Chronic Diseases System Based on Machine Learning Techniques

Sahar A. EL_Rahman [a,b,1,*], Reem Ahmed AlRashed [a,2], Duna Nasser AlZunaytan [a,3], Nada Jahz AlHarbi [a,4], Shroog Abdullah AlThubaiti [a,5], Meelaf Khaled AlHejeelan [a,6]

[a] Computer Science Dep., Princess Nourah Bint Abdulrahman University, Saudi Arabia
[b] Electrical Engineering Dep., Faculty of Engineering-Shoubra, Benha University, Cairo, Egypt
[1] sahr_ar@yahoo.com; [2]reemalrashed2@gmail.com; [3]dalzunaytan@gmail.com; [4]alharbinada22@gmail.com; [5]shroog.1997s@gmail.com; [6]melaf103@gmail.com.
* corresponding author

ARTICLE INFO

Article history
Received January 01, 2020
Revised March 17, 2020
Accepted April 15, 2020

Keywords:
chronic diseases
high blood pressure
kidney chronic diseases
diabetes
machine learning algorithms

ABSTRACT

This paper aims to improve the quality of the patient's life and provide them with the lifestyle they need. And we have the intention to obtain this by creating a mobile application that analyzes the patient's data such as diabetes, blood pressure, and kidney. Then, implement the system to diagnose patients of chronic diseases using machine learning techniques such as classification. Our idea is to recommend a lifestyle for the patient and make the doctor participate in it by writing notes. In this paper, machine learning classifiers were used to predict whether the person is prone to some chronic diseases. Blood pressure, diabetes and kidney are considered in this work. For hypertension, Tree algorithm has shown 100% accuracy, which was the best one. Chronic Kidney Disease (CKD) is a significant public health concern with rising prevalence. With a set of considered attributes such as specific gravity, albumin, serum creatinine, hemoglobin, packed cell volume and hypertension used to predict if the person has Kidney disease or not. For kidney, Random Forest algorithm has shown 100% accuracy, which was the best one among other algorithms tested. We considered attributes such as pregnancies, glucose, blood pressure, skin thickness, insulin, diabetes pedigree function, age and BMI of a person to diagnose whether a patient has diabetes based on specific diagnostic measurements or not. For diabetes, neural networks have shown the best accuracy. It was 76.3%.

1. Introduction

People with chronic diseases are at increased risk of poor health outcomes and health disparities. Due to the massive lack of health care monitoring of these diseases that needs daily control at home and hospital when they go to their doctor. Handling of patient data and records and the treatment of chronic diseases is hard and time-consuming. There is no technology employment to help doctors make diagnoses faster and much more accurate as well as recognize patients who may take advantage of the new types of treatments. As we start with health measurements recording, the patients of chronic diseases usually need to take note of daily measurements on paper. Sometimes as well Patients forget to measure or to record them. These situations said before, will get the doctor not fully to understand the case because of the lost measures. We need to rely on a better way to store the measurements for a longer time and in a safe place. Moreover, it's difficult for the patient to decide on a food diet and exercises in a way that suits his measurements and his health statuss. Given the problems before in patients, society includes the need to provide communication for people who have chronic disease the progress toward achieving the goal in higher quality health care and improved population health outcomes. The proposed system intends to streamline the process of monitoring the patient's glucose level, blood pressure and kidney by both the patient and their doctor(s). A mobile application will be developed for reading, storing, and sharing data of blood pressure, glucose level and kidney. The solution will offer to continue and remote monitoring of the patient's health. Patient's data may be handled in a more efficient way which is using machine learning and classification algorithms to help doctors in diagnosis and patient's in following their medical status. Therefore, diagnosis will be faster and more accurate. We are planning to replace paper measurement recording by saving them in the application's database, so it is kept for a longer time and in a safe place. Furthermore, we will notify patients when they forget to record measurements so the doctor will fully understand the case because of the all measures are stored. Because of the complexity of deciding the daily lifestyle, patients tend to ignore this important thing that affects health status. Moreover, the application will allow the patient to track their exercise and diet, depending on their health conditions. The patients will have an exercise recommendation that is consistent with his status and doesn't tire or hurt them. The system domain is healthcare domain where we will focus on serving patients of all ages by organizing their measurements for some chronic diseases and try to avoid problems early by giving them warnings when there is a noticeable difference in their measurements as well as providing the appropriate diet and exercise.

The patients can follow their medical condition and see any developments in the situation, through the application. Moreover, the doctor can view the patient's record and medical status, so that they can write medical notes and answer any questions or concerns sent by the patient. We will use data mining and machine learning algorithms to build the model.

The main advantage of our system is that our system combines three different major chronic diseases, while most of the systems focus on only one disease. We are planning to implement three different machine learning models for each disease dataset for the classifications. For algorithms, we may use the Decision tree, Support Vector Machine, Artificial Neural Network, k-nearest neighbor (k-NN) and Naïve Bayes classifier. Three algorithms for each disease's model and see what the best one of them is. Datasets are mostly would be separated for each disease. Accuracy will be defined in Experimental results and analysis.

The organization of the paper is as follows. Section 2 introduces related works survey. In section 3, the background information are given. In section 4, the proposed methodology for machine learning is discussed. In section 5, requirement analysis , In section 6, requirement analysis . In section 7, the results and analysis. In section 8, the conclusion and future works are presented. And finally, section 9, refences were provided.

2. Related Work

2.1. Hypertension related work survey

Hudson Fernandes Golino and others have written a paper about Predicting Increased Blood Pressure Using Machine Learning algorithm named Classification tree. The main features for prediction were obesity specifically blood mass index (BMI), waist (WC), hip circumference (HC) and waist-hip ratio (WHR). The dataset was first split into two subsets, one for each sex. For women, it contained 225 records and another for men with 153 records. Fifteen trees were calculated in the training group for each sex, using different numbers and combinations of predictors. The overall sensitivity and specificity for the best model (tree no. 15) was 72% and 86.25% for women. For men's testing sample, it showed that sensibility decreased to 52.38% and the specificity decreased to 69.70% (Golino et al. , 2014).

Zhang B and others predicted blood pressure from physiological index data. They used a support vector machine regression (SVR) algorithm to solve the key gap between the need for continuous measurement for prophylaxis and the lack of an effective method for continuous measurement. They have used a real-time Data-set with Eighteen participants (12 males, 6 females). The main features include PTT, HR, PPG, I, II, III, aVF, aVR, aVL and SpO2. The results of the algorithm were compared with those obtained from two classical machine learning algorithms, linear regression (LR) and backpropagation neural network (BP) with four different feature combination. The experimental results showed that the SVR model can accurately and effectively predict blood pressure for the combination of all the features it has achieved 99.00% and 99.43% accuracy for Ps (systolic) and Pd (diastolic) prediction (Zhang et al., 2019).

In this paper, Satyanarayana and others wrote a paper to predict high blood pressure. Main features for prediction was age, anger, and anxiety (AAA) and obesity (+), cholesterol level (+). They have used different classifiers for prediction. Used classifiers are supported by WEKA (Waikato Environment for Knowledge Analysis). Algorithms are J48, Naïve Bayes classifier, Simple logistic regression, REP, and Random Forest. a Real-time dataset in a medical diagnostic center. The Random forest algorithm has shown 87.5 % accuracy, Simple logistic regression has 71 % accuracy, Naïve Bayes has 66.25% accuracy, J48 has 83.5 % accuracy, and REP tree accuracy was 79 %. So, the Random Forest has shown the best performance (Nimmala et al., 2018).

Melvut Ture and others in this paper, the performance of classification techniques are compared to predict the risk of essential hypertension disease. The dataset was a real time data set with 694 records. Main features for prediction was Age, sex, family history of hypertension, smoking habits, lipoprotein(a), triglyceride, uric acid, total cholesterol, and BMI. They used more than one algorithm to predict risk of hypertension such as Multilayer Perceptron (MLP) and Radial Basis Function (BRF) performed better than Logistic Regression (LR), Flexible Discriminant Analysis (FDA), Multivariate Additive Regression Splines (MARS), Chi-squared Automatic Interaction Detector (CHAID), Quick Unbiased Efficient Statistical Tree (QUEST) and Classification and Regression Tree (CART) techniques. They compared the performance by Sensitivity and Specificity and Predictive Rate (PR) of three decision trees, four statistical algorithms, and two neural networks. Performance is good if the Predictive Rate (PR) is greater. In this paper, they found that Multilayer Perceptron (MLP) and Radial Basis Function (RBF) performed better than the others where the sensibility and specificity for the best two model was 71.79% and 66.67% (Ture et al., 2005).

In this paper, JIAN-HUI WU and others, they represent Risk Assessment of Hypertension in Steel Workers. In this paper learning vector quantization (LVQ) neural network algorithm and the Fisher SVM. They used the LVQ and SVM to estimate the hypertension risk of steel workers. Then the result is better with LVQ algorithm because its accuracy is 93.33%. And the accuracy result with SVM is 76.67%. Then the accuracy in LVQ is more than in SVM (WU1 et al., 2019).

Robabeh Abbasi and others wrote a paper about the Long-term Prediction of Blood Pressure Time Series Using Multiple Fuzzy Functions. The paper firstly introduces a new prediction method for time series prediction based on fuzzy functions (FF) in multi-model mode and applies it for forecasting MAP time series as a new application applies it for forecasting MAP time series as a new application. The proposed model consists of three steps. The first step is to estimate the missing values in the MAP time series by a linear interpolation method to denoise it by using the empirical mode decomposition (EMD) procedure. The second step is to reconstruct the phase space. The last step is to apply a predictive model based on fuzzy functions (FFs) and they applied it to three different Fuzzy functions FRB (ANFIS), FFs-based ANFIS and FFs-based MARS models. Results show that the proposed FF-based MARS model is more accurate than ANFIS (adaptive neuro-fuzzy inference system) and FF-based ANFIS (Abbasi et al. ,2014).

2.2. Diabetes related work survey

In this paper the Benamina et al, they presented a study about an expert's information is intense experience through practice and education in a clinical healthcare field. Case-based reasoning (CBR) has emerged as a major research area within the search for a problem-solving paradigm and decision support, and the Cases retrieval is the important step in (CBR). They were tested using fuzzy logic and data mining to improve the response time and the accuracy of the retrieval of similar cases. The goal of fuzzy logic to simplify the complexity of computing the similarity between diabetic patients who require different monitoring plans. After comparing the accuracy of the Fispro Fuzzy DT is 81%, Weka Decision tree 73% and JColibri k-NN is 66%. The results indicate how the proposed fuzzy decision tree helps to improve the accuracy of diagnosing diabetes mellitus patient's classification and retrieval step of CBR reasoning, also monitoring plan corresponding to the result, with the classification option (Benamina et al., 2018).

Fioravanti et al, wrote about Mobile health systems are assisting in the provision of continuous and personalized health services to chronic diseases such as diabetes. This paper was proposed to solve the compliance problems about how to self-monitoring and healthy lifestyles, by automatic generation of feedback messages containing clinical guidelines and patient’s healthy lifestyle. A mobile health system was implemented and tested in a small-scale exploratory study for 4 weeks. The results are improving lifestyle care, improving patient outcomes and approval using of technologies. The type of messages with total number Warnings 2,591, Reminders 32, Tips and goals 56 and the total is 2,679 of messages (Fioravanti et al., 2015).

Park et al, researched about the ease of use, proximity to the user and various health maintenance applications enable mobile tablet devices to help type 2 diabetes patients self-management. The goal of this paper was to talk about technologies are the tools that allow them to participant digital literacy, improve patient care and strengthen healthcare systems. The results from online surveys of five months to the patients’ participation in mobile health activities perceived to have improved were recording biometric data (71.4%, n = 21), tracking what they did on a daily basis (67.9%, n = 19), deciding what to eat or what not to eat (60.7%, n = 17), and gaining confidence in managing diabetes (75%, n = 21). So, the program demonstrated that patients would feel better in the continuous access to the internet that the devices provide for skilled users and unskilled users have the challenge to use digital technologies (Park et al., 2016).

Diabetes disease increased with a high rate due to the unhealthy lifestyle around the world. Habibzadeh et al, studied the consequences of two groups chosen randomly with type 2 diabetes that first group they learn to take care of themselves and the second group is the usual lifestyle. After doing Statistical analysis using independent t-test the result of self-organization (t =11.24, p < 0.001), self-adjustment (t = 7.53, p < 0.001), interaction with health experts (t = 7.31, p < 0.001), blood sugar self-monitoring (t = 6.42, p < 0.001), adherence to the proposed diet (t = 5.22, p < 0.001), and total self-management (t = 10.82, p < 0.001) were increased in the first group because they learn about take care of themselves. So, it is vital to understand and encourage the importance of patient self-management. They recommended using the self-management of society and patients with type 2 diabetes disease (Habibzadeh et al., 2017).

In this paper Tamilvanan and Bhaskaran, study about data mining is the process of discovering patterns in large data sets. They use Classification techniques to find out in which group each data instance is related within a given dataset. It is used for classifying data into different classes to proof which is more efficient of Naive Bayes, Random Forest, and NB-Tree algorithms. It is used in the medical dataset for diabetes disease. After computation techniques the result using such as Confusion Matrix and Average of the Precision 0.759, Recall 0.763, F-Measure 0.76, and Accuracy 0.763 for the Naive Bayes, Precision 0.731 Recall 0.736 F-Measure 0.733 Accuracy 0.735 for the NB-Tree and Precision 0.744 Recall 0.749 F-Measure 0.745 Accuracy 0.748 for the Random Forest. The efficient classification algorithms are Naive Bayes more accuracy 76.3% and less error rate of 23.7%. Also, there are other data mining techniques can also be used for prediction e.g. Clustering, Association Rules (Tamilvanan and Bhaskaran, 2017).

Iyer, A., S, J. and Sumbaly, R. In their study, used both Naive Bayes and J48 Decision Tree algorithms to predict diabetes. Data was divided into 70:30 of training and testing splits, respectively. The training split was used in the model development, meanwhile, the testing split was used in the model evaluation. Naïve Bayes accuracy reached 79% while J48 was 76%. They concluded that experimental results showed the adequacy of using the proposed model in the future for the diagnosis of diabetes.

Jayalakshmi, T. and Santhakumaran, A. work aimed at using artificial neural networks (ANNs) for the prediction of diabetes. The researchers identified challenges with the dataset including missing values, which ANNs have difficulty in interpreting. The paper claimed that with the use of K-Nearest Neighbor (KNN) algorithm to replace missing values, the accuracy of diabetes prediction reached 99%.

S. Mustafa, K., Watan, I., G. AND Enteesha, D.,M. work was unique in its effort to address noise in the dataset. They used K-Means Clustering to split the data into several groups, and manually eliminated the minority class in some of these groups. Decision tree was trained and tested on the remainder of the data points and reached a 98% accuracy in predicting diabetes.

2.3. Kidney related work survey

Due to the 2 Billion Riyals assigned for renal replacement therapy in Saudi Arabia, Alassaf R. and others (2018), proposed four machine learning techniques: ANN, SVM, NB, and k-NN to develop a model that seek to reduce the number of patients and the costs required for therapy, by diagnosing chronic kidney disease accurately. The dataset has been collected from King Fahd University Hospital (KFUH) in Khobar. From their experimental result, ANN, SVM, Naïve Bayes achieved a testing accuracy of 98.0% while k-NN has achieved an accuracy of 93.9%. The following are some of the earlier works in the field of using machine-learning algorithms to diagnose CKD. They used the same dataset from the UCI Machine Learning Repository with different machine learning algorithms.

Alimran A. and others (2018), performed a comparative analysis with three modern classifiers namely: Logistic Regression, feed-forward neural networks and wide & deep learning to diagnose CKD. The dataset is obtained from the UCI Machine Learning Repository. The performance of these algorithms was measured by f1-score, precision, recall and AUC score were used for logistic regression and an additional loss score was considered for the feed-forward neural networks and wide & deep model. However, Feed-forward neural network resulted in 0.99 f1-score, 0.97 accuracies, 0.99 recall and 0.99 AUC score as the best performing CKD diagnostic method. Whereas Logistic regression generated the lowest result among all and wide & deep learning with a larger number of hidden layers and neurons found to be efficient for bigger datasets.

Using the same dataset, Charleonnan A. and others (2016), developed a system which is used to predict chronic kidney disease using machine learning predictive models: including K-nearest neighbors (KNN), support vector machine (SVM), logistic regression (LR), and decision tree classifiers. The authors compared the performance of the four classifiers with SVM. Five-time averages of sensitivity and specificity were illustrated and showed that SVM's sensitivity is slightly higher than other methods at 0.99, where Logistic sensitively at 0.94, Decision Tree at 0.93 and KNN's 0.96.

Using the UCI Repository, Salekin and Stankovic (2016) have developed an automated machine learning solution to detect CKD and explore 24 parameters related to kidney disease. The dataset used for evaluation suffers from noisy and missing data. With three different classifiers, they evaluate solutions: K-NN, RF, and neural networks. To reduce over-fitting and recognize the most significant predictive characteristics for chronic kidney disease, they used two methods to perform feature reduction: wrapper method and LASSO regularization. Also, through cost analysis considering all 24 attributes they identify a cost-effective highly accurate detection classifier using only 5 attributes: specific gravity, albumin, diabetes mellitus, hypertension, and hemoglobin. By using this approach, they achieved a detection accuracy of 0.993 using F-measure.

In another study conducted by W.H.S.D Gunarathne and others (2017), Classification models with different classification algorithms used to predict the CKD and non-CKD status of the patient. Four proposed machine learning techniques: Multiclass Decision Forest, Multiclass Decision Jungle, Multiclass Logistic Regression and Multiclass Neural Network. According to the results obtained using Microsoft Azure Machine Learning Studio it shows that decision forest performs with the predictive accuracy of 99.1%, decision jungle performs with the predictive accuracy of 96.6%, logistic regression performs with the predictive accuracy of 95.0% and neural network performs with the predictive accuracy of 97.5%. It was observed from the comparison that the model with the Multiclass Decision forest algorithm performed best for the reduced dataset with the 14 attributes with an accuracy of 99.1%

In another research study by M.P.N.M. Wickramasinghe and others (2017), The aim is to control the disease by using classification algorithms to identify a suitable diet plan for the CKD patients according to their blood potassium level. The experiment is performed on different classification algorithms like: Multiclass Decision Jungle, Multiclass Decision Forest, Multiclass Neural Network and Multiclass Logistic Regression. These algorithms are applying to the test result obtained from patients' medical records using Microsoft Azure Machine Learning Studio. The results showed that Multiclass Decision Forest algorithm gives the overall accuracy as 99.17%, Multiclass Decision Jungle algorithm gives 97.50%, Multiclass Logistic Regression algorithm gives 89.17% and Multiclass Neural Network algorithm gives 82.50%.

Sinha P.K. , (2015), performed a comparative study on the performance of Support vector machine (SVM) and K-Nearest Neighbor (KNN) classifier in classifying CDK. The performance of the classifiers was evaluated based on accuracy, precision, recall, and F-measure. Their experimental results showed that KNN yielded 78% accuracy, 85% precision, 76% recall and 80% f-measure scores which were better than SVM. However, SVM produced the highest recall value of 100%.

Vijayarani D.S. , Dhayanand M.S. (2015), presented a research to predict kidney related diseases by using Artificial Neural Network (ANN) algorithm and Support Vector Machine (SVM) algorithm. To compare the performance of selected two algorithms based on their execution time and accuracy was the goal of this research work. It is concluded that the performance of ANN algorithm is better than SVM algorithm with the accuracy of 87%, from the experimental results.

Siddheshwar Tekale et al. (2018) ,Has compared results of different models. And finally, they concluded that the Multiclass decision tree algorithms gives the accuracy of 91.75% and SVM gives accuracy of 96.75% for the reduced dataset of 14 attributes.

Chakrapani et al. (2019) ,In this paper, different classification algorithms were used to predict chronic kidney disease and finally the concluded that the ANN algorithms give the accuracy 99.9% for the dataset of 25 attributes.

S.Ramya, Dr. N.Radha (2016) , In this research work, classify the different stages of chronic kidney disease according to its severity. The classification algorithms that have been considered for predicting chronic kidney disease are Backpropagation Neural Network, Radial Basis Function and Random Forest. The models are evaluated with four different measures like Kappa, Accuracy, Sensitivity and Specificity. From the experimental result, the Radial Basis Function is a better accuracy for predicting chronic kidney disease and it attains the accuracy of 85.3%.

3. Background Information

3.1 Hypertension related work survey Hypertension Background Information

Hypertension is a term used to describe high blood pressure. Blood flow is based on the beat from which blood is pumped by the heart. The pressure does not always stay at the same rate. This varies at a specific point in time depending on the activities. Hypertension results in an abnormal pressure of the main arteries for a long time. (Cunha,2011).

3.1.1. Hypertension Types

There are two main categories of hypertension flow. These include hypertension of the essential(primary) and secondary levels.

Essential (Primary) hypertension: Essential hypertension often known as primary hypertension; primary hypertension affects 95% of people suffering from the disease. (Carretero et al. 2002). Although there is no clear medical cause for essential hypertension, many factors may contribute to it, including salt sensitivity, kidney chemical imbalance, insulin resistance, family history and age. Essential hypertension is usually seen in combination with type 2 diabetes, high cholesterol, and central obesity.

Secondary hypertension: In this case, high blood pressure is caused by another illness, such as kidney disease or certain cancers (especially adrenal gland cancer). Most people with secondary hypertension are likely to have an endocrine or kidney defect that, when corrected could get blood pressure back to normal levels. Secondary hypertension can also be caused by certain medications (especially NSAIDS [Motrin/ibuprofen] and steroids) (Milechman et al., 2014).

3.1.2. Hypertension Representation:

Blood pressure is represented as the ratio of Systolic Blood Pressure (SBP) over Diastolic Blood Pressure (DBP).

$$BP = SBP / DBP \tag{1}$$

Systolic blood pressure is the pressure in the arteries as the heart contracts and pumps the blood forward into the arteries, while diastolic is the pressure resulting from the contraction of the arteries (Zareian,2004; Cunha 2011).

3.1.3. Measurements Ranges:

Table 1. Blood pressure range table

Blood pressure	Low	Normal	Borderline	High
Systolic	<90	90-130	131-140	140
Diastolic	<60	60-80	81-90	90

Impact factors: There are many reasons for increasing blood pressure as said before. The occurrence is correlated with many underlying factors. Such factors include age, heavy consumption of salt, lack of exercise, and genetic factors (Cunha et al. ,2011).

Impact of age: Among humans, ageing is a continuous and cumulative process that results among decreased physiological function across all body systems (Franceschi et al., 2008). Age affects the heart performance in pumping blood. If the age of a person increases, then pathways of the heart's pacemaker system deposit fat, which will affect the heart performance while pumping the blood. When we age, the elasticity nature of arteries also decreases, they become stiff. In such a situation, to pump the blood throughout the body through arteries, the heart has to push the blood using more force, this may, in turn, increasing blood pressure (Nimmala et al. ,2018).

Impact of Obesity: Excessive processing of body fat and weight is a significant public health concern. The Body Mass Index (BMI) can be determined based on weight and height. Additional fat in the body needs oxygen and nutrients to live. It increases the workload of the heart because it must pump more blood through additional blood vessels. The more circulating blood also means more pressure on the artery walls (Nimmala et al. ,2018).

3.2 Diabetes Background work

3.2.1 Types of Diabetes

Type 1 diabetes: Autoimmune reacts where the body's defense system attacks cells that produce insulin. As a result, the body produces very little insulin, which is known to cause type 1 diabetes. The exact causes of this are said to be linked to a combination of genetic and environmental conditions. Type 1 diabetes can develop in children and adolescents and may occur at any age. When you have type 1 diabetes, your body produces very little or no insulin, meaning we need to inject insulin daily to keep your blood glucose levels under control. If insulin cannot reach people with type 1 diabetes, they will die. People with type 1 diabetes need insulin injections daily to control their blood glucose levels. Type 1 diabetes in a family member slightly in-creases the risk of developing the disease, and environmental factors and exposure to certain viral infections are risk factors for type 1 diabetes. The search for risk factors in type 1 diabetes remains restricted (WebMD, 2019).

Symptoms of type 1 diabetes:

- Abnormal thirst and dry mouth
- Sudden weight loss
- Frequent urination
- Lack of energy, fatigue
- Continuous hunger
- Blurred vision
- Bedwetting (WebMD, 2019)

Type 2 diabetes: The most common among adults and represents about 90% is type 2 diabetes of all cases of diabetes. This type is characterized by insulin levels not working properly, blood sugar levels continue to rise, and more insulin is released due to insulin resistance as the body does not respond to insulin completely. For some people with type 2 diabetes, this may eventually lead to depletion of the pancreas, leading to less and less production of insulin, leading to high blood sugar levels (hyperglycemia). When you have type 2 diabetes, the body does not benefit from the insulin it produces. Due to high levels of obesity, physical inactivity and malnutrition it is increasing in children, adolescents and younger adults. Type 2 diabetes may be diagnosed in older adults (WebMD, 2019).

The symptoms of type 2 diabetes:

- Excessive thirst and dry mouth
- frequent urination
- Lack of energy, fatigue
- Slow wound healing
- Recurrent infections of the skin
- Blurred vision
- Tingling or numbness in the hands and feet (WebMD, 2019)

Risk factors of type 2 diabetes:

- Family history of diabetes
- Overweight
- An unhealthy diet
- Physical inactivity
- Increase age
- High blood pressure
- Ethnic origin
- Impaired glucose tolerance (IGT)
- History of gestational diabetes
- Malnutrition during pregnancy. (WebMD, 2019)

3.2.2 Patients of diabetes lifestyle

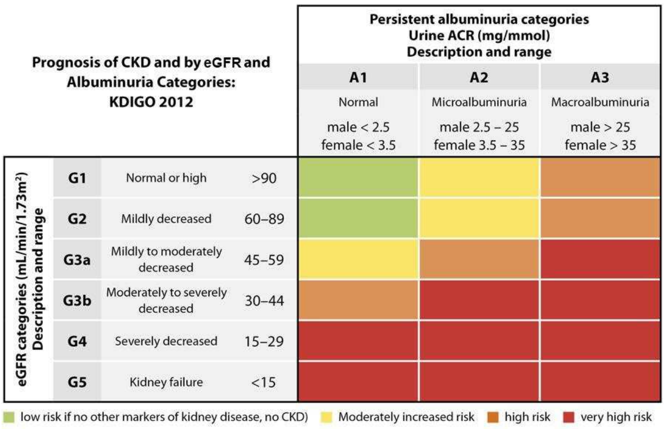
This has proved during 20 years of medical research that a healthy lifestyle can prevent diabetes type 2 from occurring in the first place and even reflect its progress, a substantial and long-term study. Patients can manage and monitor your diabetes by focusing on changes in your lifestyle. Manage stress, eat healthy, physical activity. These things affect the quality of diabetes lifestyle. Patients must be active in walking and doing some exercises. (Chen et al., 2015). Make better food choices program such as Mediterranean-an, vegetarian and lower carbohydrate. (Coughlin, 2017). Stress reduction is beneficial for the improvement of health such as yoga postures, breathing exercises (pranayama), meditation. (Yadav et al., 2015). Eating healthy is essential for people with diabetes because of its effect eating food on blood sugar (WebMD,2019). Should be focused on eating as much as the body needs by relying on specific tactics to eat well by keeping a food record. Can use a smartphone to help them regulate their diet (Monique Tello, 2019).

3.3 Chronic kidney disease background

Chronic kidney disease (CKD) is abnormal kidney function and structure. It is common, frequently unrecognized and often exists together with other conditions. It occurs when the kidneys are damaged and could not filter the blood properly. Chronic kidney disease, which is also called chronic kidney failure, describes the gradual loss of kidney function. Your kidneys perform many vital functions, such as filtering wastes from the blood, manage fluids in your blood, control blood pressure. (Noia T.D et al., 2013) There are five stages of CKD. The most serious one is stage 5 because, at this stage, the kidneys are unable to do most of their functions. The stages are determined based on the patient's Glomerular Filtration Rate (GFR).

3.3.1 Glomerular Filtration Rate (GFR)

Glomerular filtration rate (GFR) is the number used to figure out a person’s stage of kidney disease. A math formula using the person’s age, race, gender and their serum creatinine can be used to calculate a GFR. A doctor will order a blood test to measure the serum creatinine level. Creatinine is a waste product that comes from muscle activity. When kidneys are working well, they remove creatinine from the blood. As kidney function slows, blood levels of creatinine rise. (About Chronic Kidney Dis-ease, n.d)



Prognosis of CKD and by eGFR and Albuminuria Categories: KDIGO 2012				Persistent albuminuria categories Urine ACR (mg/mmol) Description and range		
				A1	A2	A3
				Normal	Microalbuminuria	Macroalbuminuria
				male < 2.5 female < 3.5	male 2.5 – 25 female 3.5 – 35	male > 25 female > 35
eGFR categories (mL/min/1.73m^2) Description and range	G1	Normal or high	>90	🟩	🟨	🟧
	G2	Mildly decreased	60–89	🟩	🟨	🟧
	G3a	Mildly to moderately decreased	45–59	🟨	🟧	🟥
	G3b	Moderately to severely decreased	30–44	🟧	🟥	🟥
	G4	Severely decreased	15–29	🟥	🟥	🟥
	G5	Kidney failure	<15	🟥	🟥	🟥

🟩 low risk if no other markers of kidney disease, no CKD) 🟨 Moderately increased risk 🟧 high risk 🟥 very high risk

Fig 1. Classification and prognostic risk of CKD (eGFR) and presence of albuminuria (mg/mmol). (KDIGO clinical guidelines, 2012)

3.3.2 Impact factors:

The most common causes of kidney disease are diabetes and high blood pressure

- **Impact of Heart Disease and Stroke**: Having kidney sickness will increase the chances of also having heart sickness and stroke.
- **Impact of diabetes:** CKD attributable to diabetes referred to as diabetic kidney disease, is described via reduced kidney characteristic or the presence of kidney injury for at least three months, regardless of kidney function (Chronic Kidney Disease Basics | Chronic Kidney Disease Initiative | CDC, n.d.).
- **Impact of high blood pressure:** High blood strain is a leading cause of CKD. Over time, high blood strain can harm blood vessels in the course of your body. This can reduce the blood provide to necessary organs like the kidneys. High blood pressure additionally damages the tiny filtering units in your kidneys. As a result, the kidneys may additionally end removing wastes and extra fluid from your blood (National Kidney Foundation, n.d).
- **Impact of Obesity:** Obesity results in complex metabolic abnormalities which have wide-ranging effects on diseases affecting the kidneys. Some of the harmful renal consequences

of obesity may be mediated by downstream comorbid conditions such as diabetes mellitus or hypertension. Still, there are also effects of adiposity which could impact the kidneys directly Being obese doubles your risk of developing CKD compared to someone who has a healthy body weight, while overweight people increase their risk of developing CKD by 1.5 times (Cva K., 2017).

4. Proposed Methodology

In this paper, we used a data-mining classification technique. Data mining is the concept of extracting knowledge from large amounts of data. Classification is the process of finding a model that describes and distinguishes data classes or concepts, to be able to use the model to predict the label of a data record or to represent a descriptive analysis of data record for taking effective decisions. The classification mod-el consists of two stages: In stage 1, the training stage, the model is trained by a set of records, whose class labels are already known. In stage 2, the testing stage, the model goes to predict class labels of a collection of files, whose class labels are unknown, also called as test records. There are different classifiers, but for experimental analysis, we used classifiers supported by Orange3 tool. Orange3 supports various machine-learning (ML) algorithms. As we have compared our experimental results with Naïve Bayes, k-NN, tree, SVM, random forest, neural network and many other classifiers, the rest of this section discusses these ML algorithms. Experimental analysis is done on three different datasets. We used 70% records to train the model, and 30% records to test the model.

Data Preprocessing: Nowadays real-world datasets are highly susceptible to noisy missing or inconsistent data due to their usually vast size (often several gigabytes or more) and probable sourcing from various, heterogeneous sources. Low-quality data would result in low-quality mining outcomes. There are a variety of pre-processing techniques. Data cleaning is one of these techniques, it may be used to eliminate noise and fix data inconsistencies. Data integration as well, merges data from different sources into a cohesive data store, such as a data warehouse. Data transformations, such as standardization, can be applied. Data reduction can minimize data size by aggregating, deleting redundant functions, or clustering. These methods are not separated from each other, they may work together. (Han, et al, 2012). We will do the needed cleaning before applying algorithms on datasets.

4.1 Naïve Bayes classifier

Bayesian classifiers are statistical classifiers. We can estimate the probability of class membership, such as the likelihood that a given tuple belongs to a specific class (Han et al. ,2012).

Naive Bayes is among the simplest probabilistic classifiers it is based on Bayes theorem. It constructs a classification model by learning the conditional probabilities of each input attribute (Nimmala et al. ,2018). The same model is used to predict the class membership of input instance using the following equation:

$$P\,(x|y) = \frac{P\,(y|x).P(y)}{P\,(y)} \tag{2}$$

where P(x|y) is defined as the probability of observing x, given that y occurs. P(x|y) is called posterior probability P(y|x), P(x), and P(y) are called prior probabilities.

4.2 K-NN classifier

K-nearest neighbor (k-NN) it's a data mining technique. It tries to classify an unknown sample based on the known classification of its neighbors. Let us say a set of samples with known classification is available. Each sample should be classified similarly to its surrounding samples. If the classification of a sample is unknown, then it could be predicted by considering the classification of its nearest neighbor. (Mascherano et al., 2009).It has been widely used in the area of pattern recognition. Nearest-neighbor classifiers are based on learning by analogy, that is, by comparing a given test tuple with training tuples that are similar to it. N attributes describe the training tuples. Each tuple represents a point in an n-dimensional space. In this way, all of the training tuples are stored in n-dimensional pattern space. When given an unknown tuple, a k-nearest-neighbor classifier searches the pattern space for the k training tuples that are closest to the unknown tuple. These k training tuples are the k “nearest neighbors” of the unknown tuple (Pearson., 2019).

4.3 Support Vector Machine classifier

Support vector machines is based on supervised learning algorithm which can be used for both classification and regression problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification to maximize the separation between data points and the hyper-plan that best separates the features into different domains. The points closest to the hyperplane are called as the support vector points and the distance of the vectors from the hyperplane are called the margins (Yadav A, 2018).

The main idea of SVM is to find the optimal hyperplane between data of two classes in the training data [2]. SVM finds the hyperplane by solving optimization problem:

$$maxQ(a) = \sum_{i-1}^{n} a_i - \frac{1}{2} \sum_{i=1}^{n} \quad \sum_{j=1}^{n} a_i a_j d_i d_i x_i^T x_i \tag{3}$$

Where 0 ≤ai ≤ C for i=1, 2, … , n

The problem is how to construct a decision boundary that correctly classifies an input pattern that is not necessarily in the training set (Teng S. , et al. , 2010).

4.4 Artificial neural network algorithm

An artificial neural network (ANN) is a computational system, where information is processed collectively, in parallel throughout a network of nodes (neuron). In ANN the individual elements of the network, the neurons (nodes), read an input, process it, and generate an output. To create an ANN is necessary to put together a number of neurons. They are arranged on layers. A network has to have an input layer (which carries the values of outside variables) and an output layer (the predictions or the result). Weighting Factors: Neurons usually receive more than one input at the same time. Each input has its own relative weight which gives the input the impact that it needs on the processing element's summation function. Weights are adaptive coefficients within the network that determine the intensity of the input signal as registered by the artificial neuron. They are a measure of an input's connection strength. These strengths can be modified in response to various training sets and according to a network’s specific topology or through its learning rules. Activation Function: ANN use various functions other than activation functions, and most of them use sigmoid functions, which are also called logistic functions. (Su-Hyun Han. at al, 2018)The sigmoid function has the advantage that it is very simple to calculate compared to other functions. The sigmoid function is expressed as the following equation:

$$\boldsymbol{\alpha}(x)= 1/(1+e\^(-x)) \tag{4}$$

5. Requirements Analysis

We have two approaches to analyze requirements and they are Structured Analysis and Object-Oriented Analysis. We decided to use a structured approach to do the requirement analysis phase. Because we are focusing on the process and data more than objects. It which requires use case, Data Flow Diagram (DFD) and Entity Relationship diagram (ERD).

5.1 Use Case Diagram

As graphic diagrams describing and displaying relationships between users and actors (usually users and external systems) (Klimek & Szwed,2010); (Back et al. ,1999).

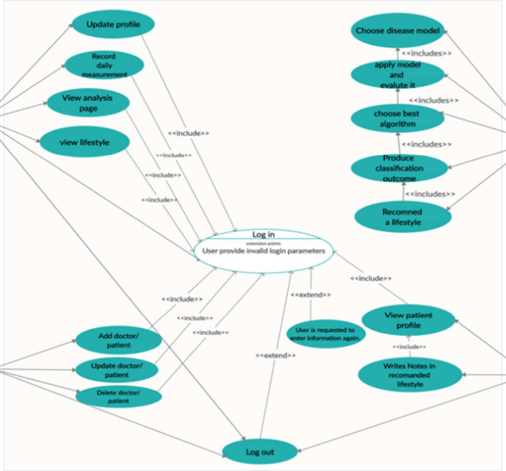


Fig 2. Use case diagram

5.2 Data Flow Diagram (DFD)

Process design can be created using a data flow diagram (DFD). DFD is used in a visual view to describe system requirements. Data flow diagram is made up of: Processes, Data Flows, Data Stores and External Entities. Process is an activity or function performed for a particular business purpose, data flow is either a single piece of data or a logical set of several pieces of data, datastore is a collection of data that is stored in a certain way, external entity is an internal individual, organization or system that is external to the system, but the system interacts with it.(Ibrahim & Yen ,2010); (Liu & Tang, 1991).

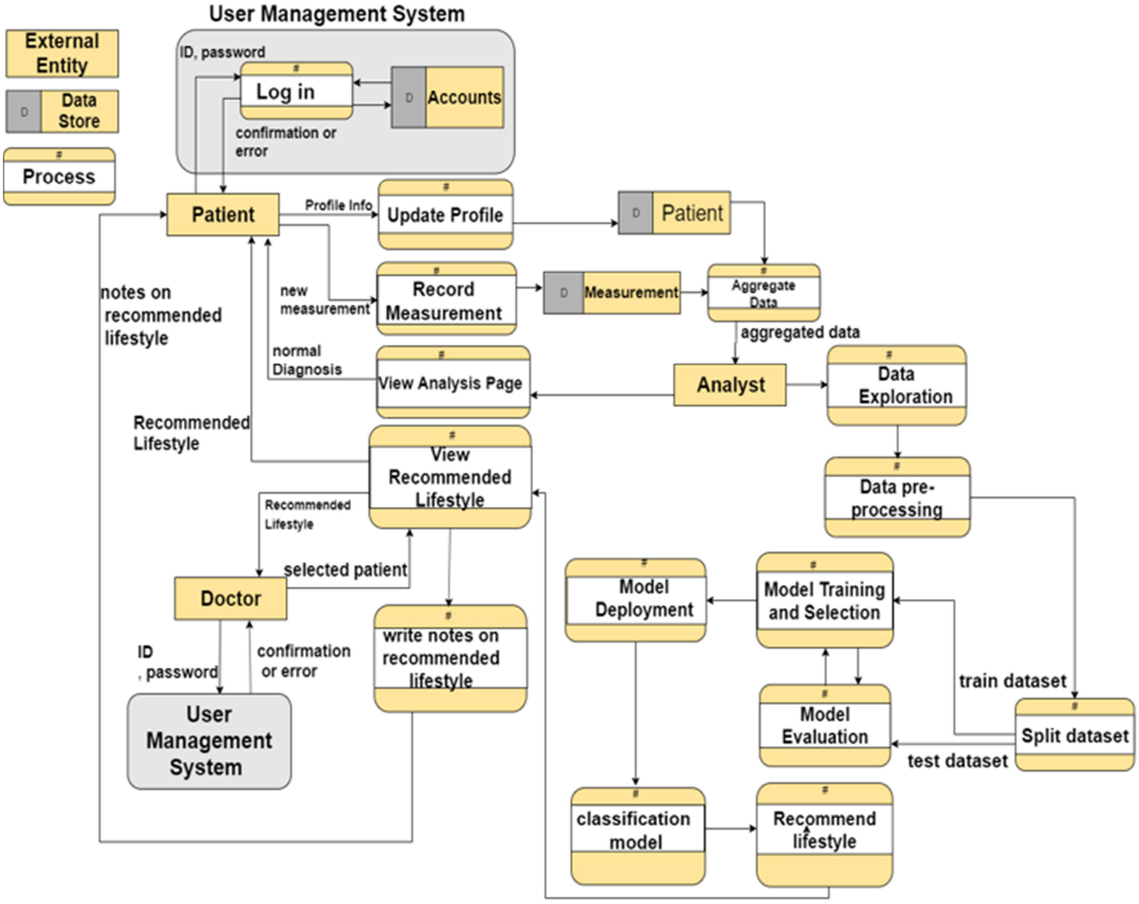


Fig 3. Data flow diagram

5.3 Entity Relationship diagram (ERD)

The Entity-Relationship diagram has been widely used in structured analysis and conceptual modeling. The ER approach is easy to understand, powerful to model real-world problems and readily translated into a database schema. The typical semantic constructs of the ER model and its variations we consider the following features:

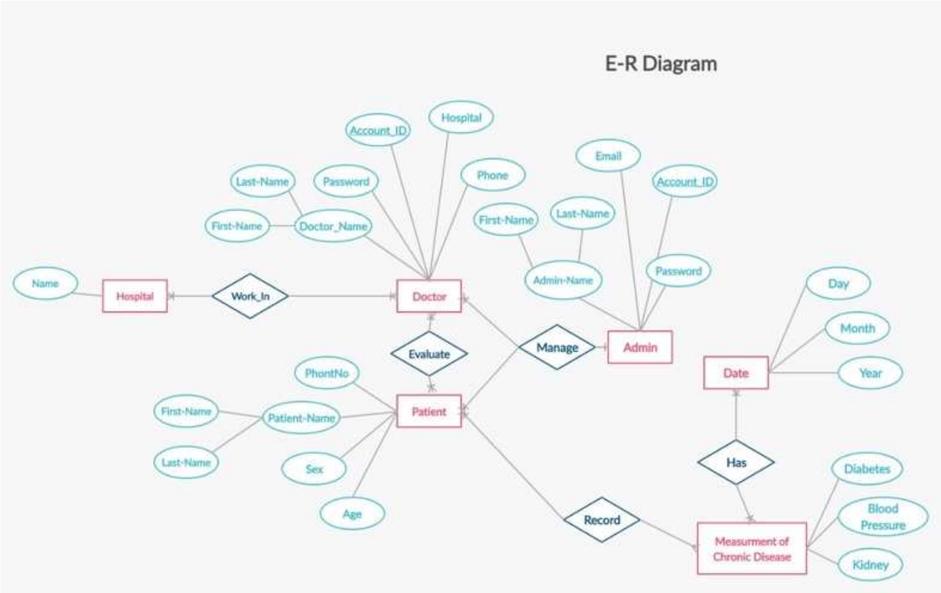


Fig 4. Entity Relationship Diagram

6. System Design

6.1 System Architecture

An architecture diagram is a graphical representation to illustrate, explain, and communicate thoughts about overall the system structure and the user requirements that the system must include using a bunch of icons and lines. Also, show formal structure, behavior of a system and fundamental organization of a system, their relationships to each other and represent database or memory representations with components that give a useful, implementable meaning.

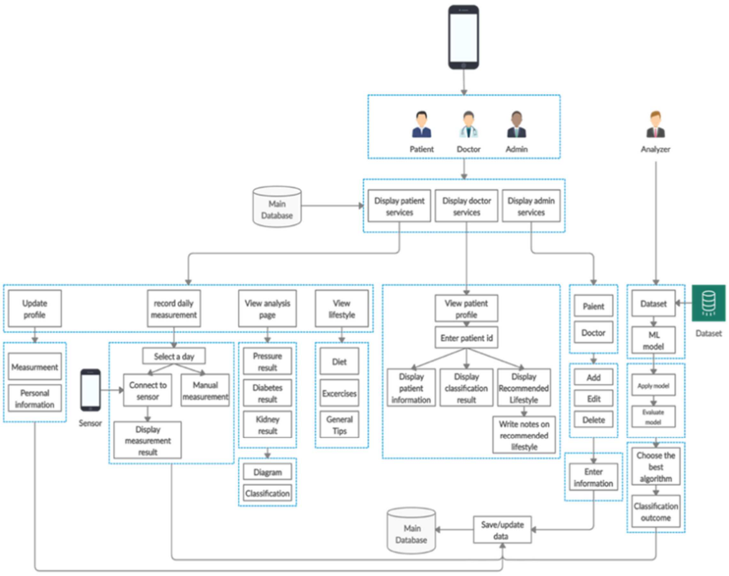


Fig 5. System architecture

6.2 System Flowchart:

We did 2 of them to describe the system. Flowchart A: describes the flow of the whole system Flowchart B: this one describes the machine learning part of the system.

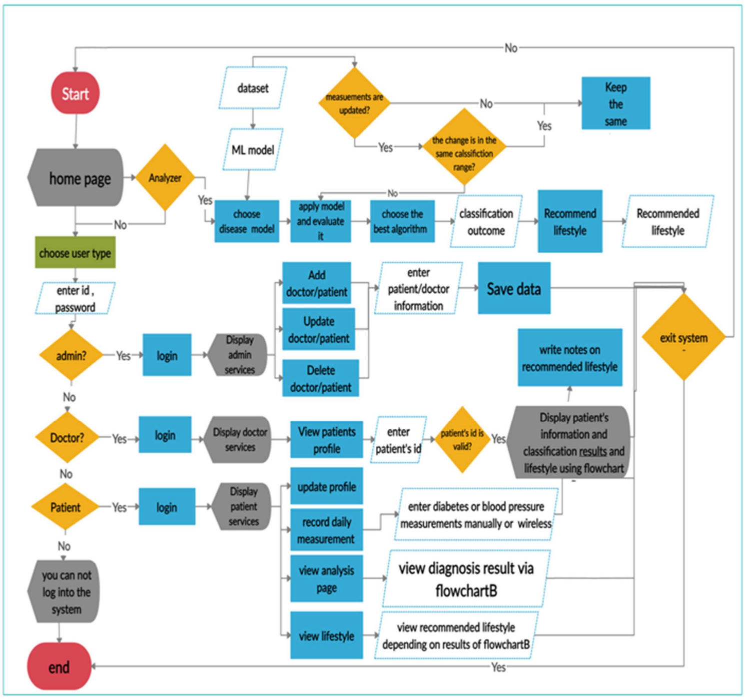


Fig.6. System flowchart A

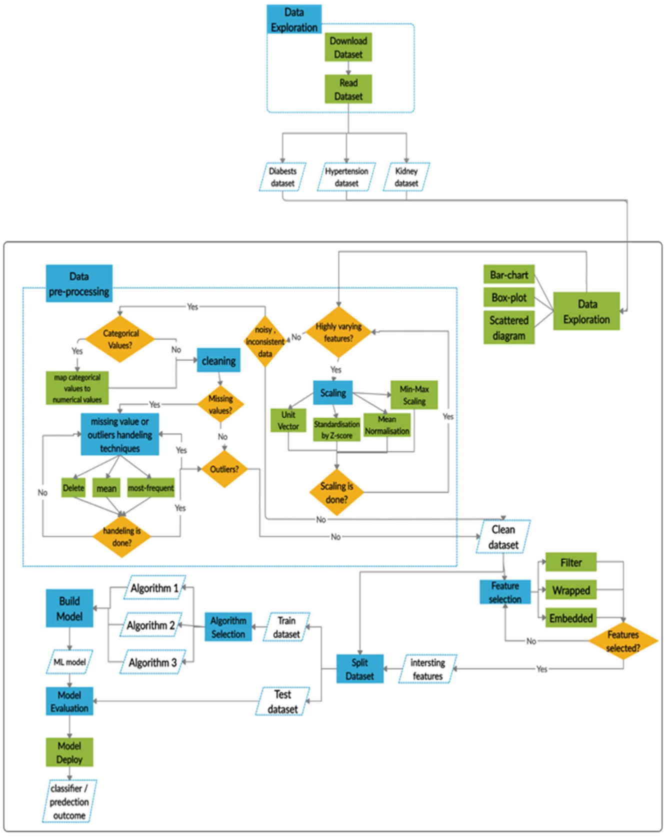


Fig.7. System flowchart B

7. Experimental results and analysis

We used Orange3 from Anaconda-Navigator; it is a data mining tool used to test many machine learning algorithms. Such as Support Vector Machine (SVM) and Artificial Neural Network (ANN) algorithms to test these algorithms on the three different datasets.

7.1 Hypertension Experimental results and analysis

We found a paper (Golino et al, 2014) that used two separated datasets .One for women and the other one for man. women's dataset can from downloaded at http://dx.doi.org/10.6084/m9.figshare.845664; men's dataset can be downloaded from http://dx.doi.org/10.6084/m9.figshare.845665. Dataset preprocessing involves a series of various steps, where each step applies a certain transformation that helps machine learning construct a better predictive model. We did categorical values handling and missing values .then we merged them into one dataset.It has 400 records. Each patient has only one record containing id, gender, age ,hc, wc,whr, sbp, dbp, BMI and hypertension.

Orange3 results for hypertension dataset:

Table 2. Hypertension results.

Model	AUC	CA	F1	Precision	Recall
kNN	0.858	0.768	0.758	0.756	0.768
Tree	1.000	1.000	1.000	1.000	1.000
SVM	0.976	0.893	0.888	0.895	0.892
Random Forest	0.997	0.983	0.983	0.983	0.983
Neural Network	0.990	0.927	0.927	0.928	0.927
Naive Bayes	0.992	0.786	0.793	0.828	0.786

We tested 6 algorithms KNN, Tree, SVM, Random forest, Neural network , and Naïve Bayes. We found the best accuracy 100% which is tree algorithm.

7.2 Diabetes Experimental results and analysis

We have used a dataset with 786 records downloaded from https://www.kaggle.com/uciml/pima-indians-diabetes-database. Each patient has only one record containing pregnancies, glucose, blood pressure, skin thickness, insulin, diabetes pe-degree function, age and BMI.

Table 3. Diabetes results.

Model	AUC	CA	F1	Precision	Recall
kNN	0.756	0.727	0.723	0.721	0.727
Tree	0.656	0.708	0.706	0.705	0.708
SVM	0.544	0.378	0.277	0.553	0.378
Random Forest	0.806	0.752	0.747	0.746	0.752
Neural Network	0.834	0.763	0.759	0.758	0.763
Naive Bayes	0.825	0.758	0.761	0.767	0.758

We tested 6 algorithms KNN, Tree, SVM, Random forest, Neural network , and Naïve Bayes. We found the best accuracy 76.3% which is neural networks.

7.3 Kidney Experimental results and analysis:

We have used a Kidney dataset from UCI repository https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease with 400 records. Each patient has only one record containing 24 features and one target which is CKD or non-CKD, it deter-mined if the patient have a kidney disease or not.

Applying algorithms using Orange3 tool:

Table 4. Kidney results.

Model	AUC	CA	F1	Precision	Recall
kNN	0.969	0.942	0.943	0.949	0.942
Tree	0.996	0.983	0.983	0.983	0.983
SVM	0.999	0.992	0.992	0.992	0.992
Random Forest	1.000	0.996	0.996	0.996	0.996
Neural Network	0.999	0.985	0.985	0.985	0.985
Naive Bayes	0.999	0.975	0.975	0.976	0.975

We tested 6 algorithms KNN, Tree, SVM, Random forest, Neural network , and Naïve Bayes. We found the best accuracy 100% which is random forest.

8. Conclusion

While writing this paper, we have used three different datasets for blood pressure id, gender, age ,hc, wc,whr, sbp, dbp, BMI and hypertension of a person were used to predict whether a person is prone to HBP or not. We used different classifiers to predict whether a person susceptible to have hypertension or not. Among all classification algorithms used for experimental Tree algorithm has shown the best accuracy when comparing with other algorithms.

For diabetes , we have used pregnancies, glucose, blood pressure, skin thickness, insulin, diabetes pedigree function, age and BMI of a person to predict whether a person is prone to diabetes or not. We used different classifiers to predict whether a person susceptible to have diabetes or not. Among all classification algorithms used for experimental analysis the neural network algorithm has shown the best accuracy when comparing with other algorithms.

For kidney diseases, the predictive models were presented by using machine learning methods including support vector machine (SVM) and artificial neural network (ANN) classifiers to predict chronic kidney disease. From the experimental results, it can be seen that random forest classifier gives the highest accuracy. Therefore, it can be concluded that random forest classifier is appropriated for predicting the chronic kidney disease in our case. In future; we would like to consider other attributes such as anger, anxiety, and cholesterol level to improve the prediction performance of the classifiers.

References

[1] Chaurasia, V.,Pal, S., Tiwari,B. (2018). Chronic Kidney Disease: A Predictive model using Decision Tree.

[2] Nimmala, S., Ramadevi, Y., Sahith, R., & Cheruku, R. (2018).High blood pressure prediction based on AAA++ using machine learning algorithms.

[3] Golino, H. F., Amaral, L. S. D. B., Duarte, S. F. P., Gomes, C. M. A., Soares, T. D. J., Reis, L. A. D., & Santos, J. (2014). Predicting Increased Blood Pressure Using Machine Learning. Journal of Obesity, 2014, 1–12.

[4] Noia T.D, Ostuni V., Pesce F.,Binetti G., Naso D., Schena F., Sciascio E. D., (2013), "An end stage kidney disease predictor based on an artificial neural networks ensemble", Expert Systems with Applications, Retrieved from https://www.sciencedirect.com/science/article/pii/S0957417413000778

[5] Kerr M., Bray B., Medcalf J., Matthews B., (2012), "Estimating the financial cost of chronic kidney disease to the NHS in England", Nephrology Dialysis Transplantation, 27(3)

[6] Alassaf R, Alsulaim K., Alroomi N., Alsharif N., Aljubeir M., Olatunji S., Alahmadi A., Imran M., Alzahrani R., Alturayeif N., (2018), "Preemptive Diagnosis of Chronic Kidney Disease Using Machine Learning Techniques", International Conference on Innovations in Information Technology (IIT), Retrieved from https://ieeexplore-ieee-org.sdl.idm.oclc.org/document/8606040/authors#authors

[7] Vijayarani D.S., Dhayanand M.S. , (2015), "Kidney Disease Prediction Using SVM and ANN Algorithms", International Journal of Computing and Business Research (IJCBR), 6(2)

[8] Yadav A, (2018), "Support vector machine", towards data sciences, Retrieved from https://towardsdatascience.com/support-vector-machines-svm-c9ef22815589

[9] Teng S. , Du H. , Wu N. , Zhang W. , Su J. , (2010), "A Cooperative Network Intrusion Detection Based on Fuzzy SVMs", Journal of Networks, 5(4), 475-483

[10] Moguerza J. M. , Javier M. ,Muñoz A. , (2006), "Support Vector Machines with Applications", Stat. Sci, 21(3), 322-336

[11] Patel S., (2017), "Chapter 2 : SVM (Support Vector Machine) — Theory", Machine Learning 101, Retrieved from https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72

[12] The Renal Association, (2019), "CKD stages", Retrieved from https://renal.org/information-resources/the-uk-eckd-guide/ckd-stages/

[13] Centers for Disease Control and Prevention, (2018), "Chronic Kidney Disease Basics", Retrieved from https://www.cdc.gov/kidneydisease/basics.html

[14] Rakel D., (2018), "Integrative Medicine 4thedition)", Philadelphia, PA: Elsevier

[15] Tests to Measure Kidney Function, Damage and Detect Abnormalities, (n.d.), National Kidney Foundation, Retrieved from https://www.kidney.org/atoz/content/kidneytests

[16] Fox R., Kim H., (2018), "Treatment of HCV in Persons with Renal Impairment", Hepatitis C Online, Retrieved from https://www.hepatitisc.uw.edu/go/key-populations-situations/treament-renal-impairment/core-concept/all

[17] An-na W., Yue Z., Yun-tao H., Yun-lu L., (2010), "A novel construction of SVM compound kernel function", 2010 International Conference on Logistics Systems and Intelligent Management (ICLSIM), Retrieved from https://ieeexplore-ieee-org.sdl.idm.oclc.org/document/5461210/authors#authors

[18] Practical Points for Use of Estimated GFR and Albuminuria (ACR) in Assessing CKD (2015),Guidelines and Audit Implementation Network

[19] Su-Hyun Han ,Ko Woon Kim, SangYun Kim &Young Chul Youn.(2018) .Artificial Neural Network: Understanding the Basic Concepts without Mathematics.

[20] Adriana Albu , Loredana Ungureanu .(n.d.)Artificial Neural Network in Medicine.

[21] Sheikh Salahuddin Ahmed, Md. Aminul Haque Khan, Tarafdar Runa Laila .(2013).Treatment and Prevention of Common Complications of Chronic Kidney Disease

[22] Chronic Kidney Disease Basics | Chronic Kidney Disease Initiative | CDC. (n.d.). Retrieved from https://www.cdc.gov/kidneydisease/basics.html

[23] of the World Kidney Day Steering Committ (2017) Obesity and kidney disease: hidden consequences of the epidemic

[24] Jawdeh, B. G. A., & Govil, A. (2017). Acute Kidney Injury in Transplant Setting: Differential Diagnosis and Impact on Health and Health Care. Advances in Chronic Kidney Disease.

[25] Kramarenko,O.M.(2015), http://economyofregion.com/current//2664/pdf/ , Economy of Region.

[26] Cva,K.(2017), https://www.medwinpublishers.com/JOBD/JOBD16000139.pdf , Journal of Orthopedics & Bone Disorders, 1(7). doi: 10.23880/jobd-16000139

[27] Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group, (2012), KDIGO Clinical practice guideline for the evaluation and management of chronic kidney disease.

[28] Banerjee A., Noor A., Siddiqua N., Uddin M., (2019), ”Food Recommendation using Machine Learning for Chronic Kidney Disease Patients”, 2019 International Conference on Computer Communication and Informatics (ICCCI), Retrieved from https://ieeexplore-ieee-org.sdl.idm.oclc.org/document/8821871/authors#authors

[29] Nutrition Division-Ministry of health, (2014), “Dietary Guidelines& Nutrition Therapy for Specific Diseases”, health.gov.lk

[30] Centers for Disease Control and Prevention. Chronic Kidney Disease Surveillance System website. https://nccd.cdc.gov/CKD. Accessed October 20, 2019.

[31] WebMD. (2019). What is a chronic disease? [online] Available at: https://www.webmd.com/cancer/qa/what-is-a-chronic-disease [Accessed 21 Oct. 2019].

[32] MedicineNet. (2019). Definition of Chronic disease. [online] Available at: https://www.medicinenet.com/script/main/art.asp?articlekey=33490 [Accessed 21 Oct. 2019].

[33] Who.int. (2019). WHO | 2. Background. [online] Available at: https://www.who.int/nutrition/topics/2_background/en/ [Accessed 21 Oct. 2019].

[34] Iyer, A., S, J. and Sumbaly, R. (2015). Diagnosis of Diabetes Using Classification Mining Techniques. International Journal of Data Mining & Knowledge Management Process, 5(1), pp.01-14.

[35] Jayalakshmi, T. and Santhakumaran, A. (2010). A Novel Classification Method for Diagnosis of Diabetes Mellitus Using Artificial Neural Networks. 2010 International Conference on Data Storage and Data Engineering.

[36] S. Mustafa, K. , Watan, I., G. and Enteesha, D., M. (2018). An Accurate Diabetes Prediction System Based on K-means Clustering and Proposed Classification Approach. 2018 International Journal of Applied Engineering Research, 5(6), pp. 4038-4041.

[37] Chen, L., Pei, J., Kuang, J., Chen, H., Chen, Z., Li, Z. and Yang, H. (2015). Effect of lifestyle intervention in patients with type 2 diabetes: A meta-analysis. Metabolism, 64(2), pp.338-347.

[38] Coughlin, S., Hatzigeorgiou, C., Anglin, J., Xie, D., Besenyi, G., Leo, G. and Wilkins, J. (2019). Healthy lifestyle intervention for adult clinic patients with type 2 diabetes mellitus. [online] Openaccessjournals.com. Available at: http://www.openaccessjournals.com/articles/healthy-lifestyle-intervention-for-adult-clinic-patients-with-type-2-diabetes-mellitus.html [Accessed 21 Oct. 2019].

[39] Solutions, C. (2019). Chronic Disease Management Software - ScienceSoft. [online] Scnsoft.com. Available at: https://www.scnsoft.com/healthcare/chronic-disease-management [Accessed 25 Oct. 2019].

[40] Shelly, G. B., & Rosenblatt, H. J. (2012). Systems analysis and design. Boston: Course Technology Cengage Learning.

[41] Dennis, A., Wixom, B. H., & Roth, R. M. (2013). System analysis and design. Hoboken, NJ: Wiley.

[42] DeMarco, T.,1978.Structured Analysis and Systems Specifications, Prentice Hall.

[43] Wieringa, R., 1998. A survey of structured and object-oriented software specification methods and techniques.

[44] Falessi, D., Cantone, G., & Grande, C. (2007). A Comparison of Structured Analysis and Object-Oriented Analysis - An Experimental Study. Proceedings of the Second International Conference on Software and Data Technologies.

[45] Klimek,R. & Szwed,P. (2010).Formal Analysis of Use Case Diagrams

[46] Back ,R. , Petre ,L. ,Porres,I.(1999).Analyzing UML Use Case as Contracts

[47] Liu, T., & Tang, C. S. (1991). Semantic specification and verification of data flow diagrams. Journal of Computer Science and Technology

[48] Ibrahim, R., & Yen, S. Y. (2010). Formalization of the Data Flow Diagram Rules for Consistency Check. International Journal of Software Engineering & Applications.