Machine Learning Implementations for Multi-class Cardiovascular Risk Prediction in Family Health Units

Mert Erkan Sozen
Business Development Chief,
Izmir Metro Company, Izmir, Turkey.
E-mail: merterkansozen@gmail.com

Gorkem Sariyer
Business Administration,
Yasar University, Izmir, Turkey.
Corresponding author: gorkem.ataman@yasar.edu.tr

Mustafa Yigit Sozen
Family Medicine Specialist,
Ayvalık No 2 Family Health Unit, Balıkesir, Turkey.
E-mail: yigitsozen@gmail.com

Gaurav Kumar Badhotiya
Operations and Decision Sciences,
Indian Institute of Management Ahmedabad (IIMA), Ahmedabad, Gujarat, India.
E-mail: gkb.choudhary@gmail.com

Lokesh Vijavargy
Jaipuria Institute of Management Jaipur, India.
E-mail: lokeshvijay79@gmail.com

Abstract
Cardiovascular disease (CVD) risk prediction plays a significant role in clinical research since it is the key to primary prevention. As family health units follow up on a specific group of patients, particularly in the middle-aged and elderly groups, CVD risk prediction has additional importance for them. In a retrospectively collected data set from a family health unit in Turkey in 2018, we evaluated the CVD risk levels of patients based on SCORE-Turkey. By identifying additional CVD risk factors for SCORE-Turkey and grouping the study patients into 3-classes "*low risk*," "*moderate risk,*" and "*high risk*" patients, we proposed a machine learning implemented early warning system for CVD risk prediction in family health units. Body mass index, diastolic blood pressures, serum glucose, creatinine, urea, uric acid levels, and HbA1c were significant additional CVD risk factors to SCORE-Turkey. All of the five implemented algorithms, k-nearest neighbour (KNN), random forest (RF), decision tree (DT), logistic regression (LR), and support vector machines (SVM), had high prediction performances for both the K4 and K5 partitioning protocols. With 89.7% and 92.1% accuracies for K4 and K5 protocols, KNN outperformed the other algorithms. For the five ML algorithms, while for the "*low risk*" category, precision and recall measures varied between 95% to 100%, "*moderate risk,*" and "*high risk*" categories, these measures varied between 60% to 92%. Machine learning-based algorithms can be used in CVD risk prediction by enhancing prediction performances and combining various risk factors having complex relationships.

Keywords- Cardiovascular diseases, Machine learning, Risk prediction, Family health units, SCORE-Turkey.

1. Introduction
Cardiovascular diseases (CVDs) are currently the primary cause of morbidity and mortality worldwide. In 2017, 17.79 million deaths took place globally from cardiovascular diseases (GBD, 2017), and 160,409

deaths were in Turkey (TSI, 2021). These numbers represented 31.77% of all global deaths (GBD, 2017) and 36.8% of deaths in Turkey (TSI, 2021). Since CVDs are the leading and most common causes of death, identifying the main factors increasing cardiovascular risks and predicting patients' risk levels has received growing attention in medicine and multidisciplinary research. There exist several determined systems for CVD risk assessment, such as the American College of Cardiology/American Heart Association (ACC/AHA) scales (Goff et al., 2014), the Framingham scales (D'Agostino et al., 2008), Reynolds (Ridker et al., 2007), QRISK (Hippisley-Cox et al., 2007), QRISK2 (Hippisley-Cox et al., 2008), and SCORE (Conroy et al., 2003). Hypertension, cholesterol, age, and smoking are the most common CVD risk factors identified by these scales.

Although these scales effectively reference clinicians, they have various weaknesses and limited prediction powers due to the increased complexity and randomness of CVD patient profiles (Damen et al., 2016; Obermeyer and Emanuel, 2016). Standard CVD risk assessment scales assume that identified risk factors are related to CVD risk outcomes, and thus they oversimplify complex nonlinear relations. Improved approaches that combine various risk factors with linear or nonlinear results are required. On the other hand, since researchers have been presenting new CVD risk factors (Ahmad et al., 2018; Lee et al., 2016; Parrinello et al., 2015), the continuous growth of these factors is another reason for a more precise CVD prediction model requirement.

As a leading emerging technology, machine learning (ML) presents a comprehensive approach to these established CVD risk assessment scales to handle current limitations. ML technologies have become popular in the various contexts of medical literature (Haug and Drazen, 2023; Hu et al., 2023; Sidey-Gibbons and Sidey-Gibbons, 2019). Many researchers and practitioners benefit from the power of these technologies in the COVID-19 context (Ibrahim et al., 2023; Kamalov et al., 2023; Sariyer et al., 2023a; Sariyer et al., 2023b; Solayman et al., 2023; Sozen et al., 2022). While implementation of these technologies in emergency departments' operations is widely studied in the literature (Ataman et al., 2023; Ataman and Sariyer, 2021; Sariyer and Ataman, 2021; Sariyer and Ocal Tasar, 2020), these technologies have also been well adopted in cardiology research (Liu et al., 2023; Avram, 2023; Huang et al., 2022; Jamthikar et al., 2022; Kanagarathinam et al., 2022; Cho et al., 2021; Dey et al., 2019).

As a subset of artificial intelligence, ML is characterized by the ability of software to learn from data and generate predictions without explicit prior programming. ML techniques and technologies depend on the computer to understand all complicated nonlinear relationships between variables and develop an efficient model to estimate the model target given the values of independent features.

There exist studies presenting the implementation of these ML technologies to predict the CVD risk levels of patients. While some of these studies used ACC/AHA guidelines as a baseline risk prediction model for evaluating CVD risk (Quesada et al., 2019; Kakadiaris et al., 2018; Weng et al., 2017), others used Framingham scores (Du et al., 2020; Narain et al., 2016). There exist other studies using QRISK (Cho et al., 2021; Georga et al., 2019), Reynolds (Han et al., 2020), and SCORE (Dimopoulos et al., 2018; Quesada et al., 2019) scales as baseline models. However, these studies defined CVD risk level as a binary outcome variable: low risk and high risk. Although an increasing number of categories in outcome variables makes it harder for ML algorithms to learn and thus may reduce the prediction performance, it is more valuable from the clinical viewpoint (Sariyer and Ataman, 2022; Ataman and Sariyer, 2021; Sariyer et al., 2019). A higher number of outcome variable categories in these models classifies patients into more homogenous and smaller groups, ensuring more detailed and applicable predictions.

On the other hand, existing studies applied ML technologies for CVD risk prediction to data sets received from second- and third-level health services, such as public and private hospitals. However, although first-level health services are designed to offer the most basic services to patients, this level is critical for the early diagnosis of patients at risk of different diseases. First-level health services have more significance in Turkey since the elderly population regularly visits family health units (Eser et al., 2019; Akman et al., 2017). Although first-level health services may fail to treat such risky patients when they are early diagnosed, responsible doctors may direct and transfer them to second- and third-level hospitals.

Therefore, we aimed to develop a CVD risk prediction model that may be used in first-level services. Thus, we believe this study contributes to the literature by 1) developing a multi-class CVD risk prediction model and 2) ensuring early diagnosis of CVD risk groups in first-level health services/family health units.

This study is organized as follows: section 2 describes the materials and methods by presenting the dataset and the methodology of the research. In section 3, we summarize the background information on the implemented ML algorithms. In section 4, we present the findings of the study. In section 5, we highlight the discussions and critical differences between the proposed model of this study and existing studies and demonstrate this study's contributions to the literature. Finally, we present the concluding remarks in section 6.

2. Materials and Methods

We present the study design of this paper in Figure 1. We first collected the data set and transformed it into a structured data set by accomplishing the required pre-processing tasks. Then, we applied statistical analysis to select input variables for the proposed CVD risk model for family health units. We validated the feature selection process by presenting the Pearson correlation coefficients between selected factors and CVD risk levels. We divided the data set into train and test data sets using the train-test split. Learning was done by different machine learning algorithms in the train data set. We evaluated and compared the prediction performances of implemented algorithms in test sets and decided on the best-performing algorithm. Finally, we presented the CVD risk prediction model for family health units (FHU) based on the selected input variables and the best-performing ML algorithm.
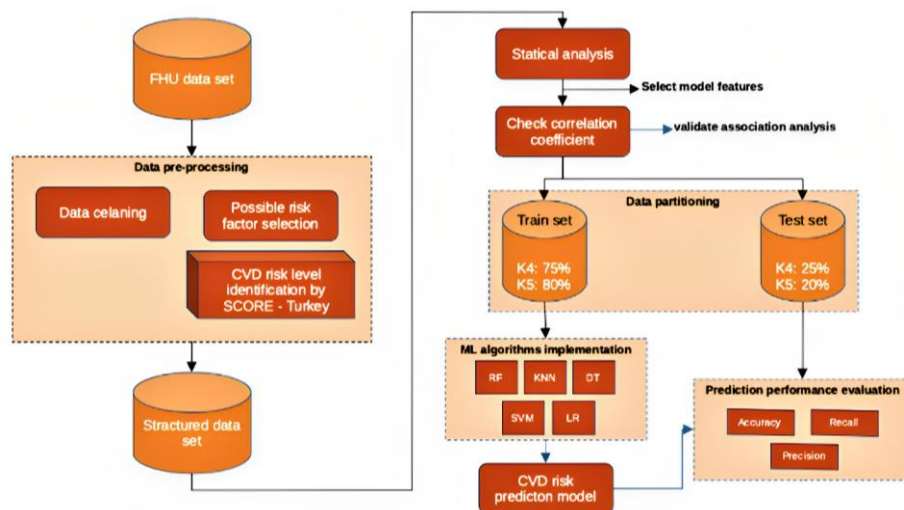

Figure 1. Flow chart of proposed ML-implemented system for CVD risk prediction in FHUs.

2.1 Data Source

We received the data for this study from the family health unit with number 1002009 in Balıkesir City, Turkey. Data on all patients arriving at this unit during 2018 were extracted from this health system's electronic databases. Since the middle-aged and elderly constitute the main age profile of health units and SCORE-Turkey evaluates the risk levels starting with age 40, we eliminated patients below 40.

2.2 Risk Factor Variables

We used SCORE-Turkey as a standard guideline for predicting current CVD risk levels of family health unit patients to catch early warnings and direct the risky patients to higher health services. Thus, we started forming our data set with five core variables (gender, age, smoking behavior, total cholesterol, and systolic blood pressure) defined in SCORE-Turkey. We defined additional variables which may affect CVD based on their inclusion in published CVD risk algorithms and within the literature on other potential CVD risk factors. However, our main limitation in this study was due to the variable definition. Since health units are restricted by some basic tests and treatments and cannot make detailed diagnostic tests, some of those previously documented risk factors were not available in our data set. Thus, based on the existence of these factors, we identified the risk factors of our model. The defined factors were further reviewed by practicing clinicians.

The final data set included 21 factors. Two of them (gender and smoking behavior) were binary categorical variables. The remaining factors (age, diastolic blood pressure, body mass index, total cholesterol, ALP, ALT, AST, serum iron, serum glucose, HbA1c, HDL cholesterol, serum Ca, serum creatinine, serum potassium, serum Na, serum urea, serum uric acid, and vitamin B12) were continuous variables.

2.3 Outcome

FHU's databases do not include detailed data on patients' histories, such as whether they have had any heart conditions before. Therefore, we could not define our outcome of interest based on the cardiovascular disease status of patients. Consistent with the study objective, we expressed the outcome variable as the potential CVD risk levels of patients and calculated it using the baseline SCORE-Turkey model. Initially, SCORE-Turkey had seven different classes based on risk calculations. However, since learning a model with such a large number of different categories is complicated, mainly when we are limited in sample size and have to define some risk factors, we decided to reduce the number of types appropriately. By interpreting these original categories with the help of clinicians, we decided to describe the outcome variable (potential CVD risk level) in three categories: "low risk", "moderate risk", and "high risk" patients. Table 1 presents the transformation of SCORE-Turkey risk groups and the defined outcome categories.

Table 1. CVD risk level 3-class definition based on SCORE-Turkey.

Class #	Class label	Used SCORE-Turkey classes
1	low-risk patients	0% & 1% & 2% & 3-4%
2	moderate-risk patients	5-9% & 10-14%
3	high-risk patients	15% and above

2.4 Statistical Analysis

We present the main properties of the patients for each of the outcome classes as mean ± SD (standard deviation) for continuous variables and number (%) for categorical variables. The significance of differences between low-risk, moderate-risk, and high-risk patient groups are analyzed by one-way

ANOVA for categorical variables and the Chi-square test for categorical variables. We used two-tailed tests and 95% confidence intervals (p-values ≤0.05 were noted as significant differences). We used IBM SPSS statistics version 22.0 (IBM, Armonk, NY, USA) in the statistical analysis.

To evaluate and compare the prediction powers of ML algorithms, we used accuracy, precision, and recall as the primary decision criteria. Accuracy showed the percentage of correctly classified instances compared to the total instances in the test data set. Precision and recall present detailed information on the prediction powers of the ML algorithms for each class of the outcome variable.

2.5 Feature Selection

In statistics and ML, feature selection is used to select the important factors or input variables of the proposed models. To efficiently integrate the model, explain model findings, avoid the curse of dimensionality, have computing advantages, decrease overfitting, and improve prediction accuracy, data scientists use feature selection tasks as a pre-stage in modelling. Various techniques for feature selection are used in the literature, such as principal component analysis, logistic regression, Fisher's discriminant ratio, mutual information, Pearson correlation, and ANOVA. We used ANOVA to select significant factors in our model.

2.6 Data Pre-processing

The data set included missing information in at least one too many input variables for some patients' records. In addition, there exist redundant entries. Both of them were excluded from the data set. Patients below 40 were excluded from the analysis to have more meaningful and applicable interpretations for family health units in Turkey.

For defining the outcome variable, a transformation process was accomplished. Binary variables were coded with 0's and 1's. Females were coded with 0's, and males were coded with 1's. For smoking behavior, "not smoking" ones were coded with 0's and smoking with 1's.

2.7 Data Partitioning

Data partitioning was integrated into ML models to divide the study data set into the train and test/validation data sets. Different partitioning protocols have been well documented in the literature. Based on our data set's sample size limitation, we used the K5 protocol to increase the learning ability of the implemented ML algorithms. In this protocol, a randomly chosen 80% of the data sets are used for learning (the training data set), and the remaining 20% are used for validation. We presented the results of the K4 protocol, in which 75% of the data set is used for training and 25% for validation. We repeated this protocol 20 times for different random partitions, calculated the prediction power in each trial's validation tests, and reported the average accuracy.

3. Used ML Algorithms

3.1 Decision Trees (DT)

The decision tree is one of the most commonly used and easily interpretable ML algorithms in the medical literature. This algorithm is tree-structured (Quinlan, 1986). This is a supervised learning technique. This algorithm produces a regression tree for a continuous-type outcome variable and a classification tree for a categorical outcome variable. For both regression and classification-type models, any measurement scale of the input variables is allowed. The resulting flowchart-like tree has inner nodes defining a test on the attribute. Each branch of this tree displays output from the test, and each leaf node has a class label. Frequently, a top-down approach is used in which the input variable space is attempted to be converted to the highest purity in the leaves.

3.2 Random Forest (RF)

This algorithm was proposed by Breiman (2001). This is a widely known ensemble technique combining multiple decision trees. These trees are generated based on the best split point found in a random subset of features rather than all (Ayyoubzadeh et al., 2020). The main parameters of this algorithm are the number of features in the random subset and the number of trees in the forest (Krauss et al., 2017).

3.3 Support Vector Machines (SVM)

In the SVM algorithm, samples constituting the boundaries of the classes are obtained, and based on these samples, the most appropriate linear decision boundary is calculated to distinguish these classes. Boundary samples are labelled as support vectors. Linear classification methods attempt to partition data by creating a hypersurface. This algorithm, which is one of the linear classification methods, finds the best hypersurface separating the class samples with the maximum margin (Han et al., 2023). It is essential to convert nominal variables to numeric ones when running this algorithm.

3.4 Logistic Regression (LR)

LR is a supervised learning technique for a categorical outcome variable with input variables (discrete or continuous). This algorithm predicts the probability of outcome 2 being in any category based on the model inputs. It also measures the relationships between outcome and input variables by defining a logit function and estimating the probability of this function.

3.5 K Nearest Neighbour

K nearest neighbour classification is an algorithm in instance-based learning. The nearest neighbour method is to find a pre-determined number of training samples closest in distance to a new point and predict the label from these instances. The number of samples can be user-defined based on the local density of the point. The training sample data points are assigned weights according to their distances from the sample data point. This technique is straightforward to implement.

4. Results

4.1 Data Set Characteristics

After pre-processing, we formed the structured data set for this study. This set included a total of 501 patient records. While 351 (70.06%) of these patients were female, 150 (29.94%) were male. The mean and deviation statistics of the age variable for all these patients were 57.82 ± 10.50. While 468 (86.43%) of these patients were not smoking, 68 of them (13.57%) of them were smoking.

According to CVD risk level estimations, 224 (44.71%) of these patients were labelled as "low-risk" patients with 0% to 5% risk estimations based on SCORE-Turkey. While 182 (36.33%) of these patients were labelled as "moderate-risk" patients (5% to 15% estimates), the remaining 95 (18.96%) of them were labelled as "high-risk" patients with at least 15% risk estimations based on SCORE-Turkey.

We present the patient's demographic and clinical characteristics for each risk group definition separately in Table 2.

Table 2 showed that the average age, body mass index, systolic and diastolic blood pressures, total cholesterol, glucose level, HbA1C, ALT, serum creatinine, serum calcium, serum potassium, serum urea, serum uric acid, B12, and percentages of males and smokers were increasing with increasing CVD risk levels.

Table 2. CVD risk level 3-class definition based on SCORE-Turkey.

Risk factor variables	Low-risk (n=224)	Moderate-risk (n=182)	High-risk (n=95)	p-value
Age	47.38 (6.00)	65.63 (3.66)	67.47 (1.94)	<0.001
Body mass index	27.41 (10.38)	29.93 (24.05)	30.23 (20.39)	<0.001
Systolic blood pressure	123.85 (18.80)	142.39 (22.73)	152.76 (17.69)	<0.001
Diastolic blood pressure	77.35 (10.73)	80.99 (7.50)	83.14 (8.56)	<0.001
Total cholesterol	193.05 (36.50)	205.62 (37.51)	211.22 (39.50)	0.048
HDL Cholestrol	66.55 (17.82)	65.28 (19.06)	63.95 (17.72)	0.436
Glucose level	93.45 (16.67)	104.57 (25.15)	114.12 (25.37)	<0.001
HbA1C	5.31 (0.95)	5.63 (0.95)	6.23 (1.54)	<0.001
ALP	66.23 (21.74)	75.65 (27.58)	71.40 (19.14)	0.696
ALT	18.09 (11.92)	18.30 (15.18)	16.80 (10.22)	0.926
AST	16.65 (8.85)	18.50 (11.80)	16.46 (4.98)	0.402
Serum creatinine	0.88 (0.18)	1.03 (0.26)	1.08 (0.29)	<0.001
Serum iron	81.27 (35.73)	87.54 (40.30)	79.91 (27.91)	0.971
Serum calcium	8.96 (0.66)	9.14 (0.67)	9.22 (0.75)	0.145
Serum potassium	4.09 (0.47)	4.19 (0.44)	4.23 (0.47)	0.100
Serum sodium	139.60 (2.95)	140.24 (3.16)	139.76 (3.09)	0.261
Serum urea	26.99 (8.42)	35.20 (11.08)	38.27 (11.80)	<0.001
Serum uric acid	4.33 (1.45)	5.34 (1.67)	5.47 (1.44)	0.019
B12	382.82 (163.19)	401.05 (160.09)	409.58 (152.56)	0.128
Female* (%)	80.36	75.14	70.53	<0.001
Smoking* (%)	13.39	19.04	28.42	<0.001

The three risk level categories had significant differences based on age, body mass index, systolic and diastolic blood pressure, total cholesterol, glucose level, HbA1C, serum creatinine, serum urea, and serum uric acid. We also observed that gender and smoking behavior distributions had statistically significant differences in defined risk level categories. By excluding the remaining nine variables (HDL cholesterol, ALP, ALT, AST, serum iron, serum calcium, serum potassium, serum sodium, and B12), which have insignificant differences between the risk group categories, we selected 12 input variables for CVD risk modelling.
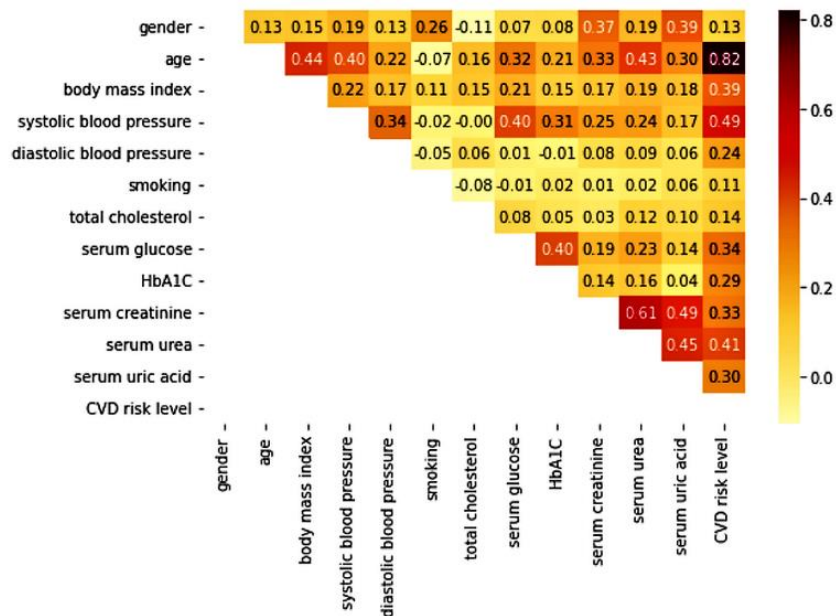

Figure 2. Pearson correlation coefficients between the study variables.

Thus, the structured data set included 12 input features (age, body mass index, systolic and diastolic blood pressures, total cholesterol, glucose level, HbA1C, serum creatinine, serum urea, serum uric acid, gender, smoking behavior), and a 3-class CVD risk level outcome variable for 501 patients.

In Figure 2, we present the Pearson correlation coefficients between these 13 variables of our data set in a heat-map form.

Based on the Pearson correlation coefficients shown in Figure 2 between the input variables and the CVD risk level, we observed that age has the highest degree of correlation with the CVD risk level. Systolic blood pressure, serum urea, body mass index, serum glucose, serum creatinine, serum uric acid, and HbA1C were moderately correlated with the CVD risk. The remaining four features (diastolic blood pressure, total cholesterol, gender, and smoking behavior) have lower degrees of relation to the CVD risk level.

4.2 Performance Evaluation

In Figure 3, we present the accuracy levels (percentage of correctly classified instances) of five implemented ML algorithms for K4 and K5 data partitioning protocols.

Based on Figure 3, we observed that for the K4 partitioning protocol, the first four algorithms (KNN, RF, DT, LR) had accuracy levels ranging from 85% to 90%, and SVM had a slightly lower performance of around 82%. While a similar result holds for the K5 protocol since the learning abilities of the algorithms improved with the increased percentage of the train data set, the accuracy levels of all algorithms increased. For the K5 protocol, while SVM achieves an accuracy level of 85%, the DT and LR algorithms reached approximately 90%. KNN had the highest accuracy level (92%) and RF had a very similar performance (91%).
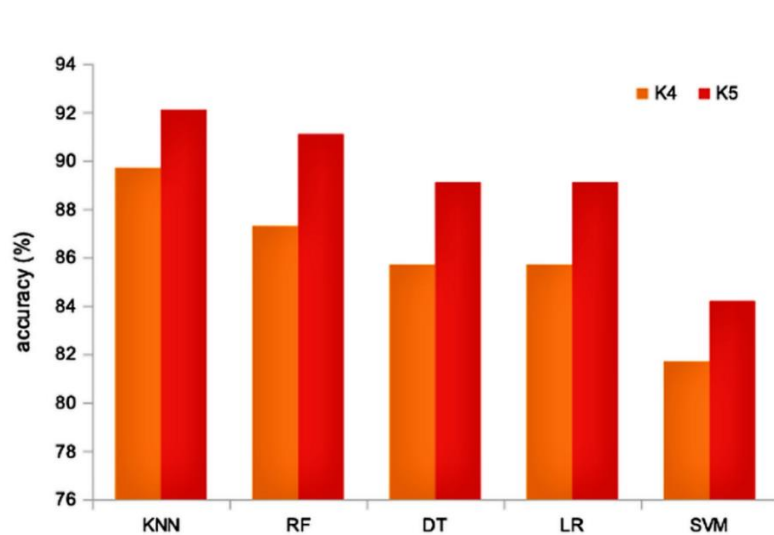

Figure 3. Prediction accuracies of ML algorithms for K4 and K5 protocols.

Figure 4 shows precision statistics of the implemented ML algorithms for each risk level category for K4 and K5 protocols. For the "low-risk" category, all algorithms had superior performances varying between 95% to 100% for both K4 and K5 protocols. Thus, among the patients predicted as "low-risk," almost all of them were actually "low-risk," and they were correctly predicted.
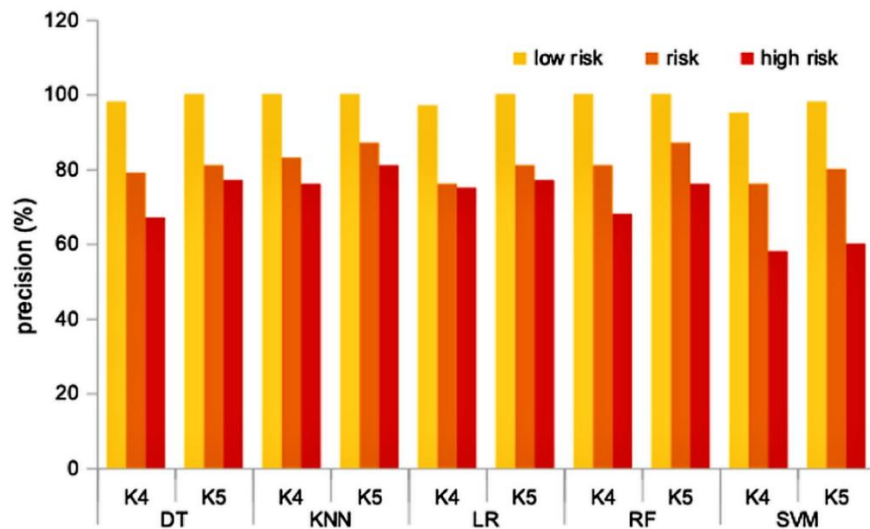
Figure 4. Precision levels of ML algorithms for K4 and K5 protocols.

The algorithms' "moderate-risk" category prediction performances varied between 76% to 83% for the K4 protocol and 80% to 87% for the K5 protocol. Finally, precision levels for "high-risk" varied between 58 to 76 for the K4 protocol and between 60% to 81% for the K5 protocol. The SVM algorithm obtained the lowest precision levels, and the highest ones were observed in KNN. By the implementation of KNN, among the patients predicted as "moderate-risk", 83% and 87% of them were also actually labelled as "moderate risk" for the K4 and K5 protocols, respectively. Similarly, within the "high-risk" predictions, while 76% of them were actually "high-risk" in the K4 protocol, 81% of them were actually "high-risk" in the K5 protocol.

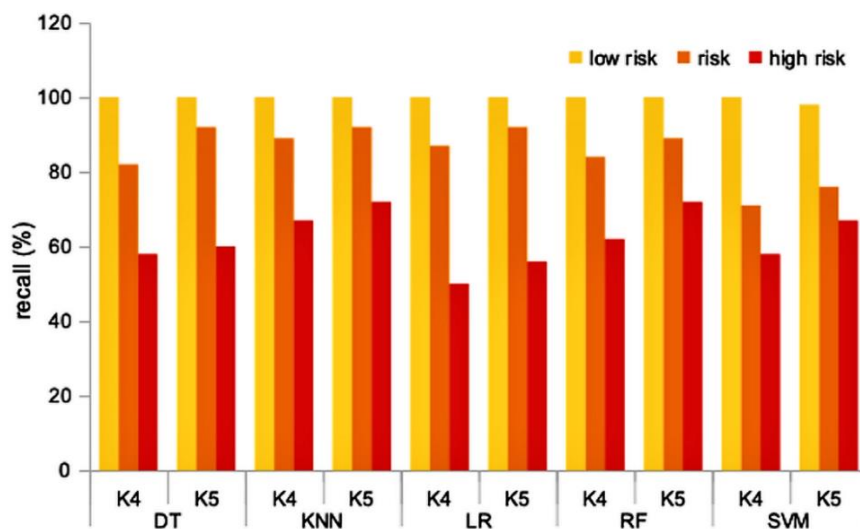For each outcome category, we also present recall statistics in Figure 5.

Figure 5. Recall levels of ML algorithms for K4 and K5 protocols.

We observed that for the patients who were actually "low-risk", all algorithms had perfect performances in classifying them as "low-risk" patients (except SVM for the K5 protocol). Except for the SVM algorithm, all the others had high performances while correctly predicting the patients at the "moderate-risk" level. However, the algorithms performed relatively poorly and incorrectly predicted the "high-risk" patients. The best-performing algorithm, KNN, correctly predicted 67% of patients for the K4 protocol and 72% for the K5 protocol.

We showed detailed statistics on precision and recall measures for each combination of actual and predicted CVD risk levels as a final analysis. Although all incorrect classifications are costly in prediction models, some errors can have higher costs. In our model, the worst situations are predicted as "low risk," where the actual risk levels are moderate risk or high risk." If a patient is predicted to be "low risk", the responsible doctor in family health units will not direct him to a higher level of health services. So, these cases can be deemed to have the worst or highest cost in our analysis.

On the other hand, if a patient is actually "high-risk" but the model predicts it as "moderate-risk", the cost of this error is lower in our model, as patients predicted as "moderate-risk" and "high-risk" will be referred to higher-level services for detailed diagnosis. Based on this logic, we showed the detailed statistics (percentage of incorrectly classified instances for each actual and predicted value combination) on precision and recall measures in Tables 3 and 4.

Table 3. Detailed statistics (in %) on precision measure.

Cost of error	Actual	Predicted		ML algorithms				
			protocol	KNN	RF	DT	LR	SVM
Lower costs of error	low-risk	moderate-risk	K4	0	0	0	0	0
			K5	0	0	0	0	2.86
	high-risk	moderate- risk	K4	16.67	19.15	21.28	23.53	23.81
			K5	12.82	13.16	19.05	19.05	17.14
	low-risk	high risk	K4	0	0	0	0	0
			K5	0	0	0	0	0
	moderate-risk	high-risk	K4	32.81	31.82	33.33	25	41.67
			K5	18.75	23.53	23.08	23.08	40
Higher costs of error	moderate-risk	low-risk	K4	0	0	1.72	3.39	5
			K5	0	0	0	0	2.17
	high-risk	low-risk	K4	0	0	0	0	0
			K5	0	0	0	0	0

Table 4. Detailed statistics (in %) on recall measure.

Cost of error	Actual	Predicted		ML algorithms				
			protocol	KNN	RF	DT	LR	SVM
Lower costs of error	low-risk	moderate-risk	K4	0	0	0	0	0
			K5	0	0	0	0	2.18
	low-risk	high-risk	K4	0	0	0	0	0
			K5	0	0	0	0	0
	moderate-risk	high-risk	K4	11.11	15.56	15.55	8.89	22.22
			K5	8.11	10.82	8.11	8.11	21.62
	high-risk	moderate-risk	K4	33.33	37.5	41.67	50	41.67
			K5	27.77	27.78	44.44	44.44	33.33
Higher costs of error	moderate-risk	low-risk	K4	0	0	2.22	4.44	6.67
			K5	0	0	0	0	2.7
	high-risk	low-risk	K4	0	0	0	0	0
			K5	0	0	0	0	0

Based on Tables 3 and 4, we mainly observed that for the two combinations having the higher error cost, incorrect prediction percentages were 0 for KNN and RF and were precisely or very close to 0 for the other ML algorithms' K4 and K5 protocols. Thus, although average recall and precision measures were relatively higher in some of the ML algorithms, we concluded that, for the higher costs of error situations, our ML-implemented system had almost perfect prediction performances.

5. Discussion

By taking the established SCORE-Turkey protocol as a baseline model for CVD risk prediction for family health units, we found that machine learning algorithms had high performance for classifying patients based on their CVD risk levels. Besides, these ML technologies have the power to integrate many factors with complex nonlinear relations. Thus, unlike established protocols, models integrating ML algorithms were not limited to a small set of risk factors and incorporated more pre-existing medical conditions. In addition to the five risk factors (gender, age, smoking behavior, total cholesterol, and systolic blood pressure) defined in SCORE-Turkey, we identified additional seven factors (body mass index, serum glucose, HbA1c, serum creatinine, serum urea, serum uric acid, and diastolic blood pressure) that had significant differences in the defined CVD risk level groups. When all these factors were integrated into the proposed CVD risk prediction model for FHU, 85% to 92% accuracies were obtained for the K5 protocol in different ML algorithms.

Although all of these algorithms have high levels of accuracy, KNN outperforms the other four algorithms in this data set by having higher accuracy, precision, and recall measures. Besides, for the worse situation in which "moderate-risk" and "high-risk" patients were actually classified as "low-risk", the KNN model had perfect performance and did not have any incorrect classification.

5.1 Key Difference Between the Proposed ML-based System and Existing Studies

First-level health services follow up with a specific group of patients, particularly in the middle-aged and elderly groups, at regular intervals. Thus, family health units play a crucial role in health systems since receiving early signals can identify the CVD risk of patients and direct them to higher-level health services for early diagnosis and treatment, which may save their lives. Although family health units have an early warning function in health systems, existing studies for CVD risk evaluations have not given enough attention to these services. This is the first study to present a CVD risk prediction model specifically for family health units. To fulfil this gap in the literature, proposing an emerging ML technologies-based prediction system is the other strength of this study.

Additionally, existing research uses patients' electronic records received from higher-level health services (second- or third-level health services, such as different types of hospitals). These higher-level health services are responsible for the diagnosis and treatment of patients. These data sets do not include information about patients' CVD status, such as if a patient's history has included myocardial infarction, heart attack, stroke, angina pectoris, hyperlipidemia, or any related diseases in a defined period (as in the past ten years). Thus, existing researchers attempt to present binary classification models where the outcome of interest (CVD risk) is labelled as fatal or non-fatal, low risk or high risk, etc. Due to the structure of family health units, our data set does not include this information. However, in family health units, the CVD risk levels of patients are evaluated based on pre-established protocols categorizing patients into more than one risk level. Thus, the multi-class CVD risk prediction system proposed in this study is another essential contribution to the literature.

Existing studies presenting CVD risk prediction systems for second- and third-level health services use established protocols as a baseline model to show the superiority of implemented ML algorithms. Using

routine clinical data, Weng et al. (2017) showed that four applied ML algorithms outperformed the established ACC/AHA protocol by 1.7% to 3.6% in prediction accuracy. By comparing the prediction powers of the Framingham risk score and logistic regression, Kennedy et al. (2013) showed that with the same predictors, logistic regression had a 2% higher accuracy compared to the Framingham score, and further accuracies (increasing +7%) were obtained when additional parameters were used in the model. Using SCORE as a baseline model, Quesada et al. (2019) tested the performance of 15 ML algorithms. They observed that seven of those algorithms had better accuracies, achieving +7% s compared to the baseline protocol. Besides comparing the performances of the ML algorithms, Navarini et al. (2020) compared the implementations of different protocols such as SCORE, QRISK, etc. They concluded that all the compared protocols had poor prediction abilities except for SCORE, which was adequately performed. They also showed that ML algorithms improved prediction abilities. Dimopoulos et al. (2018) used SCORE as a baseline model and implemented KNN, decision tree, and random forest as ML algorithms. They showed that while accuracy, sensitivity, and specificity values of the SCORE baseline model were the highest since ML algorithm results were close to them, the authors concluded that the advantages of ML algorithms should be taken into consideration. Narain et al. (2016) proposed a neural network-based ML system and concluded that this system's prediction power was higher than the established Framingham protocol.

Similarly, Han et al. (2020) concluded that the ML-based prediction system had around 10% improvement compared to traditional Framingham scores. While showing that ML algorithms significantly surpassed standard risk scales, Du et al. (2020) also showed that nonlinear models had higher accuracy than linear models. There exist some other studies using these baseline protocols only for ML-based systems' input identification that do not present the prediction powers of these protocols (Kim et al., 2017). Compared to these existing studies, we used an established SCORE-Turkey protocol for risk factor identification or performance comparison and for identifying CVD risk levels in family health unit patients.

The sample size, or the number of patients included in the study, is the other point that distinguishes our analysis from the literature. By using second- or third-level health services data or population health survey results, a large body of literature was able to validate their model in medium- to large-scale data sets with 2,000 to 90,000 patients (Du et al., 2020; Quesada et al., 2019; Dimopoulos et al., 2018; Weng et al., 2017; Kim et al., 2017; Kennedy et al., 2013), which increased the learning abilities of their model. Although our sample size was smaller (501), compared to these studies, by achieving around 90% accuracy, our ML-based system reached a prediction power exceeding or at least competing with the systems presented in the literature. In addition, while literature studies achieved performances ranging between 75% and 95% for binary class (fatal/non-fatal or low-risk/high-risk) prediction models, our system achieved such accuracy levels for a multi-class CVD risk prediction system (low-risk, moderate-risk, and high-risk).

When different ML algorithms were compared, we observed that a large body of literature either just applied neural networks (Kakadiaris et al., 2018; Kim et al., 2017; Narain et al., 2016; Bandyopadhyay et al., 2015; Dybowski et al., 1996) or showed that neural networks were outperforming the other algorithms (Weng et al., 2017). While some other studies used linear models (Kennedy et al., 2013), some of the other studies showed that nonlinear models such as KNN and RF outperformed the others (Du et al., 2020). In this study, by applying DT, LR, SVM, KNN, and RF, we showed that KNN outperformed the others.

5.2 Limitations and Future Research Directions

The limitations of this study are that it includes a single institution setting, and the findings may not be generalizable to other settings. These findings may be validated by future multi-center studies applying the proposed early warning ML-based CVD risk assessment model. Another main limitation of this study is its sample size. Since the subject of this study is the FHUs and the numbers of patients are not as high in these services as in the second- or third-level health services, the sample size is limited. The definition of the study variables is the other limitation. Since FHUs are restricted by some basic tests and treatments, some of the previously documented CVD risk factors are not available in our data set.

For future research directions, neural and deep belief networks should be implemented. Explainable artificial intelligence techniques may also be adapted to present a detailed understanding of the impact of all model parameters. In addition, a multicenter study can be designed and the number of patients can be increased with data collected from more than one center for future studies. This will increase the generalizability and reliability of the study.

6. Conclusion

In this study, we proposed an ML-implemented early warning system for multi-class CVD risk prediction in family health units. This model mainly contributes to the literature in two folds; by developing a multi-class CVD risk prediction model and by ensuring early diagnosis of CVD risk groups in first-level health services, or FHUs.

By taking the established SCORE-Turkey protocol as a baseline model for CVD risk prediction for FHUs, we found that ML algorithms had high performance for classifying patients based on their CVD risk levels. The accuracy levels of the implemented ML algorithms varied between 80% to 89% according to the K4 protocol and 85% to 92% according to the K5 protocol. In addition to the five risk factors (gender, age, smoking behavior, total cholesterol, and systolic blood pressure) defined in SCORE-Turkey, we identified additional seven factors (body mass index, serum glucose, HbA1c, serum creatinine, serum urea, serum uric acid, and diastolic blood pressure) that had significant differences in the defined CVD risk level groups. Although all of these algorithms have high levels of accuracy, KNN outperforms the other four algorithms (DT, SVM, RF, and LR) in this data set by having higher accuracy, precision, and recall measures.

By integrating various risk factors with nonlinear relationships with CVD risk classes and obtaining high prediction accuracies, we concluded that ML-implemented systems are helpful in clinical practices. We labelled this system an "early warning system," since by detecting signals of CVD risks, responsible doctors in family health units will be able to direct risky patients to higher-level health services for early diagnosis and treatment.

The proposed ML-implemented early warning system for multi-class CVD risk prediction in family health units can be integrated as a decision support system (DSS) in this first-level health service. Therefore, by entering the values of the identified 12 input factors of the patients into this DSS, responsible doctors can evaluate the CVD risk levels of patients. When the obtained value of CVD risk is "moderate risk" or "high risk," the responsible doctor can direct and transfer these patients to higher-level health services for detailed evaluations. We also concluded that by having the additional characteristics of combining various risk factors with complex relationships and superior prediction performances, ML-based emerging technologies could be quickly adopted in clinical decision-making.

References

Ahmad, S., Moorthy, M.V., Demler, O.V., Hu, F.B., Ridker, P.M., Chasman, D.I., & Mora, S. (2018). Assessment of risk factors and biomarkers associated with risk of cardiovascular disease among women consuming a Mediterranean diet. *JAMA Network Open*, *1*(8), e185708. https://doi.org/10.1001/jamanetworkopen.2018.5708.

Akman, M., Sakarya, S., Sargın, M., Ünlüoğlu, İ., Eğici, M.T., Boerma, W.G., & Schäfer, W.L. (2017). Changes in primary care provision in Turkey: A comparison of 1993 and 2012. *Health Policy*, *121*(2), 197-206. https://doi.org/10.1016/j.healthpol.2016.11.016.

Ataman, M.G., & Sarıyer, G. (2021). Predicting waiting and treatment times in emergency departments using ordinal logistic regression models. *The American Journal of Emergency Medicine*, *46*, 45-50. https://doi.org/10.1016/j.ajem.2021.02.061.

Ataman, M.G., Sariyer, G., Saglam, C., Karagoz, A., & Unluer, E.E. (2023). Factors relating to decision delay in the emergency department: Effects of diagnostic tests and consultations. *Open Access Emergency Medicine*, *15*, 119-131.

Avram, R. (2023). Revolutionizing cardiovascular risk prediction in patients with chronic kidney disease: Machine learning and large-scale proteomic risk prediction model lead the way. *European Heart Journal*, *44*(23), 2111-2113. https://doi.org/10.1093/eurheartj/ehad127.

Ayyoubzadeh, S.M., Ghazisaeedi, M., Kalhori, S.R.M., Hassaniazad, M., Baniasadi, T., Maghooli, K., & Kahnouji, K. (2020). A study of factors related to patients' length of stay using data mining techniques in a general hospital in southern Iran. *Health Information Science and Systems*, *8*(1), 1-11. https://doi.org/10.1007/s13755-020-0099-8.

Bandyopadhyay, S., Wolfson, J., Vock, D.M., Vazquez-Benitez, G., Adomavicius, G., Elidrisi, M., Johnson, B.E., & O'Connor, P.J. (2015). Data mining for censored time-to-event data: A Bayesian network model for predicting cardiovascular risk from electronic health record data. *Data Mining and Knowledge Discovery*, *29*, 1033-1069.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5-32. https://doi.org/10.1023/A:1010933404324.

Cho, S.Y., Kim, S.H., Kang, S.H., Lee, K.J., Choi, D., Kang, S., Park, S.J., Kim, T, Yoon, C.H, Youn, T.J, & Chae, I.H. (2021). Pre-existing and machine learning-based models for cardiovascular risk prediction. *Scientific Reports*, *11*(1), 8886. https://doi.org/10.1038/s41598-021-88257-w.

Conroy, R.M., Pyörälä, K., Fitzgerald, A.P., Sans, S., Menotti, A., De Backer, G., De Bacquer, D, Ducimetière, P., Jousilahti, P., Keil, U., Njølstad, I., Oganov, R.G., Thomsen, T., Tunstall-Pedoe, H., Tverdal, A., Wedel, H., Whincup, P., Wilhelmsen, L., & Graham, I.M. (2003). Estimation of ten-year risk of fatal cardiovascular disease in Europe: The SCORE project. *European Heart Journal*, *24*(11), 987-1003.

D'Agostino Sr, R.B., Vasan, R.S., Pencina, M.J., Wolf, P.A., Cobain, M., Massaro, J.M., & Kannel, W.B. (2008). General cardiovascular risk profile for use in primary care: The Framingham heart study. *Circulation*, *117*(6), 743-753. https://doi.org/10.1161/circulationaha.107.699579.

Damen, J.A.A.G., Hooft, L., Schuit, E., Debray, T.P.A., Collins, G.S., Tzoulaki, I., Lassale, C.M., Siontis, G.C.M., Chiocchia, V., Roberts, C., Schlüsse, M.M., Gerry, S., Black, J.A., Heus, P., Schouw, Y.T., Peelen, L.M., & Moons, K.G.M. (2016). Prediction models for cardiovascular disease risk in the general population: Systematic review. The *BMJ*, *353*. https://doi.org/10.1136/bmj.i2416.

Dey, D., Slomka, P.J., Leeson, P., Comaniciu, D., Shrestha, S., Sengupta, P.P., & Marwick, T.H. (2019). Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. *Journal of the American College of Cardiology*, *73*(11), 1317-1335. https://doi.org/10.1016/j.jacc.2018.12.054.

Dimopoulos, A.C., Nikolaidou, M., Caballero, F.F., Engchuan, W., Sanchez-Niubo, A., Arndt, H., Ayuso-Mateos, J.L., Haro, J.M., Chatterji, S., Georgousopoulou, E.N., Pitsavos, C., & Panagiotakos, D.B. (2018). Machine learning methodologies versus cardiovascular risk scores, in predicting disease risk. *BMC Medical Research Methodology*, *18*, 1-11. https://doi.org/10.1186/s12874-018-0644-1.

Du, Z., Yang, Y., Zheng, J., Li, Q., Lin, D., Li, Y., Fan, J., Cheng, W., Chen, X.H., & Cai, Y. (2020). Accurate prediction of coronary heart disease for patients with hypertension from electronic health records with big data and machine-learning methods: Model development and performance evaluation. *JMIR Medical Informatics*, *8*(7), e17257. https://doi.org/10.2196/17257.

Dybowski, R., Gant, V., Weller, P., & Chang, R. (1996). Prediction of outcome in critically ill patients using artificial neural network synthesised by genetic algorithm. *The Lancet*, *347*(9009), 1146-1150. https://doi.org/10.1016/S0140-6736(96)90609-1.

Eser, E., Çevik, C., Baydur, H., Güneş, S., Esgin, T.A., Öztekin, Ç.S., Eker, E., Gumussoy, U., Eser, G.B., & Özyurt, B. (2019). Reliability and validity of the Turkish version of the WHO-5, in adults and older adults for its use in primary care settings. *Primary Health Care Research & Development*, *20*, e100. https://doi.org/10.1017/S1463423619000343.

GBD. (2017). Global burden of disease collaborative network, global burden of disease study 2017 (GBD 2017) results. Seattle, United States: Institute for health metrics and evaluation (IHME), http://ghdx.healthdata.org/gbd-results-tool received, 20.10.2021.

Georga, E.I., Tachos, N.S., Sakellarios, A.I., Kigka, V.I., Exarchos, T.P., Pelosi, G., Parodi, O., Michalis, L.K., & Fotiadis, D.I. (2019). Artificial intelligence and data mining methods for cardiovascular risk prediction. In: Spyretta, G., Konstantina, N. (eds) *Cardiovascular Computing—Methodologies and Clinical Applications* (pp. 279-301) Series in Bioengineering. Springer, Singapore. https://doi.org/10.1007/978-981-10-5092-3_14.

Goff, D.C., Lloyd-Jones, D.M., Bennett, G., Coady, S., D'agostino, R.B., Gibbons, R., Greenland, P., Lackland, D.T., Levy, D., Donnel, C.J.O., Robinson, J.G., Schwartz, S., Shero, D.T., Smith, S.C., Sorlie, P., Stone, N.J., & Wilson, P.W.F. (2014). 2013 ACC/AHA guideline on the assessment of cardiovascular risk: A report of the American college of cardiology/American heart association task force on practice guidelines. *Circulation*, *129*(25_suppl_2), S49-S73. https://doi.org/10.1161/01.cir.0000437741.48606.98.

Han, D., Kolli, K.K., Gransar, H., Lee, J.H., Choi, S.Y., Chun, E.J., Han, H.W., Park, S.H., Sung, J., Jung, H.O., Min, J.K., & Chang, H.J. (2020). Machine learning based risk prediction model for asymptomatic individuals who underwent coronary artery calcium score: Comparison with traditional risk prediction approaches. *Journal of Cardiovascular Computed Tomography*, *14*(2), 168-176. https://doi.org/10.1016/j.jcct.2019.09.005.

Han, J., Pei, J., & Tong, H. (2023). *Data mining: concepts and techniques*. Morgan Kaufmann. ISBN:978-0-12-811760-6.

Haug, C.J., & Drazen, J.M. (2023). Artificial intelligence and machine learning in clinical medicine, 2023. *New England Journal of Medicine*, *388*(13), 1201-1208. https://doi.org/10.1056/nejmra2302038.

Hippisley-Cox, J., Coupland, C., Vinogradova, Y., Robson, J., May, M., & Brindle, P. (2007). Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: Prospective open cohort study. *The BMJ*, *335*(7611), 136. https://doi.org/10.1136/bmj.39261.471806.55.

Hippisley-Cox, J., Coupland, C., Vinogradova, Y., Robson, J., Minhas, R., Sheikh, A., & Brindle, P. (2008). Predicting cardiovascular risk in England and Wales: Prospective derivation and validation of QRISK2. The *BMJ*, *336*(7659), 1475-1482. https://doi.org/10.1136/bmj.39609.449676.25.

Hu, Y., Cato, K.D., Chan, C.W., Dong, J., Gavin, N., Rossetti, S.C., & Chang, B.P. (2023). Use of real-time information to predict future arrivals in the emergency department. *Annals of Emergency Medicine, 81*(6), 728-737. https://doi.org/10.1016/j.annemergmed.2022.11.005.

Huang, W., Ying, T.W., Chin, W.L.C., Baskaran, L., Marcus, O.E.H., Yeo, K.K., & Kiong, N.S. (2022). Application of ensemble machine learning algorithms on lifestyle factors and wearables for cardiovascular risk prediction. *Scientific Reports*, *12*(1), 1033. https://doi.org/10.1038/s41598-021-04649-y.

Ibrahim, Z., Tulay, P., & Abdullahi, J. (2023). Multi-region machine learning-based novel ensemble approaches for predicting COVID-19 pandemic in Africa. *Environmental Science and Pollution Research*, *30*(2), 3621-3643.

Jamthikar, A., Gupta, D., Johri, A.M., Mantella, L.E., Saba, L., & Suri, J.S. (2022). A machine learning framework for risk prediction of multi-label cardiovascular events based on focused carotid plaque B-Mode ultrasound: A Canadian study. *Computers in Biology and Medicine*, *140*, 105102.

Kakadiaris, I.A., Vrigkas, M., Yen, A.A., Kuznetsova, T., Budoff, M., & Naghavi, M. (2018). Machine learning outperforms ACC/AHA CVD risk calculator in MESA. *Journal of the American Heart Association*, *7*(22), e009476.

Kamalov, F., Cherukuri, A.K., Sulieman, H., Thabtah, F., & Hossain, A. (2023). Machine learning applications for COVID-19: A state-of-the-art review. *Data Science for Genomics*, 277-289. https://doi.org/10.1016/B978-0-323-98352-5.00010-0. (In press).

Kanagarathinam, K., Sankaran, D., & Manikandan, R. (2022). Machine learning-based risk prediction model for cardiovascular disease using a hybrid dataset. *Data & Knowledge Engineering, 140*, 102042. https://doi.org/10.1016/j.datak.2022.102042.

Kennedy, E.H., Wiitala, W.L., Hayward, R.A., & Sussman, J.B. (2013). Improved cardiovascular risk prediction using nonparametric regression and electronic health record data. *Medical Care*, *51*(3), 251-258.

Kim, J., Kang, U., & Lee, Y. (2017). Statistics and deep belief network-based cardiovascular risk prediction. *Healthcare Informatics Research*, *23*(3), 169-175. https://doi.org/10.4258/hir.2017.23.3.169.

Krauss, C., Do, X.A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, *259*(2), 689-702.

Lee, J.J., Pedley, A., Hoffmann, U., Massaro, J.M., & Fox, C.S. (2016). Association of changes in abdominal fat quantity and quality with incident cardiovascular disease risk factors. *Journal of the American College of Cardiology*, *68*(14), 1509-1521. https://doi.org/10.1016/j.jacc.2016.06.067.

Liu, W., Laranjo, L., Klimis, H., Chiang, J., Yue, J., Marschner, S., Quiroz, J.C., Jorm, L., & Chow, C.K. (2023). Machine-learning versus traditional approaches for atherosclerotic cardiovascular risk prognostication in primary prevention cohorts: A systematic review and meta-analysis. *European Heart Journal-Quality of Care and Clinical Outcomes*, *9*(4), 310-322. https://doi.org/10.1093/ehjqcco/qcad017.

Narain, R., Saxena, S., & Goyal, A.K. (2016). Cardiovascular risk prediction: A comparative study of Framingham and quantum neural network based approach. *Patient Preference and Adherence*, *10,* 1259-1270.

Navarini, L., Caso, F., Costa, L., Currado, D., Stola, L., Perrotta, F., Delfino, l., Sperti, M., Deriu, M.A., Ruscitti, P., Pavlych, V., Corrado, A., Di Benedetto, G., Tasso, M., Ciccozzi, M., Laudisio, A., Lunardi, C., Cantatore, F.P., Lubrano, E., Giacomelli, R., Scarpa, R., & Afeltra, A. (2020). Cardiovascular risk prediction in ankylosing spondylitis: From traditional scores to machine learning assessment. *Rheumatology and Therapy, 7,* 867-882. https://doi.org/10.1007/s40744-020-00233-4.

Obermeyer, Z., & Emanuel, E.J. (2016). Predicting the future-big data, machine learning, and clinical medicine. *The New England Journal of Medicine*, *375*(13), 1216-1219. https://doi.org/10.1056/NEJMp1606181.

Parrinello, C.M., Lutsey, P.L., Ballantyne, C.M., Folsom, A.R., Pankow, J.S., & Selvin, E. (2015). Six-year change in high-sensitivity C-reactive protein and risk of diabetes, cardiovascular disease, and mortality. *American Heart Journal*, *170*(2), 380-389. https://doi.org/10.1016/j.ahj.2015.04.017.

Quesada, J.A., Lopez-Pineda, A., Gil-Guillén, V.F., Durazo-Arvizu, R., Orozco-Beltrán, D., López-Domenech, A., & Carratalá-Munuera, C. (2019). Machine learning to predict cardiovascular risk. *The International Journal of Clinical Practice*, *73*(10), e13389. https://doi.org/10.1111/ijcp.13389.

Quinlan, J.R. (1986). Induction of decision trees. *Machine Learning*, *1*, 81-106. https://doi.org/10.1007/BF00116251.

Ridker, P.M., Buring, J.E., Rifai, N., & Cook, N.R. (2007). Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: The Reynolds risk score. *JAMA*, *297*(6), 611-619. https://doi.org/10.1001/jama.297.6.611.

Sariyer, G., & Ataman, M.G. (2021). How machine learning facilitates decision making in emergency departments: Modelling diagnostic test orders. The *International Journal of Clinical Practice*, *75*(12), e14980. https://doi.org/10.1111/ijcp.14980.

Sarıyer, G., & Ataman, M.G. (2022). The likelihood of requiring a diagnostic test: Classifying emergency department patients with logistic regression. *Health Information Management Journal*, *51*(1), 13-22. https://doi.org/10.1177/1833358320908975.

Sarıyer, G., & Öcal Taşar, C. (2020). Highlighting the rules between diagnosis types and laboratory diagnostic tests for patients of an emergency department: Use of association rule mining. *Health Informatics Journal*, *26*(2), 1177-1193. https://doi.org/10.1177/1460458219871135.

Sariyer, G., Ataman, M.G., Mangla, S.K., Kazancoglu, Y., & Dora, M. (2023a). Big data analytics and the effects of government restrictions and prohibitions in the COVID-19 pandemic on emergency department sustainable operations. *Annals of Operations Research*, *328*, 1073-1103 https://doi.org/10.1007/s10479-022-04955-2.

Sariyer, G., Kahraman, S., Sözen, M.E., & Ataman, M.G. (2023b). Fiscal responses to COVID-19 outbreak for healthy economies: Modelling with big data analytics. *Structural Change and Economic Dynamics*, *64*, 191-198.

Sariyer, G., Öcal Taşar, C., & Cepe, G.E. (2019). Use of data mining techniques to classify the length of stay of emergency department patients. *Bio-Algorithms and Med-Systems*, *15*(1), 20180044. https://doi.org/10.1515/bams-2018-0044.

Sidey-Gibbons, J.A., & Sidey-Gibbons, C.J. (2019). Machine learning in medicine: A practical introduction. *BMC Medical Research Methodology*, *19*, 1-18. https://doi.org/10.1186/s12874-019-0681-4.

Solayman, S., Aumi, S.A., Mery, C.S., Mubassir, M., & Khan, R. (2023). Automatic COVID-19 prediction using explainable machine learning techniques. *International Journal of Cognitive Computing in Engineering*, *4*, 36-46. https://doi.org/10.1016/j.ijcce.2023.01.003.

Sozen, M.E., Sarıyer, G., & Ataman, M.G. (2022). Big data analytics and COVID-19: Investigating the relationship between government policies and cases in Poland, Turkey and South Korea. *Health Policy and Planning*, *37*(1), 100-111. https://doi.org/10.1093/heapol/czab096.

TSI. (2021). *Turkish statistical institute*. Death Cause statistics. https://data.tuik.gov.tr received, 25.09.2021.

Weng, S.F., Reps, J., Kai, J., Garibaldi, J.M., & Qureshi, N. (2017). Can machine learning improve cardiovascular risk prediction using routine clinical data? *PloS One*, *12*(4), e0174944. https://doi.org/10.1371/journal.pone.0174944.

Publisher's Note- Ram Arti Publishers remains neutral regarding jurisdictional claims in published maps and institutional affiliations.