Patient care classification using machine learning techniques

Shatha Melhem   
*dept. of Computer Information Systems*  
*Jordan University of Scinece and Technology* Irbid, Jordan  
szmelhem18@cit.just.edu.jo

Ahmad Al-Aiad  
*dept. of Computer Information Systems*  
 *Jordan University of Scinece and Technology* Irbid, Jordan  
aiaiad@just.edu.jo

Muhammad Saleh Al-Ayyad  
*dept. of Biomedical Engineering*

*Al-Ahliyya Amman University*Amman, Jordan  
mayyad@ammanu.edu.jo

*Abstract*— Doctors and Specialists using the results of tests labs of patients to classify their needed care into inpatient care or outpatient care, which is time-consuming and efforts with high potentials from doctors to decide whether the patient needs to be in the hospital and monitored or not, in addition to, the probability of a wrong decision is high, which It results in endangering the patient's life. the purpose of this study is to utilize machine learning in classifying patient care into inpatient or outpatient to reduce the efforts and time expanded by the doctors which reflect on the type of services provided to the patient, also this kind of studies can help in reducing the human errors that result in risks to the patient's life and may increase the total bell of patients which led to harm wallet. machine-learning was utilized to build four models which are Support Vector Machine, Decision Tree, Random Forest, and K-Nearest Neighbors (KNN), that could predict whether the patient should be classified as inpatient or outpatient-based on their conditions and lab test results, and choosing the best model based on the highest accuracy, sensitivity, specificity, precision, false-negative rate, and false-positive rate. (EHR) the dataset has been used that consists of patients’ laboratory test results from a private hospital in Indonesia to build these models and test them. The results show that Random Forest achieved the highest accuracy (77%), Sensitivity (65%), and Precision (72%), respectively, the model also has the lowest false-negative rate (35%), and almost the lowest false positive rate (16%).

Keywords— Machine Learning, Support Vector Machine, Decision Tree, (EHR), Random Forest, K-Nearest Neighbors(KNN).

# Introduction

Health Information System (HIS) appeared with the use of computers and automation in the healthcare sector instead of paper health records, (HIS) is an automated system that uses hardware, software, to record health information about patients that have been collected and share it when needed, and that in order to improve the quality of services that offered to the patients[1].

The rise of (HIS) has transformed the healthcare industry, benefiting both patients and providers, the impacts of using (HIS) on patient care are many, it provides actual care, creates evidence-based medicine and practice guidelines, and guarantees patient safety, and engages the patient in the healthcare process. It also aides communication, increases efficiency, and reduces costs,Electronic Health Record (EHR) and Electronic Medical Record(EMR) are an examples on Health Information Systems that almost used interchangeably ,(EMR) is electronic medical documents or records that replaced by the paper based records, its used in the same health organization and contains only health data, the problem of (EMR) is it can be viewed only to the healthcare organization that has the record so (EHR) created to solve this problem and enabled many different health organizations to view and edit the record with global and national standards.

Automation of healthcare leads to produce very much data or is considered to be a source of big data , Which means integrate machine learning that used health data and health analytics to make conclusions and decisions instead healthcare providers by building models that achieve the highest accuracy with less time and costs, these days machine learning plays a main role in enhancing the efficiency of health sector, it can be used in many subregions of health , machine learning is used to achieve best services with less costs ,it can be used as predictive models to predict illness and treatments to help doctors and physicians to interfere earlier ,predict population health risks by identifying the patterns and flatting high risks signs .

Doctors and specializes one of their missions is to decide whether the patient needed to stay in the hospital or not, In hospitals, treatments, and surgeries are categorized into inpatient and outpatient, Inpatient care and outpatient care are different in the length of staying in the hospital and the kind of medicines, inpatient care like serious cases require overnight hospitalization, patients should stay in the hospital where the procedure has done for at least one night for treatment, during this time, they stay under supervision of nurses or doctors, the services and types of treatments of the inpatient care are maybe complex like delivering a baby ,complex surgeries ,serious illness or some medical issues that requiring monitoring from the doctors, and sever injuries. Outpatient care or ambulatory care is not require hospitalization, an annual exam with the main and primary doctor and a consultation with neurologist are examples on outpatient care ,the emergency cases may considered as outpatient care if the patient leaves the hospital in the same day of arrived , the services and types of treatments of outpatient care are regular procedures like X-ray,MRI’s imaging , Minor surgeries, lab tests and bloodwork, routine physical exams colonoscopies and mammograms,

Patient care classification process considered to be a very important and early step because of its consequence on a patient's health progress, whereas the length of staying affects the eventual bill. and as a doctor who makes the decision sometimes, it could be imprecise and delayed due to human errors, also this decision is time-consuming and efforts whereas the doctors may not have time or effort, and sometimes the doctor may not be available at the right moment to make that important decision which could cause death. so, it a need to search for some techniques that could help both doctors and patients.

Machine learning is an active field in healthcare and has much effectiveness, especially with the appearance of HIS , it improves the delivery of services and faster, and provide the healthcare facility the ability to the singularity, and enhance the ability of medical staff to work , by reducing the effort and time , finally, it helps the patients like trying to protect their wallets from useless spending.

In this paper ,we used machine learning to build four models which are Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF) ,and k-Nearest Neighbors(KNN),that could predict whether the patient should be classified as inpatient or outpatient based on their conditions and lab test results, and training the models on (EHR) dataset that contains patients’ laboratory test results from a private hospital in Indonesia and choosing the best model based on the highest accuracy achieved.

The rest of the paper is organized as follows: section II discussed some related work, section III deals with the Algorithms applied in this work , section IV shows the research method and the system workflow ,section V shows the results of the experiment ,section VI discussed the results, and in section VII the conclusion was made.

# Related Work

Machine learning in healthcare is adopted because of its effectiveness in enhancing the health sector, Bhardwaj et.al [2] ,presented a study of using machine learning in healthcare and concluded that the healthcare is a fast growing sector in world’s economy these days ,more people need care with high costs on patients ,but with adopted technologies like bigdata and machine learning ,better healthcare with low costs will be achieved, Wiens and Shenoy [3] defined the machine learning and proposed a study of (ML) and its effect in transforming healthcare epidemiology with mentioned some examples and applications on it like Predicting Risk of Nosocomial Clostridium difficile Infection and Predicting Reservoirs of Zoonotic Diseases, paper also touched on some considerations and challenges faced the using of (ML ) like the good choice of the target and how to deal with missy health data.

K et al[4] defined various diseases that machine learning could solve and predict it like heart disease , breast cancer, diabetic disease and thyroid disease, the results showed that. SVM gives 96.40% of accuracy for the breast cancer diagnosis, naive Baye provides 86% of accuracy for the diagnosis of heart disease and CART provides 79% of accuracy of detecting diabetic disease.

Alloghan et al[5].applied machine learning techniques to recognize patterns of re-admission of diabetes patients , the models that applied on diabetes dataset are Linear Discriminant Analysis, SVM, k–Nearest Neighbor, Naïve Bayes, J48 and Random Forest, through these models they defined the patterns of re-admission and found that the diabetes patients are more likely to be re-admission are either women, or Caucasians, or outpatients, or those who receive less medication or those who undergo less rigorous lab procedures, treatment procedures.

Beaulieu-Jones et al[6],analyzed submissions to a popular ML for health venue to determine the current state of researches ,the results showed that the trends of machine learning are focus on health because of the easy-to-access, well-annotated data and would benefit from greater clinician involvement to develop into translational applications .

Electronic Health Record (EHR) is an type on health information systems, Heart et al[7],defined EHR and EMR and conclude that when combined the medical information with health records ,thus providing complete view that matches with the definition of patient-centered medical care .it can lead to dramatic amelioration in personalized care as well as public health decision-making, resulting in improved health and wellness. EHR which is considered as source of data that used in analysis and build prediction models to predict many kind of diseases Wu et.al[8] built a model to predict heart failure by using EHR data to train the models and compare the performance by AUC between SVM ,logistic regression and boosting ,the results showed that the prediction models could predict the heart failure before 6 months of the clinical diagnosis.

Zheng et al.[9] applied machine learning and feature engineering on EHR dataset to discover diverse genotype-phenotype associations affiliated with Type 2 Diabetes Mellitus (T2DM) ,the models that had been used in this paper are k-Nearest-Neighbors, Naïve Bayes, Decision Tree, Random Forest, Support Vector Machine and Logistic Regression , the resulted showed that the framework achieved high identification performances (∼0.98 in average AUC), which are much higher than the state-of-the-art algorithm (0.71 in AUC) .

Wang et al. [10] developed a deep learning model to predict the risk of advanced colorectal cancer in Taiwanese adults because it has been ranked as the third cause of death in Taiwan since 2008,the model achieved AUC of 0.922 ,sensitivity of 0.837,specifity of 0.867. Reddy and Delen [11],used RNN-LSTM-based deep-learning methodology to Predict hospital readmission for lupus patients, since Hospital readmission is one of the serious metrics used for measuring the type of services and performance of hospitals, the results showed that RNN-LSTM has a significantly better performance (with an AUC of .70) compared to ANN (with an AUC of 0.66) and logistic regression (with an AUC of 0.63),this is due to that RNN has ability to capture the disease from temporal relationships of the illness status in patients with time.

the previous works , some didn’t give a piratical experiment with results about the effectiveness of using machine learning in healthcare or draw the confusion matrix to specify the sensitivity of prediction of complex and life deadliest human diseases ,and other are considered as backdated analysis which resulted in biases.

This paper shows a practical experiment about the valuable data of EHR and using it as a source to train multiple machine learning algorithms in a way to enhance the services and reduce the loading efforts on the doctors and trying make life easier, the experiment has multiple evaluations including the false-negative rate to make it more robust.

# Algorithms Used

## Support Vector Machine (SVM)

SVM is a supervised machine learning algorithm that used in classification and regression problems, but it most used in classification [12]. SVM mixes various ideas like the VC-theory, maximum margin optimal hyperplane, statistical learning, different kernels, and so on. which makes it especially powerful over the orthodox empirical risk minimization-based methods like the neural networks[13]. It’s work principle about plotting each data item as a point in n-dimensional space where the n is the number of features of the dataset, then finding a hyperplane that differentiates the two classes very good.

## Decision Tree (DT)

It is a non-parametric machine learning algorithm, used in classification and regression ,its learns from decision rules which derived from the data features and then used these rules in predicting the value of the target variable , (DT) consists of root node ,some interior nodes, and a number of terminal nodes, root node has the highest information gain ,both root and interior nodes are represent the overall number of attributes in the (DT) ,and terminal nodes are represent the final classifications[14].

## Random Forest

Machine learning model used in classification, and one of the most used algorithms due to the easy use and often produces a great result, it consists of a large number of decision trees that work together as an ensemble, the total number of decision trees specified at training time by a hyperparameter called “n\_estimators” witch can be any number , at testing time each individual decision tree or estimator predicts the class label, and the class with the highest number of votes becomes the model’s prediction. the voting used to improve the accuracy [15].

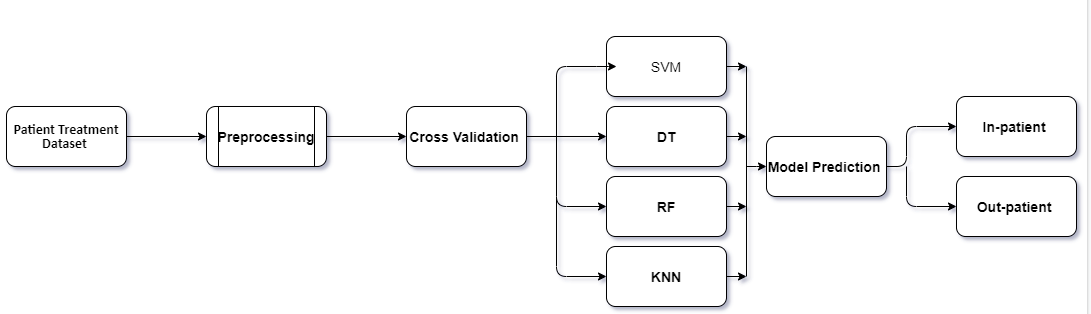
## K Nearest Neighbors (KNN)

KNN can be used in both classification and regression problems , but it is widely used in classification ,in this model for each testing point ,Euclidian distances or similarly between the resting data point and all the training points are calculated ,then determining the class label of the testing point that the most of k nearest training points that have the same class label[16].

# Research Method

In this section ,the major steps of the research were defined which include the dataset ,preprocessing it, then input it to the four models to train them , and finally testing them and choose the best model based on its performance , the system workflow of the research method is illustrated in figure1.

This section is presented with four subsections starting with the description of the dataset and it characteristics that used in this work , the preprocessing that applied on it, finally the four models that been used which are (SVM,DT,RF,KNN) The selection of these four models was based on the ability of dealing with our small dataset, and having relatively memory efficiency, and the easy to implement as shown with KNN which only needed to deal with two parameters (K and the distance function), the evaluation metrices of these models to choose the best model based on it .



1. The system workflow

## Dataset Description

“Patient Treatment” is an Electronic Health Records(EHR) from the Kaggle website created and published in 2020-11-19 , which is collected from a private hospital in Indonesia.it contained 11 attribute as shown in table 1 ,and 3309 records of patients' laboratory tests results which divided into 1992 record as out patients and 1317 record as inpatients ,the dataset used in determining whether the patient care is classified as inpatient care or outpatient care .

1. description of the dataset attributes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| id | Attribute | meaning | Type | Rang of values |
| 1 | HAEMATOCRIT | Patient laboratory test result of haematocrit | Continuous | 13.7-69 |
| 2 | HAEMOGLOBINS | Patient laboratory test result of haemoglobins | Continuous | 3.8-18.9 |
| 3 | ERYTHROCYTE | Patient laboratory test result of erythrocyte | Continuous | 1.4-7.86 |
| 4 | LEUCOCYTE | Patient laboratory test result of leucocyte | Continuous | 1.1-76.6 |
| 5 | THROMBOCYTE | Patient laboratory test result of thrombocyte | Continuous | 10-1121 |
| 6 | MCH | Patient laboratory test result of MCH | Continuous | 14.9-40.8 |
| 7 | MCHC | Patient laboratory test result of MCHC | Continuous | 26-38.4 |
| 8 | MCV | Patient laboratory test result of MCV | Continuous | 54-116 |
| 9 | AGE | Patient age | Continuous | 1-99 |
| 10 | SEX | Patients gender | Binary | M  F |
| 11 | SOURCE | The class target | Nominal | 1=in care patient,  0 = out care patient |

## Data preprocessing

Data preprocessing is an important step before analyzing the data, it enhancing the performance of the models since some models are sensitive and affected by the range of features like SVM and KNN because they are using distances between the value points to determine the similarity , and for efficient representation of data .first data normalization is applied on the dataset before use it to rescale the distribution of numerical attributes to shift the data to be between 0 and 1 and eliminate the units of measurement to easily comparing the features in many places , second step was standardization to transform data to have the Gaussian distribution (mean of Zaro and standard deviation of 1 ) to bring all features with the same distribution that make the models train faster ,the previous two steps done with “sklearn” library in python ,and the third was drop “age” column because the classification depends only on tests results.

## classifiers used

Four classification models were used on “Patient treatment” dataset to classify the patients into inpatient and outpatient using “colab” notebook and python programming language three of the models (DT,SVM,RF) are eager and the fourth(KNN) is lazy , knn has less training time since it saved the training data and wait ,until the testing data is used ,when it does the classification process conducted based on the most related or nearest training data points to the testing points , this type of models have less training time but more time in testing unlike the eager models which construct a model within the training and use it in testing ,this kind of models has more time in training but less time in testing .

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Method | Accuracy % | Sensitivity % | Specificity % | Precision % | FN  Rate | FP Rate |
| SVM | 72 | 45 | 85 | 64 | 55 | 15 |
| DT | 66 | 64 | 69 | 55 | 36 | 31 |
| RF | 77 | 65 | 84 | 72 | 35 | 16 |
| K-NN (K=13) | 74 | 57 | 86 | 71 | 43 | 14 |

“Patient treatment” dataset was divided into training data and testing data using k-cross validation with k=10,cross validation is a statistical method and resampling procedure used to evaluate the machine learning models on specific data samples , it measures the powerful of a model’s generalization and how it will behave with new data that has not seen it before [17].

The following six measures were calculated : Accuracy which defined as the ability of a model to predict the positive and negative records correctly , Sensitivity as the true positive rate and is a metric that evaluates the ability of a model to predict correctly the positive cases , Specificity as the true negative rate [18] is a metric that evaluates the ability of a model to predict correctly the negative cases. Precision also called positive predictive value [19] , False Negative rate (FNR) the percentage of wrong predictions of positive labels as negative, equations (1-6) represent these measures.

Accuracy= (TP+TN)/(TP+FP+TN+FN) (1)

Sensitivity =TP/(TP+FN) (2)

Specificity=TN/(FP+TN) (3)

Precision=TP/(TP+FP) (4)

False Negative Rate=100%-Sensitivity (5)

False Positive Rate= 100%-Specificity (6)

The confusion matrix which is a table with some values that represent the performance of the model as shown in table2 , True positive (TP) is number of correctly predicted records as positive (inpatient),True negative (TN) is number of correctly predicted records as negative (outpatient ), False positive (FP) is number of records that are negative but the model predicted them as positive, and False negative (FN) is number of positive records ,the model predicted them as negative.

1. Confusion Matrix

|  |  |  |
| --- | --- | --- |
|  | Predictive Negative (outpatient) | Predictive Positive (inpatient) |
| Actual Negative (outpatient) | TN | FP |
| Actual Positive (inpatient) | FN | TP |

# Results

Four different classifiers were applied three of them are eager (DT,RF,SVM) and one is lazy (KNN) , after divided the dataset into training and testing using cross validation method with k=10 , then trained the models and test them , the results shown in the table 3 for the six measures (Accuracy, Sensitivity, Specificity, Precision, FNR, FPR) , and as it is noticed ,the best performance based on the accuracy, Sensitivity, Precision was for the Random Forest model with (77%,65%,72%) respectively , and has the lowest False Negative Rate with 35% ,and almost the lowest False Positive Rate with (16%), KNN with K=13 achieved accuracy very close to RF model ,but still the RF model is super passed the KNN in False Negative Rate . table 4 illustrates the accuracy of the KNN model on different values of K, the highest accuracy was at K=13. K values increase the accuracy started to increase accordingly, but at K=15 the accuracy started to decrease and continued decreasing as K increased.

1. measure matrix accuracy
2. KNN accuracy with different values of K

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| k-values | 1 | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 |
| Accuracy | 65 | 66 | 68 | 70 | 73 | 75 | 77 | 73 | 72 |

# Discussion

Medical staffs exert high efforts and time to categorize patients into inpatients or outpatient using the results of tests labs, beside the high efforts and time ,this process has a high probability of a wrong decision due to the human errors , which It results in endangering the patient's life , machine learning has the ability to be effected in the healthcare field in many ways for both doctors and patients, it reducing the efforts and human errors of taken decisions by doctors that may result in endangering the patient's life, in addition, to contribute to speeding of the process of hospitalization, Four classifiers were applied to the patient treatment dataset, the results of the experiment show that the RF model has the highest accuracy (77%), sensitivity (65%), and precision (72%), the model has the lowest false-negative rate (35%) indicates that RF model has the lowest error to Discriminate the class inpatient since this rate is very important to be as little as possible because when the patient is should be hospitalized and classified as an outpatient this means that may his life is in danger and be infected with a serious illness that causes him dead at the end, so the less the false-negative rate is , the better the performance of the model, finally the false positive rate is (16%) this value means that the model mistake in classify (16%) of the negative cases as positive

KNN model has accuracy of (74%) , lower sensitivity ,precision than RF but higher specificity (86%) and lower false positive rate , this is means that KNN model is outperform RF model in its ability to distinguish the negative class label . SVM model has a good accuracy of (72%), lower sensitivity, precision than RF but higher specificity (85%) and lower false positive rate, this is means that SVM model is outperform RF model in its ability to distinguish the negative class label like KNN. DT model has the lowest accuracy of (66%), compared with the previous three models it is bad, but on the other hand, is much better than SVM and KNN in its ability to distinguish the positive class.

Choosing the best model has multiple measures, only accuracy cannot judge the effectiveness of a model, here, the DT has the lowest accuracy but also has high sensitivity and low false-negative rate after RF, although, KNN model has higher accuracy than DT but has a higher false-negative rate, in our domain problem we care about the false negative and positive rates to be as lowest as possible. especially the false-negative which has priority, since it’s more important to predict inpatient care as it has more negative effects on the patient’s life, which could cause death if the patient really needed to be hospitalized and the classifier predicted the opposite rather than predicting outpatient care that has effects on the total budget of the patient and The amount of hospital capacity of patients, Which does not compare with the importance of the patient's life, the RF achieved the lowest false-negative rate and the highest accuracy and sensitivity, and this is the priority measures that take into consideration when the best model was chosen.

# The accuracy is low, and also for the rest five matrix measures, this is due to the low variance of the features, the results should be improved to make the system more robust, and it also needs to be trained on more data, whereas, the dataset that was used contains only 3309 records

# Conclusion and Future work

The process of using the results of test labs of patients to classify their needed care into inpatient care or outpatient care by doctors , is time-consuming and efforts ,also it has a high probability of a wrong decision due to the human errors, which It results in endangering the patient's life, in this paper, machine learning was utilized in a way as a part of the hospitalization systems, by building a system that can classify the patient care into inpatient care or outpatient care, and then take the necessary measures based on the classifier decision, this system can help doctors and Specialists in making their work easier by reducing the time and efforts that expended on taking the decisions, reducing the probability of a wrong decision that is taken by doctors, which could endangering the patient's life ,and also reflects the length of staying that affects the eventual bill of the patient. so, this system can improve the services and lives. In future we want to complete the work and improve the six matrix measures ,and utilizing other machine and deep learning techniques in addition to try to collect more data.

##### References

1. Sayles, N. and Burke, L., 2018. Introduction To Computer Systems For Health Information Technology. 3rd ed. American Health Information Management Association.
2. R. Bhardwaj, A. R. Nambiar and D. Dutta, "A Study of Machine Learning in Healthcare," 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), Turin, pp. 236-241, doi: 10.1109/COMPSAC.2017.164.
3. Wiens, J.,and Shenoy, E. S. (2017). Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology. Clinical Infectious Diseases, 66(1), 149–153. doi:10.1093/cid/cix731h.
4. K. Shailaja, B. Seetharamulu and M. A. Jabbar, "Machine Learning in Healthcare: A Review," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, pp. 910-914, doi: 10.1109/ICECA.2018.8474918.
5. Alloghani, M., Aljaaf, A., Hussain, A., Baker, T., Mustafina, J., Al-Jumeily, D., & Khalaf, M. (2019). Implementation of machine learning algorithms to create https://doi.org/10.1186/s12911-019-0990-x
6. Beaulieu-Jones, B., Finlayson, S. G., Chivers, C., Chen, I., McDermott, M., Kandola, J., ... & Naumann, T. (2019). Trends and focus of machine learning applications for health research. JAMA network open, 2(10), e1914051-e1914051.‏
7. Heart, T., Ben-Assuli, O., & Shabtai, I. (2017). A review of PHR, EMR and EHR integration: A more personalized healthcare and public health policy. Health Policy and Technology, 6(1), 20-25.‏
8. Wu, J., Roy, J., & Stewart, W. F. (2010). Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches. Medical care, 48(6 Suppl), S106–S113. doi.org/10.1097/MLR.0b013e3181de9e17
9. Zheng, T., Xie, W., Xu, L., He, X., Zhang, Y., You, M., ... & Chen, Y. (2017). A machine learning-based framework to identify type 2 diabetes through electronic health records. International journal of medical informatics, 97, 120-127.‏
10. Wang, Y. H., Nguyen, P. A., Islam, M. M., Li, Y. C., & Yang, H. C. (2019, August). Development of Deep Learning Algorithm for Detection of Colorectal Cancer in EHR Data. In MedInfo (pp. 438-441).‏
11. Reddy, B. K., & Delen, D. (2018). Predicting hospital readmission for lupus patients: An RNN-LSTM-based deep-learning methodology. Computers in biology and medicine, 101, 199-209.‏
12. Alaiad, A., Najadat, H., Mohsen, B., & Balhaf, K. (2020). Classification and Association Rule Mining Technique for Predicting Chronic Kidney Disease. Journal of Information & Knowledge Management, 19(01), 2040015.‏
13. Ukil, A. (2007). Support vector machine. In Intelligent Systems and Signal Processing in Power Engineering (pp. 161-226). Springer, Berlin, Heidelberg.‏
14. Swain, P. H., & Hauska, H. (1977). The decision tree classifier: Design and potential. IEEE Transactions on Geoscience Electronics, 15(3), 142-147.‏
15. S. Gharibeh, S. Melhem and H. Najadat, "Classification on Web Application Requests," 2020 11th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 2020, pp. 1-5, doi: 10.1109/ICICS49469.2020.239537.
16. Sarkar, M., & Leong, T. Y. (2000). Application of K-nearest neighbors algorithm on breast cancer diagnosis problem. In Proceedings of the AMIA Symposium (p. 759). American Medical Informatics Association.‏
17. Varoquaux, G. (2018). Cross-validation failure: small sample sizes lead to large error bars. Neuroimage, 180, 68-77.‏
18. Banerjee, P., Dehnbostel, F. O., & Preissner, R. (2018). Prediction is a balancing act: Importance of sampling methods to balance sensitivity and specificity of predictive models based on imbalanced chemical data sets. Frontiers in chemistry, 6, 362.‏
19. Forsyth, A. W., Barzilay, R., Hughes, K. S., Lui, D., Lorenz, K. A., Enzinger, A., ... & Lindvall, C. (2018). Machine learning methods to extract documentation of breast cancer symptoms from electronic health records. Journal of pain and symptom management, 55(6), 1492-1499.‏