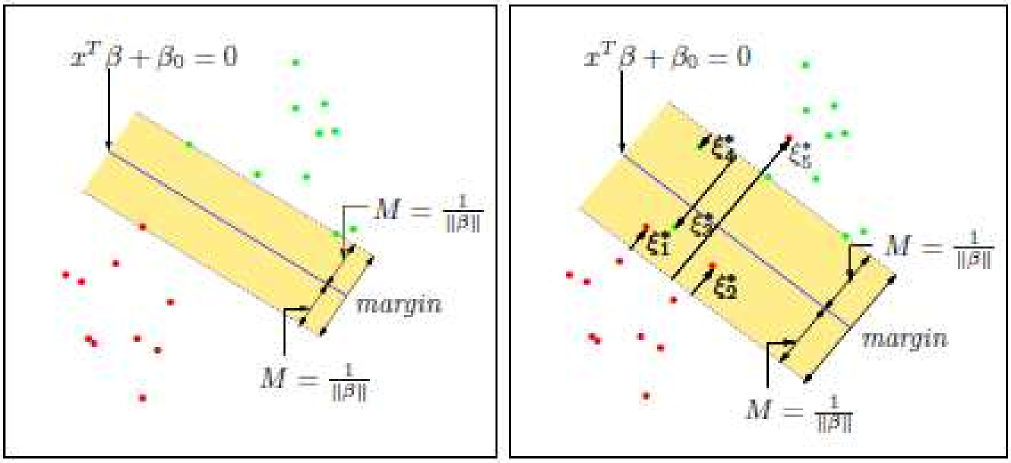

Figure 4. Support vector classifiers: separable (left panel) and non-separable (right panel) cases [61]

Linear, quadratic, polynomial and radial basis (RBF) function represent the four basic kernels. Since the kernel function defines the feature space in which the training set examples will be classified, the selection of an appropriate kernel function is important [62].

In this research PUK kernel was used for SVM classifier with $\omega = 1$ and $\sigma = 1$. These values gave the optimal performance. Different values of ω and σ from the literature were applied [63], however, the classification accuracy decreased. Linear and RBF kernels were also tested, but classification accuracy was significantly reduced.

Before the RF algorithm was proposed, SVM was usually used as benchmark algorithm in research studies in various fields. Therefore, both of them are considered for the classification.

e) Naïve Bayes (NB)

The Naïve Bayes (NB) classifier is in family of "probabilistic classifiers" based on Bayes' rule. According to Bayes' rule, the probability is calculated as [64,65]:

$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)} \qquad [9]$$

where:

- P(h): is independent probability of h (hypothesis) – prior probability,

- P(D): is independent probability of D (data),
- P(D|h): is conditional probability of D given h – likelihood,
- P(h|D): is conditional probability of h given D – posterior probability.

Using conditional independence assumption, it reduces complexity of Bayesian classifiers, which lowers number of parameters to be estimated when modeling P(D|h). It is very useful for large datasets since it does not have complicated iterative parameter estimation. Regardless of its simplicity it surpasses more sophisticated classifiers [64,65].

2.3 Performance Evaluation

Performances of the proposed approaches were evaluated through several performance metrics: true positive (TP) rate, false positive (FP) rate, accuracy and precision. TP rate is the rate of correctly classified instances, whereas FP rate represents the classification error. Accuracy and precision can be calculated according to the formula [65]:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad [10]$$

$$Precision = \frac{TP}{TP + FP} \quad [11]$$

where TN is a true negative rate which represents the number of correctly classified instances belonging to a group of negatively classified instances, whereas FN is a false negative rate which represents the classification error.

3. Results and discussion

Table 2 shows parameters for machine learning algorithms applied to dataset divided into groups and subset as explained in section 2.1. of this paper.

Table 2. Machine learning algorithm parameters

Algorithm	Parameters	
Decision tree (DT)	splitting criterion	Gain Ratio
	pruning	true
	confidence factor	0.25
Random Forest (RF)	random vectors	6
	trees	20
	voting	majority vote
k-Nearest Neighbor (k-NN) algorithm	k	5
	distance	Euclidian
Support Vector Machine algorithm	ω	1
	σ	1
	kernel	P_{uk}

3.1 Results on the dataset with all features

Table 3 presents performance overview of 5 different machine learning algorithms applied on dataset containing all features. As it can be seen from the Table 3, RF performed had the best predicition/classification accuracy. Moreover, RF was successful in detecting all defibrillators with positive inspection outcome whereas it misclassified only one defibrillator with negative inspection. K-NN and SVM were the second-best performance where k-NN performed better than SVM in classification of defibrillator with positive inspection outcome and SVM was better in classification of medical devices with negative inspection outcome. The lowest classification accuracy was obtained using DT algorithm. However, it performed the best in detecting defibrillators with negative inspection outcome among all tested classifiers.

Table 3. Results obtained on dataset with all features

Classifier	Accuracy	True Positive (TP)			False Positive (FP)			Precision		
		Pass	Fail	Average	Pass	Fail	Average	Pass	Fail	Average
DT	98.38	98	100	98.4	0	2	0.4	100	92.7	98.5
RF	**99.60**	100	98	99.6	2	0	1.6	99.5	1	99.6
k-NN	99.19	100	96.1	99.2	3.9	0	3.1	99	100	99.2
SVM	99.19	99.5	98	99.2	2	0.5	1.7	99.5	98	99.2
NB	98.79	100	94.1	98.8	5.9	0	4.7	98.5	100	98.8

3.2 Results on the dataset with features extracted by Info Gain algorithm

When it comes to the feature selection performed with Info Gain, 26 out of 38 features were selected. Considering mentioned selection performance of the four classifiers namely DT, RF, k-NN and SVM did not change when compared to the case when all features were considered; the slight change occurred in detecting defibrillator with negative inspection outcome with NB. The results are summarized in Table 4.

Table 4. Results obtained on dataset with Info Gain Applied

Classifier	Accuracy	True Positive (TP)			False Positive (FP)			Precision		
		Pass	Fail	Average	Pass	Fail	Average	Pass	Fail	Average
DT	98.38	98	100	98.4	0	2	0.4	100	92.7	98.5
RF	99.60	100	98	99.6	2	0	1.6	99.5	100	99.6
k-NN	99.19	100	96.1	99.2	3.9	0	3.1	99	100	99.2
SVM	99.19	99.5	98	99.2	2	0.5	1.7	99.5	98	99.2
NB	**97.98**	100	90.2	98	9.8	0	7.8	97.5	100	98

3.3. Results on the data set with features extracted by Genetic Algorithm

Genetic algorithm with population size 20 selected 11 attributes with the highest impact on the classification result. Among those attributes were visual inspection attributes, performance error and measured values of safety and performance inspection.

As shown in Table 5, the best result was obtained using RF with the accuracy of 100%. DT and SVM gave the accuracy of 99.19 % where DT showed to be better than SVM in classification of defibrillators with positive inspection outcome and SVM was better in classification of defibrillator with negative inspection outcome. The lowest accuracy rate was achieved with NB with accuracy of 95.95%. NB was not comparable to other tested classifiers in classifying negative inspection outcome of defibrillators, achieving TP rate of 80.4%.

Table 5. Results obtained on dataset with Genetic Algorithm Applied

Classifier	Accuracy	True Positive (TP)			False Positive (FP)			Precision		
		Pass	Fail	Average	Pass	Fail	Average	Pass	Fail	Average
DT	99.19	100	96.1	99.2	3.9	0	3.1	99	100	99.2
RF	**100.0**	100	100	100	0	0	0	100	100	100
k-NN	98.79	100	94.1	98.8	5.9	0	4.7	98.5	100	98.8
SVM	99.19	99.5	98	99.2	2	0.5	1.7	99.5	98	99.2
NB	95.95	100	80.4	96	9.6	0	15.6	95.1	100	96.1

3.4 Results on the data set with features extracted by Wrapper algorithm

Wrapper feature selection method combined with k-NN and RF achieved the highest classification accuracy of 99.6%. It showed that measurement errors, measured values were dominant attributes for accurate prediction/classification. RF was better in classifying negative inspection outcome of defibrillator while k-NN was better in classifying positive inspection outcome of defibrillators as shown in Table 6. The lowest classification accuracy was achieved by DT.

Table 6. Results obtained on dataset with Wrapper Applied

Classifier	Accuracy	True Positive (TP)			False Positive (FP)			Precision		
		Pass	Fail	Average	Pass	Fail	Average	Pass	Fail	Average
DT	98.38	98.5	98	98.4	2	1.5	1.9	99.5	94.3	98.4
RF	**99.6**	99.5	100	99.6	0	0.5	0.1	100	98.1	99.6
k-NN	**99.6**	100	98	99.6	2	0	1.6	99.5	100	99.6
SVM	99.19	99.5	98	99.2	2	0.5	1.7	99.5	98	99.2
NB	98.78	100	94.1	98.8	5.9	0	4.7	98.5	100	98.8

Overall, comparative analysis of applied machine learning algorithms is summarized in Figure 5. From Figure 5, it can be seen that the results showed the advantage of the RF method over DT, k-NN, SVM and NB. Feature selection performed with Genetic Algorithm and Random Forest classification algorithm provided the best prediction/classification results for defibrillator performance. Moreover, RF showed to be the most appropriate classifier when it comes to the classification of defibrillator safety and performance, achieving classification accuracy of 99.6% with Info Gain and Wrapper feature selection methods, and data set with all features included. Number of trees and random features did not affect the performance of the RF classifier.

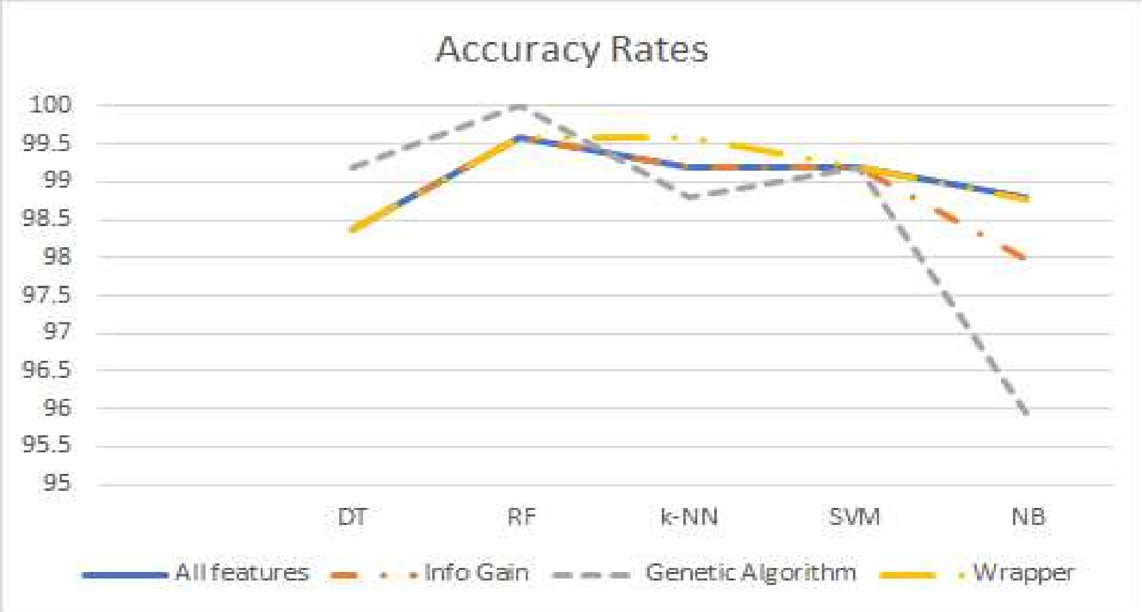

Figure 5. Comparison of accuracy rates for DT, RF, k-NN, SVM and NB

The classification performance of SVM showed that it can also be an applicable method for prediction of defibrillator safety and performance inspection outcome. SVM misclassified only one instance from each group achieving classification accuracy of 99.19%. The results were consistent in all mentioned approaches. The number of support vectors did not affect the performance of the classifier. However, the best performance was achieved by using the P_{uk} kernel in all approaches.

DT achieved the highest classification accuracy of 99.19% with Genetic Algorithm feature selection method, misclassifying only two instances belonging to group of negative inspection outcome medical devices. The performance of DT decreased in other three approaches, where the classification accuracy was 98.38%. These results indicate that DT is also suitable for the defibrillator inspection outcome prediction. Even though the classification accuracy is the same for three approaches (all features, Info Gain and Wrapper), TP and FP rates differ.

The optimal performance of k-NN was achieved with Wrapper feature selector of 99.6%, misclassifying only one instance of negative inspection outcome devices. Additionally, k-NN achieved good performance when Info Gain is applied, and all features are considered, with classification accuracy of 99.19%. The lowest performance was achieved with Genetic Algorithm with classification accuracy of 98.79%. Different values of k were tested but the accuracy value reached the highest number in all four cases with 5 nearest neighbors.

NB achieved the lowest performance results when compared to other four classifiers. However, it performed perfectly in classifying defibrillator with positive inspection outcome. NB with Genetic Algorithm misclassified 20% of the instances which is a huge number when compared with the performance of other four classifiers. However, it can be applicable for the prediction of medical devices with positive inspection outcome.

The characteristics of classifiers together with feature selection algorithms determine the performance of the classifier. However, the accuracy may depend on numerous different factors such as precision of measurements, subjectively examining physical appearance of the device, as well as the division of the data set into training and testing.

It is noticed that this research topic has not been popular in the previous studies. Therefore, this a novel approach in prediction of medical device inspection outcome. Predictive analysis has been recognized as one of the three main areas in which healthcare will benefit from artificial intelligence. [66] The algorithms like proposed one can be developed to identify risk medical devices and order relevant preventive service actions for them. With a predictive maintenance and supervision mechanism for medical devices in healthcare the maintenance events may be scheduled in such a way as to avoid resource crunches, and maintenance events may skip connected parts.

5 Conclusion

In this paper an automated system for defibrillator performance prediction was developed. Such systems can be used to detect hardware deviations which can potentially lead to inaccurate diagnosis and wrong treatments applied to patients.

The study was based on 1221 defibrillator inspection that were carried out according to Legal metrology framework for medical devices implemented in Bosnia and Herzegovina. Prediction/Classification of defibrillator performance was evaluated using 5 different machine learning algorithms on dataset with different number of attributes. The results show that among all tested classifiers RF classifier yielded highest accuracy and proved its significant role in classification and prediction. The performance of RF showed that it can be an applicable in expressing the valuable knowledge to healthcare institutions and laboratories for inspection outcome and testing of medical devices.

This automated system when combined with database of real-time measurements of medical devices, acquired as a result of periodical safety and performance inspections can be powerful tool of post-market surveillance by National Notified Bodies as instructed by new EU Medical Device Regulation. Such system can optimize the costs of medical device maintenance in healthcare institutions.

This study is based on limited number of defibrillators that are mostly used all over the world and that were available for inspection during the three-year period. However, the variety of types of devices was sufficient to derive general conclusion on developed system's performance and significance of prediction of MD performance.

Reference

[1] European Comission – Medical Device Directives, available at: ec.europa.eu/growth/single-market/european-standards/harmonised-standards/medical-devices_en

[2] Food and Drug Administration (FDA), available at: www.fda.gov
[3] International Organization for Standardization, IEC 60601, available at: https://www.iso.org/standard/65529.html
[4] International Electrotechnical Commission, IEC 62353, available at: https://webstore.iec.ch/publication/6913
[5] International Organization for Standardization, available at: https://www.iso.org/standard/59752.html
[6] MAUDE - Manufacturer and User Facility Device Experience, available at: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm
[7] EUDAMED: European Database on Medical Devices, available at: http://ec.europa.eu/idabc/en/document/2256/5637.html
[8] UK government: Vigilance system alerts; Salma M (2009) Vigilance reporting for medical devices in the EU.
[9] European Commission DG Health and Consumers (2013) Guidelines on a medical devices vigilance system. MEDDEV
[10] Kramer DB, Tan YT, Sato C, Kesselheim AS. Postmarket surveillance of medical devices: a comparison of strategies in the US, EU, Japan, and China. PLoS Med. 2013;10(9):e1001519.
[11] Badnjević A, Cifrek M, Magjarević R, Džemić Z, (2018), Inspection of medical devices for regulatory purposes, Series in Biomedical Engineering ISBN 978-981-10-6649-8
[12] Gurbeta, L., Dzemic, Z., Bego, T., Sejdic, E., Badnjevic, A. „Testing of Anesthesia Machines and Defibrillators in Healthcare Institutions“, J Med Syst (2017) 41: 133. https://doi.org/10.1007/s10916-017-0783-7
[13] Gurbeta L., D. Vukovic, Z. Dzemic, A. Badnjevic, Legal metrology procedures for increasing safety and performance characteristics with cost benefits analysis: Case study dialysis machines, IUPESM – The World Congress on Medical Physics & Biomedical Engineering in Prague, June 3—8, 2018
[14] Gurbeta L., Z. Dzemic, A. Badnjevic, Establishing traceability chain of infusion and perfusor pumps using legal metrology procedures in Bosnia and Herzegovina, IUPESM – The World Congress on Medical Physics & Biomedical Engineering in Prague, June 3—8, 2018
[15] Gurbeta L, Badnjević A., „Inspection process of medical devices in healthcare institutions: software solution,“ Health Technol. (2017) Volume 7, Issue 1, pp 109–117, doi:10.1007/s12553-016-0154-2
[16] Sharareh Taghipour, Dragan Banjevic and Andrew K. S. Jardine, Reliability Analysis of Maintenance Data for Complex Medical Devices.
[17] Nichols, T., & Linberg, K. (2002). Apparatus and method to automatic remote software updates of medical device systems.
[18] Linberg, K. R. (2002). 6442433. US Patent. USPTO. Retrieved from
[19] Das S, Dey A, Pal A, Roy N (2015) Applications of Artificial Intelligence in Machine Learning: Review and Prospect. International Journal of Computer Applications 115(9): 31-41.
[20] Horvitz Eric (2006) Machine learning, reasoning, and intelligence in daily life: Directions and challenges. USA.
[21] Le Cun Y, Bengio Y, Hinton G.Deep learning. Nature. 2015 May 28;521(7553):436-44.
[22] Kramer, O. (2017). Genetic Algorithm Essentials. Springer.

[23] Ghamisi, P., & Benediktsson, J. A. (2015). Feature Selection Based on Hybridization of Genetic Algorithm and Particle Swarm Optimization. IEEE Geoscience and Remote Sensing Letters, 12(2), 309–313.

[24] Cruz, J. A., & Wishart, D. S. (2006). Applications of Machine Learning in Cancer Prediction and Prognosis. Cancer Informatics, 2, 117693510600200.

[25] Guidi, G., Pettenati, M. C., Miniati, R., & Iadanza, E. (2013). Random Forest for automatic assessment of heart failure severity in a telemonitoring scenario. Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2013, 3230–3233.

[26] Malhotra, R. (2014). Comparative analysis of statistical and machine learning methods for predicting faulty modules. Applied Soft Computing, 21, 286–297.

[27] Manivannan, M., Najafi, B., & Rinaldi, F. (2017). Machine Learning-Based Short-Term Prediction of Air-Conditioning Load through Smart Meter Analytics. Energies, 10(11), 1905.

[28] Ye, X. (2015). Fault Prediction of Electronic Equipment. https://doi.org/10.14257/astl.2015.83.02

[29] Gurbeta L, Badnjević A., „Inspection process of medical devices in healthcare institutions: software solution," Health Technol. (2017) Volume 7, Issue 1, pp 109–117, doi:10.1007/s12553-016-0154-2

[30] Fluke Biomedical Impulse 7000DP Defibrillator analyzer. [online] Available at: www.flukebiomedical.com

[31] Fluke Biomedical Electrical Safety Analyzer. [online] Available at: www.flukebiomedical.com

[32] Badnjević A., Cifrek M., Magjarević R., Džemić Z. (eds) Inspection of Medical Devices. Series in Biomedical Engineering. Springer, Singapore

[33] Badnjevic A, Gurbeta L, Boskovic D, Dzemic Z. „Measurement in medicine – Past, present, future", Folia Medica Facultatis Medicinae Universitatis Saraeviensis Journal (2015) 50(1): 43-46

[34] Gurbeta, L., Dzemic, Z., Bego, T., Sejdic, E., Badnjevic, A. „Testing of Anesthesia Machines and Defibrillators in Healthcare Institutions", J Med Syst (2017) 41: 133. https://doi.org/10.1007/s10916-017-0783-7

[35] Gurbeta L, Badnjevic A, Dzemic Z, Jimenez E.R, Jakupovic A, "Testing of therapeutic ultrasound in healthcare institutions in Bosnia and Herzegovina", 2nd EAI International Conference on Future Access Enablers of Ubiquitous and Intelligent Infrastructures, 24-25 October 2016, Belgrade, Serbia

[36] Gurbeta L., Alic B., Dzemic Z., Badnjevic A. (2017) Testing of infusion pumps in healthcare institutions in Bosnia and Herzegovina. In: Eskola H., Väisänen O., Viik J., Hyttinen J. (eds) EMBEC & NBC 2017. EMBEC 2017, NBC 2017. IFMBE Proceedings, vol 65. Springer, Singapore

[37] Gurbeta L., Alic B., Dzemic Z., Badnjevic A. (2017) Testing of dialysis machines in healthcare institutions in Bosnia and Herzegovina. In: Eskola H., Väisänen O., Viik J., Hyttinen J. (eds) EMBEC & NBC 2017. EMBEC 2017, NBC 2017. IFMBE Proceedings, vol 65. Springer, Singapore

[38] L. Gurbeta, D. Vukovic, Z. Dzemic, A. Badnjevic, Legal metrology procedures for increasing safety and performance characteristics with cost benefits analysis: Case study dialysis machines, IUPESM – The World Congress on Medical Physics & Biomedical Engineering in Prague, June 3—8, 2018
[39] L. Gurbeta, Z. Dzemic, A. Badnjevic, Establishing traceability chain of infusion and perfusor pumps using legal metrology procedures in Bosnia and Herzegovina, IUPESM – The World Congress on Medical Physics & Biomedical Engineering in Prague, June 3—8, 2018
[40] Visalakshi, S., & Radha, V. (2014). A literature review of feature selection techniques and applications: Review of feature selection in data mining. In 2014 IEEE International Conference on Computational Intelligence and Computing Research. https://doi.org/10.1109/iccic.2014.7238499
[41] Calix, R., & Sankaran, R. (2013). Feature Ranking and Support Vector Machines Classification Analysis of the NSL-KDD Intrusion Detection Corpus. TwentySixth International Florida Artificial Intelligence Research Society Conference.
[42] Osanaiye, O., Raymond Choo, K.-K., & Dlodlo, M. (2016). Analysing Feature Selection and Classification Techniques for DDoS Detection in Cloud. Southern Africa Telecommunication Networks and Applications Conference (SATNAC).
[43] Dimitris, G., Ioannis, T., & Evangelos, D. (2014). Feature Selection for Robust Detection of Distributed Denial-of-Service Attacks Using Genetic Algorithms. Hellenic Conference on Artificial Intelligence.
[44] Kabir, M., Islam, M., & Murase, K. (2010). A new wrapper feature selection approach using neural network. Neurocomputing, 73(16-18), 3273-3283
[45] Panthong, R., & Srivihok, A. (2015). Wrapper Feature Subset Selection for Dimension Reduction Based on Ensemble Learning Algorithm. Procedia Computer Science, 72, 162–169.
[46] Novakovic, J., Strbac, P., & Bulatovic, D. (2011). Toward optimal feature selection using ranking methods and classification algorithms. Yugoslav Journal of Operations Research. An International Journal Dealing with Theoretical and Computational Aspects of Operations Research, Systems Science, and Management Science, 21(1), 119–135.
[47] Shaikhina, T., Lowe, D., Daga, S., Briggs, D., Higgins, R., & Khovanova, N. (2017). Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation. Biomedical Signal Processing and Control. https://doi.org/10.1016/j.bspc.2017.01.012
[48] Lai, D., Palaniswami, M., & Begg, R. (2007). Computational Intelligence in Biomedical Engineering.
[49] Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. Annals of Translational Medicine, 4(11), 218–218.
[50] Üstün, B., Melssen, W. J., & Buydens, L. M. C. (2006). Facilitating the application of Support Vector Regression by using a universal Pearson VII function based kernel. Chemometrics and Intelligent Laboratory Systems, 81(1), 29–40.
[51] Chaudhary P, Kaul P. Factors affecting utilization of medical diagnostic equipment: A study at a tertiary healthcare setup of Chandigarh. CHRISMED J Health Res 2015;2:31623.
[52] Virk, S., Muhammad, A., & Martinez-Enriquez, A. (2008). Fault Prediction Using Artificial Neural Network and Fuzzy Logic [Ebook] (1st ed., p. 154). Mexico: Mexican International Conference.

[53] Reddi, N., Hariharan, S., &Mathur, G. (2007). Toward a Fuzzy Logic Control of the Infant Incubator [Ebook] (1st ed., pp. 1-8). USA: Research Gate.

[54] Amer, G., & Al-Aubidy, K. (2005). Novel Technique To Control The Premature Infant Incubator System Using Ann (1st ed., p. 1). Jordan: Science Direct.

[55] Liao, S. (2004). Expert system methodologies and applications—a decade review from 1995 to 2004 (1st ed., p. 93). Science Direct. Retrieved from https://www.sciencedirect.com/science/article/pii/S0957417404000934

[56] Olivera, J.M., Rocha, L.A., Rotger, V.I. and Herrera, M.C., "Acoustic Pollution in Hospital Environments", Journal of Physics: Conference Series, 332:1–10, (2011).

[57] Sharareh Taghipour, Dragan Banjevic and Andrew K. S. Jardine, Reliability Analysis of Maintenance Data for Complex Medical Devices.

[58] Osmanovic, A., Abdel-Ilah, L., Hodzic, A., Kevric, J., Fojnica, A. (2017). Ovary Cancer Detection using Decision Tree Classifiers based on Historical Data of Ovary Cancer Patients. CMBEBIH 2017, 503-510.

[59] Masetic, Z., & Subasi, A. (2016). Congestive heart failure detection using random forest classifier. Computer Methods and Programs in Biomedicine, 130, 54–64.

[60] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.

[61] Chun-Fu Lin, Lin, C.-F., & Wang, S.-D. (2002). Fuzzy support vector machines. IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council, 13(2), 464–471.

[62] Vanitha, C. D. A., Devi Arockia Vanitha, C., Devaraj, D., & Venkatesulu, M. (2015). Gene Expression Data Classification Using Support Vector Machine and Mutual Information-based Gene Selection. Procedia Computer Science, 47, 13–21.

[63] Üstün, B., Melssen, W. J., & Buydens, L. M. C. (2006). Facilitating the application of Support Vector Regression by using a universal Pearson VII function based kernel. Chemometrics and Intelligent Laboratory Systems, 81(1), 29–40.

[64] Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (2013). Machine Learning: An Artificial Intelligence Approach. Springer Science & Business Media.

[65] Maimon, O., & Rokach, L. (2005). Data Mining and Knowledge Discovery Handbook. New York: Springer Science+Business Media

[66] Compliance navigator for medical devices. Available at: https://compliancenavigator.bsigroup.com/en/medicaldeviceblog/ai-and-the-medical-devices-sector/