International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 01 | Jan 2021 www.irjet.net p-ISSN: 2395-0072

MACHINE LEARNING TECHNIQUES FOR CARDIOVASCULAR SCORE PREDICTION – A REVIEW

Anagha A Kashyap[1], Anusha.L[2], Rachana.S[3], Sahana.V[4], Jayasudha.BSK[5]

[1-4]*Student, Dept. of ECE, KSIT, Karnataka, India*
[5]*Asst. Professor, Dept. of ECE, KSIT, Karnataka, India*

Abstract - *Cardiovascular disease (CVD) is one of the most censorious human diseases in the world and affects human life very badly. Accurate and on time diagnosis of CVD is very important. To evaluate CVD risk status and prevalence of CVD, noninvasive-based methods such as machine learning (ML) and signal processing are reliable and efficient. The probability of CVD developing over years in humans was predicted with knowledge of their age, blood pressure, cholesterol concentration, and smoking habits as documented at the time of initial screening examination. In addition, elevated CVD risk factors are associated with a higher risk of developing heart conditions in humans, variations in their ECG signals and also clinical parameters leads to CVD. Consequently, CVD risk score came into picture where QRISK score will decide whether the patient is at low, moderate or high risk of developing CVD in the next 10 years. . A developed machine-learning-based CVD risk score system is used for CVD prediction by using real time ECG signals, and dataset. Also, some popular machine learning algorithms, feature selection algorithms, and classifiers performance evaluation is used to estimate metrics such as classification accuracy, specificity, sensitivity, and confusion matrix. This also tells the status of the patient being at low/moderate/high risk with respect to CVD. In this study, existing cardiovascular risk predictors are studied in detail. The implementation of these predictors is based only on the features extracted from ECG. This understanding of the predictor in terms of its accuracy, specificity and other results led to adding few vital parameters along with ECG in our proposed work. The proposed machine-learning-based system will assist doctors to diagnose better.*

***Key Words*: CVD-Cardiovascular diseases, RA-rheumatoid arthritis, CAC-coronary artery calcification, FRS-Framingham Risk Score, ACC/AHA-American College of Cardiology/American Heart Association, RRS-Reynolds risk score, ML-Machine Learning model.**

INTRODUCTION

The general population has become more concerned with health, especially in the area of cardiovascular disease (CVD). It's been said that a careful history and physical examination can lead to an accurate diagnosis in the majority of patients. But cardiac events may rarely present with symptoms that are more gastrointestinal than cardiovascular. Hence physical cardiac examination becomes seldom to common people. CVD is a type of disease that affects the heart and blood vessels. Any changes in the regular rhythm of the heart, pacing, number of beats, and other irregularities cause cardiac problems. Few are considered to be minor and some are very serious conditions that lead to death if immediate medication is not inclined. It is the main cause of mortality in humans. Most CVD can be prevented by addressing behavioral risk factors that are present in the human body. Approximately 90% of individuals with CVD have at least 1 antecedent, traditional risk factors such as smoking, diabetes, hypertension, and hypercholesterolemia which has been regarded as one of the complex and life-threatening human diseases in the world. Providing correct treatment at the right time can only save the life of a person.

Many national and international authorities recommend that CVD risk assessment using CV risk scores could be used to identify individuals at high risk of CVD. However, not many developing countries in the world have implemented this approach and different countries are in progress in adopting this concept where India is one of those. To intensify the adoption of such a policy in India, new CVD risk score prediction charts are done which tells the status of the people being at low/moderate/high risk with respect to CVD. One of the important objectives of Engineering in cardiology is to find the CVD risk score of an individual which would bring an awareness to the human community. Machine learning with statistical analysis and signal processing helps achieve these goals. Our motto behind this is to assess the risk score of people for better management of their heart health.

Literature Review

[1] The aim of this paper is to compare various calculators for CVD risk assessment and statin eligibility. Successive patients who came across their myocardial infraction where their CVD risk also included was calculated using Framingham Risk score- Coronary heart disease (FRS-CHD), Framingham Risk Score- Cardiovascular Disease (FRS-CVD), QRISK2, Joint British Society risk calculator 3 (JBS3), American College of Cardiology/American Heart Association (ACC/AHA), atherosclerotic cardiovascular disease (ASCVD) and WHO risk charts. Inter heart and Interstroke survey results showed that more than 86% of CVD led to key risk factors like smoking, lipids, hypertension, diabetes, obesity, diet, physical activity, alcohol consumption and psychosocial factors. Unlike other factors, diabetes mellitus was found to be the highest risk factor in South Asia. They have considered HDL level and triglyceride level as one of the data parameters. Framingham Risk Score-Cardiovascular Disease (FRS-CVD), ACC/ AHA Atherosclerotic Cardiovascular Disease (ASCVD) risk score,

QRISK2, Joint British Society calculator-3 (JBS3), Framingham Coronary Heart-Disease Risk Score (FRS-CHD) were also used for the calculations. In FRS-CVD, diabetes was considered as a high-risk factor for score calculation. In QRISK2 the presence of chronic kidney disease, atrial fibrillation, rheumatoid arthritis, family history of CVD, ethnicity along with body mass index were also considered along with the classical risk factors. JBS3 used the same risk factors for risk score calculation as QRISK2. Many associations developed Atherosclerotic Cardiovascular Disease (ASCVD) risk score calculator which is used in clinical practice to find and treat high-risk populations as well as to communicate risk effectively to people. After the study, they concluded that FRS-CVD risk assessment model has performed the best as it could identify the highest number of patients (51.9%) to be at high risk of CVD while WHO and ASCVD calculators have performed the worst only 16.2% and 28.3% patients respectively were stratified into high CVD risk. Also, QRISK2, JBS3 and FRS-CHD have performed intermediately.

[2] The aim of this paper is to estimate the prevalence of subclinical CVD and its associations with endogenous estradiol levels and demographic, anthropometric, and metabolic variables in postmenopausal women. The study protocol was performed by using the parameters such as body weight, height, waist circumference (WC) and Blood pressure. All the data was expressed in terms of mean and SD. The regression model was used to find the independent associations between common carotid IMT (IMT-CC) and age, BMI, and estradiol levels. Adjustments were made according to systolic blood pressure and LDL-cholesterol levels. All analyses were performed using the Statistical Package for the Social Sciences. A low/medium FRS was present in 97.9% of participants. Using atherosclerotic cardiovascular disease (ASCVD) risk estimator 56.6% of participants presented a risk lower than 7.5% and all others had an intermediate risk. In conclusion, they found a moderately high prevalence of subclinical atherosclerosis in a sample of postmenopausal women with low/medium risk of CV as determined by the FRS, and recommended that BMI, age and endogenous estrogen, as well as cardio- vascular risk factors, such as blood pressure and LDL-cholesterol levels may affect on the presence of subclinical CVD in recently postmenopausal women. Also, survey results that postmenopausal women experience a high transition of CV risk.

[3] This paper talks about the increased risk of atherosclerotic cardiovascular disease in rheumatoid arthritis patients which is estimated with the help of Framingham Risk Score (FRS). The most important hypothesis of this paper is to test whether American College of Cardiology/American Heart Association (ACC/AHA) 10year risk score is able to perform better than the FRS and RRS. Along with this, the main aim of this paper is to categorize the elevated cardiovascular risk based on high coronary artery calcification (CAC) scores. The dataset used consist of 98 RA patients for predicting the risk with the help of ACC/AHA 10-year risk score. FRS capable of finding the increased risk of cardiovascular diseases in RA patients, but due to the limitations of this FRS, the model is trained with C-reactive protein which leads to the formation of RRS model. At this stage, one more cardiovascular risk score prediction was released from ACC/AHA which seeks to identify the increased risk of CVD in patients without the help of diabetes or clinical parameters. Different methodologies and tests like CAC assessment, laboratory test was conducted to identify which is the best CVD risk score predictor. From the obtained results, they set a threshold values for FRS, RRS and ACC/AHA risk score and performed the statistical analysis to identify the best and the accurate model. From the analysis, FRS, RRS and ACC/AHA was able to predict the presence of highCACwere0.65,0.66, and 0.65, respectively. From this we got the conclusion that, the ACC/AHA does not offer much advantage in identifying the elevated cardiovascular risk in RA patients compared to FRS and RRS which was determined by high CAC.

[4] The main aim of this paper is to acknowledge the difference or the variation between the produced 10-year cardiovascular risk and 10-year cardiovascular risk obtained from the raw data using the Framingham Risk Score for patients in the Systolic Blood Pressure Intervention Trial (SPRINT). In April 2017, NEJM launched the SPRINT challenge to identify the CVD risk in patients. At the end of this challenge results were obtained like 61% of the patients were identified having ≥15% 10-year cardiovascular risk based on the FRS. Now, in this paper they are identifying whether the results obtained from the previous research matches with the results obtained from the raw data available from BioLINCC. On doing the analysis, they got to know that, that result does not match with what they calculated from the raw data. The mistake made in the calculation was coefficients of treated systolic blood pressure and untreated systolic blood pressure reversed. The problem was identified and rectified in this paper, again the FRS risk score was predicted for the previous data along with the raw data and identified the variation of results with the help of scatter plot. From this we can conclude that, with the help of SPRINT Data Analysis Challenge, data sharing helped in increased knowledge for operation in scientific understanding was a benefit and the analysis of NEJM correction illustrate a secondary benefit to data sharing.

[5] Cardiovascular risk scoring systems are widely-used for clinical practice including the Framingham and Q risk Score. These estimate the 10-year risk of atherosclerotic cardiovascular events including acute coronary syndrome and stroke. Non-alcoholic fatty liver disease (NAFLD) is estimated to affect 25% of the world's population and represents a spectrum of liver disease which ranges from simple steatosis (SS) to steatohepatitis (NASH) which is found in 30%-70% on biopsy, with or without fibrosis. patients with higher risk of NAFLD is considered to be associated with various markers of subclinical atherosclerosis and high-risk coronary disease. Mean platelet volume (MPV) is provided with every complete blood count result and has related to patients with atherothrombotic disease and insulin resistance. Higher MPV levels are found in

patients with more advanced fibrosis compared to earlier fibrosis, and it is associated with an increased risk of cardiovascular events in patients with NAFLD. Cardiovascular risk score for patients with NAFLD would identify patients at higher risk for major acute cardiovascular events (MACE) compared to current standard cardiovascular risk scores. Cardiovascular death, acute coronary syndrome (ACS), stroke and transient ischemic attack (TIA) were defined as major acute cardiovascular events (MACE). The distribution of variables are explored using the Shapiro-Wilk test and are normally distributed. Binary logistic regression is then used to generate a formula for the prediction of the risk of MACE. The Brier Score was used to assess the accuracy of the prediction of the formula with values which ranges from 0 (best accuracy) to 1 (lowest accuracy). Further, the Hosmer-Lemeshow test was conducted to estimate the goodness of fit for the logistic regression model with values ranging from 0 (lowest fit) to 1 (best fit). ROC (receiver operating characteristic) curves are used to assess the diagnostic performance of this algorithm and MPV compared to the established cardiovascular risk scoring systems. MPV has been showed to be associated with higher mortality within a population of patients requiring hemodialysis, a group who are at particularly high risk of atherosclerotic cardiovascular events.

[6] Cardiovascular disease (CVD) is a leading cause of death. Estimating heart disease risk in individuals, using risk calculators, is a well-respected method used to identify the patients with those who had CVD. Various CVD risk calculators are Framingham Risk Score (FRS), Reynolds Risk Score (RRS) and the Atherosclerotic Cardiovascular Disease (ASCVD) Risk Score. CVD risk factors vary widely in different group of peoples and also varied as a function of residency. The purpose of this study is to compare risk scores as a function of calculator to determine which would better discriminate CVD risk in patients. The majority people had a family history of heart disease or CVD risk factors. FRS was designed to predict CHD, stroke, heart failure, and peripheral arterial disease in the next 10 years. The 10-year Reynolds Risk Score (RRS) was designed to predict CHD and stroke but not heart failure and peripheral artery disease. It also incorporates the same risk factors as the FRS adding family history of heart disease. The 10-year ASCVD risk score predicts similar events as the RRS (CHD and stroke) and is calculated using similar risk factors as FRS. All participants were classified as low risk and high risk. This study conclude that variations in predicted CVD risk as a function of both 10-year and long-term risk calculators in this cohort of FAW. These results suggest that the 10-year FRS, RRS, and ASCVD may underestimate CVD risk whereas the lifetime ASCVD may overestimate CVD risk. CVD risk calculators are essential for primary and secondary prevention of CVD, but they must correctly determine risk. Variations in risk scores as a function of the risk score calculator and lack of risk stratification especially for the 10-year risk calculators indicate the need for studies of outcomes in relationship to risk score and determination of the contribution of individual risk factors to CVD.

[7] CVD risk factors combined with combination antiretroviral therapy (cART) related metabolic complications increase CVD risk fold in HIV infected patients compared with HIV-uninfected patients. CVD is the primary cause of mortality in the general population and a major cause of mortality in HIV infected patients. CVD risk can be calculated using one of the available CVD risk prediction models, such as the Systematic Coronary Risk Evaluation (SCORE), the Framingham Heart Study (FHS) risk scores or the atherosclerotic cardiovascular disease risk score (ASCVD). These popular CVD risk prediction models are based on distinct prospective patient cohorts and assess multiple cardiovascular end points. Prediction model of the FHS is most suitable for use. Multiple cardiovascular risk prediction models have been derived from the FHS cohort, including models of the 10-year risks of stroke, coronary heart disease (CHD) and general CVD. They have analyzed the predictions of various CVD risk prediction models, namely the D: A: D model, the FHS coronary heart disease (FHS-CHD) and general CVD (FHS-CVD) models, ASCVD and SCORE-NL. The cumulative CVD risk distribution in our cohort was calculated for each model by adding all individually calculated CVD risks. The D: A: D, ASCVD and SCORE-NL models gave similar cumulative risk distributions. This in contrast to FHS-CVD and FHS-CHD, which attributed an overall higher cumulative CVD risk to our cohort. In conclusion, in this study in HIV-infected patients, the D: A: D, ASCVD and SCORE-NL models had similarly low CVD risk predictions when compared with both FHS models (FHS-CVD and FHS-CHD). Therefore, it's advisable to predict the CVD risk for their HIV-infected patients by using HIV-specific D: A: D model or the conventional ASCVD model and possibly SCORE-NL in low CVD risk populations. It is also important to consider the different endpoints used by these models when deciding which CVD risk prediction model to use.

[8] A novel risk calculator AtheroEdge Composite Risk Score (AECRS1.0), designed by fusing CCVRF with ultrasound image-based phenotypes. Ten-year risk was computed using the Framingham Risk Score (FRS), United Kingdom Prospective Diabetes Study 56 (UKPDS56), UKPDS60, Reynolds Risk Score (RRS), and pooled composite risk (PCR) score. AECRS1.0 was computed by measuring the 10-year five carotid phenotypes such as IMT (average., max., min.), IMT variability, and total plaque area (TPA) by fusing eight CCVRFs and then compositing them. AECRS1.0 was then benchmarked against the five conventional cardiovascular risk calculators by computing the receiver operating characteristics (ROC) and area under curve (AUC) values with a 95% CI. The patients were risk-stratified into low, moderate, and high risk using the standardized thresholds. The AECRS1.0 demonstrated the best performance on a Japanese diabetes cohort when compared with five conventional calculators. 90% of CVD/strokes are attributed to conventional cardiovascular risk factors (CCVRFs) such as sex, ethnicity, age, smoking, lipids, diabetes, physical

inactivity, obesity, and hypertension. Cardiovascular risk calculators (CCVRC) completely rely on CCVRF that do not take into consideration the morphological changes in arterial walls due to atherosclerotic plaque deposition. Two important image-based phenotypes of carotid arteries such as carotid intima-media thickness (cIMT) and total plaque area (TPA) are directly associated with a 10–15% increased risk in CV events. It should also be noted that progressions in cIMT and TPA are dependent on the variations in CCVRF. This is mainly focused on the development of a novel technique called the AECRS1.0, which fuses the CCVRF with current automated carotid ultrasound (CUS) image-based phenotypes to provide a 10-year risk of CVD. However, these calculators did not utilize image-based phenotype information in their risk computation algorithms. Benchmarks the proposed AECRS1.0 against all such CCVRC. These thresholds may depend upon the types of covariates included in the risk prediction model, or patients' baseline characteristics. The proposed AECRS1.0 is the fusion of five carotid phenotypes and eight CCVRFs. Thus, risk thresholds used for CCVRC may not be suitable to initiate the statins using AECRS1.0. However, in our study, since most of the patients' baseline characteristics indicated moderate to high risk in nature, the threshold of 50% was selected for risk stratification of the patients into the high-risk class. This may not always be true in the general population and hence needs to be explored further. . ROC analyses indicated high area under the curve values for the proposed AECRS1.0 against all traditional risk factors. The current results were encouraging and more cohorts should be tried in order to validate the AECRS1.0 calculator.

[9] Aims to examine the validity of the Framingham general cardiovascular disease (CVD) risk chart in a primary care setting. 967 patients' records were randomly selected from patients who were attending follow-up in the clinic. Baseline demographic data, history of diabetes and smoking, blood pressure (BP), and serum lipids were captured from patient records in 1998. Each patient's Framingham CVD score was computed from these parameters. All atherosclerotic CVD events occurring between 1998 and 2007 were counted. In 1998, mean age was 57 years with 33.8% men, 6.1% smokers, 43.3% diabetics and 59.7% hypertensive. Median BP was 140/80 mm Hg and total cholesterol 6.0 mmol/L (1.3). The predicted median Framingham general CVD risk score for the study population was 21.5% (IQR 1.2–30.0) while the actual CVD events that occurred in the 10 years was 13.1% (127/967). The median CVD points for men was 30.0, giving them a CVD risk of more than 30%; for women it is 18.5, a CVD risk of 21.5%. Our study found that the Framingham general CVD risk score to have moderate discrimination with an area under the receiver operating characteristic curve (AUC) of 0.63. It also discriminates well for Malay (AUC 0.65, p=0.01), Chinese (AUC 0.60, p=0.03), and Indians (AUC 0.65, p=0.001). There was good calibration with Hosmer-Lemeshow test $\chi 2$ =3.25, p=0.78. ▪ the strength of our study is that it is done in a primary care setting where the cardiovascular disease (CVD) risk profiles are different from that in secondary care. ▪ strength is that it examines the validity and applicability of the Framingham general CVD risk tool in a multi-ethnic primary care population. The Framingham General CVD risk score to have a good calibration and hence, it can be used by all the three different ethnic groups. All adult patients 30 years and older, without any cardiovascular events and with documented blood pressure (BP) readings whether they were on or not on treatment, total cholesterol, high-density lipoprotein (HDL) cholesterol, smoking status, and presence or absence of diabetes mellitus (DM) were eligible for this study. These parameters were needed to compute each individual's general CVD risk level. There were 1536 patients in our original cohort. We excluded 563 patients as they did not have all the variables needed to calculate the Framingham general CVD risk score. Another six patients were excluded as we could not ascertain their CVD status by the end of 2007. Thus, a total of 967 patients (62.9%) were eligible for analysis (figure 1). Of these, 912 (94.3%) patients completed 10 years follow-up and only 55 (5.7%) patients' defaulted follow-up. Of these, 26 (2.7%) had died with 19 (73%) deaths from non-CVD causes. The Framingham general CVD risk prediction tool is valid for use in primary care population. In the absence of local risk prediction tools, the Framingham CVD risk tool can be used as a surrogate to stratify risk and hence determine the indication for pharmacotherapy. Furthermore, this Framingham general CVD risk prediction tool which is meant for use in primary care provides a good fit for use in a multiethnic population in Asia.

[10] The Framingham risk score is widely used to identify patients at increased cardiovascular risk, and women with systemic lupus erythematous (SLE) have a marked increased prevalence of cardiovascular events. Thus, we examined the hypothesis that cardiovascular risk scores would identify women with SLE who had asymptomatic coronary atherosclerosis. Ninety-three women with SLE and 65 control subjects were studied. The Framingham score and a score for younger populations developed from the Path biological Determinants of Atherosclerosis in Youth (PDAY) study were compared in both groups. Coronary atherosclerosis was ascertained by electron beam computed tomography. There were no significant differences in the median (interquartile range) Framingham [5 (2–10) compared to 7 (0–10), P 0.88] and PDAY [15 (14–18) compared to 16 (13–18), P 0.99] scores in patients with SLE and controls, respectively. Coronary atherosclerosis was associated with higher Framingham [12 (3–15) compared to 4 (1–8), P 0.008] and PDAY [17 (15–19 compared to 15 (12–18), P 0.03)] scores in patients with SLE; however, 99% of patients were classified as low-risk with a 10-year predicted risk of 1% (1–3%). Our data indicate that cardiovascular risk scores are not adequate for risk stratification in women with SLE. Measurement of coronary calcification may add information to identify asymptomatic women with lupus who might benefit from aggressive preventive measures. The overall cardiovascular risk is increased more than two-fold in patients with SLE of any age, but in young women aged 35–44 years the risk is

increased as much as 50-fold compared to women of similar age without SLE. Assessment of overall cardiovascular risk based on risk scores is a strategy that is widely used in the general population. The Framingham risk score is a composite model and uses age, sex, smoking, blood pressure, cholesterol concentrations and diabetes to estimate the risk of coronary events and stratify individuals into risk categories: low-risk (10% risk of an event in the next 10 years), intermediate-risk (10–20%) and high risk (20%),6 that determine prognosis and the need for clinical interventions. Framingham NCEP. The composite simplified coronary prediction model includes age, total and HDL cholesterol, blood pressure, antihypertensive medications and smoking. We also applied a modified Framingham risk score described by Schisterman and Whitcomb in patients with coronary artery calcification. This modified score replaces chronological age with coronary age equivalents based on coronary artery calcification percentiles according to age and sex.10 PDAY score. The PDAY score was developed to identify young individuals (aged 15–34) with a high probability of having advanced atherosclerotic lesions. Scores range from 0 to 40 and are based on age, sex, non-HDL and HDL cholesterol concentrations, smoking, hypertension, obesity and hyperglycemia.8, 9 We used a modified PDAY score by defining hyperglycemias as the presence of fasting glucose 110 mg/dL instead of the original description based on glycohemoglobin concentrations, and by including people older than 34 years of age as part of the oldest age category. The characteristics of 93 patients with SLE and 65 control subjects are shown in Table 1. Several traditional cardiovascular risk factors differed in the two groups. There were seven (7.5%) patients with SLE and five (7.7%) control subjects taking statins. Patients with lupus had significantly lower concentrations of total cholesterol, with a median (interquartile range) of 163 (139–203) mg/dL in patients and 184 (161–206) mg/dL in controls (P 0.02). Concentrations of low-density lipoprotein were also lower in patients [94 (77–118) mg/dL] than in controls [108 (89–137) mg/dL], P 0.01. Median triglyceride concentrations were higher in patients with lupus [90 (70–144) mg/dL] than controls [79 (61–108) mg/dL], P 0.03, as were concentrations of homocysteine [9.0mol/L (7.1–11.0)] compared to 7.5mol/L (6.3–8.6) in controls (P 0.001). Patients with SLE had median disease duration of six (3–11) years. Their median SLEDAI score was 4 (1–6) and the median SLICC score was 0 (0–1); 43 patients (46%) had some degree of disease specific damage.

Proposed Methodology

As we know that CVD requires extra care to improve patients' quality of life as it is causing the number one cause in the death globally. Among many CVDs, Sudden Cardiac death, Cardiac arrest etc. are the main factors for increase in cardiac mortality. Therefore, it becomes equally important to predict the CVD risk in individual people as these diseases are occurring in normal people without any pre-history of CVDs or not even any symptoms related to CVDs. From the above-mentioned literature survey, we got to know about the CVD risk score prediction in CVD's patients using different machine learning models like Logistic Regression, Neural networks, Decision trees, SVM etc. CVD risk score models are named as Framingham risk score, Reynold's risk score etc. which is capable of predicting the risk score of each individual for 10-year, after the occurrence of CVD's disease to just monitor the patient's heart in proper way to keep him away from getting risk in future.

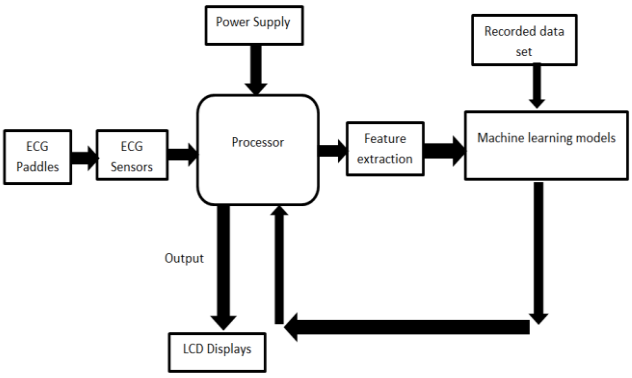

Figure 1: Block diagram of cardiovascular risk predictor

Recent studies have shown the importance of predicting risk score in CVD patients that can help on the assessment of effectiveness in preventive treatments for monitoring the heart or retrieving the heart back to its normal state in different methods either through medicines, exercises, yoga, food diet etc. The main aim of this research is to build a ML model that can predict the risk score in normal people along with the CVD patients.

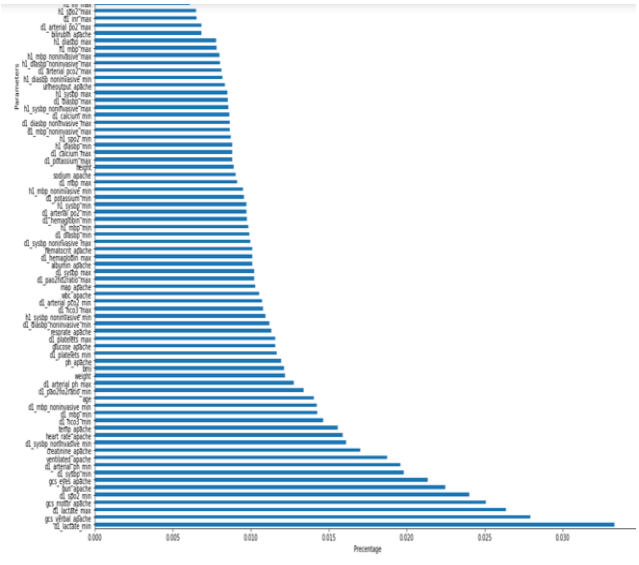

Figure 2: Vital Parameters of a patient

The goal is to extract the relevant features from long-term ECG data along with other vital parameters to increase the accuracy of the model to predict the risk of CVD's in each individual with a score. The risk score of the heart in each individual patients varies like this: (0-3) --normal range, (4-

6)--moderate range (chances of getting CVD's in future), (7-10)--high risk (confirmation of getting CVD's in future). We used the Physio net platform for data collection. Two ECG datasets from the physio net platform are selected. The first set contains the information about CVD patients and the second set about healthy people. In a similar manner, vital parameters of both CVD and normal patients are collected to further continue with the process.

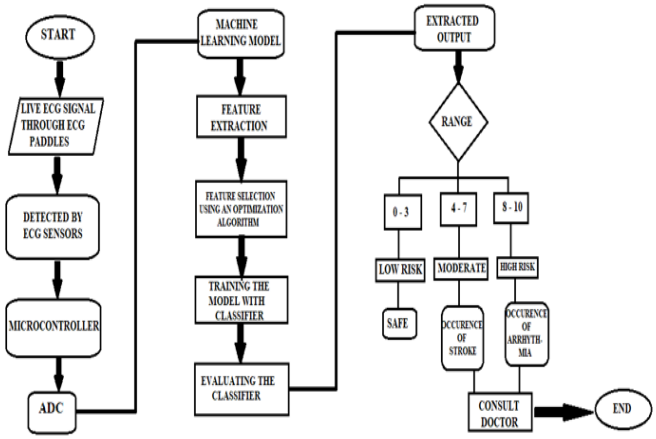

Figure 3: Flow chart of the proposed system

Conclusions

One of the major contributions of this research work is the cardiovascular risk prediction in a patient. A combination of machine learning algorithm and ECG signal features helped us to build a model of risk predictor. This predictor would definitely help the patient to understand his cardiovascular risk. At the same time helps manage heart health better if he is under a high-risk profile. The results of this predictor also favor the doctor's decision better.

The datasets used for training and testing the models have been sourced from the physio net Platform and WIDS datathon. Logistic regression is chosen as the machine learning model. Upon developing this model, it helps the people to take care of their health with preventive measures in abnormal conditions through medicines, diet, exercises etc. by checking their heart risk score, along with this; it also helps in diagnosing the heart at any condition which represents the controlled environment for a specific segment of CVD patients or normal people.

Acknowledgment

This research was supported by Kammavari Sangham Institute of Technology. We are thankful to our guide Asst.Prof.Jayasudha.BSK who provided expertise that greatly assisted the research.

References

[1] Maturana, M. A., Franz, R. F., Metzdorf, M., da Silva, T. R., & Spritzer, P. M. (2015). Subclinical cardiovascular disease in postmenopausal women with low/medium cardiovascular risk by the Framingham risk score. Maturitas, 81(2), 311-316.

[2] Garg, N., Muduli, S. K., Kapoor, A., Tewari, S., Kumar, S., Khanna, R., & Goel, P. K. (2017). Comparison of different cardiovascular risk score calculators for cardiovascular risk prediction and guideline recommended statin uses. Indian heart journal, 69(4), 458-463.

[3] Kawai, V. K., Chung, C. P., Solus, J. F., Oeser, A., Raggi, P., & Stein, C. M. (2015). The ability of the 2013 ACC/AHA cardiovascular risk score to identify rheumatoid arthritis patients with high coronary artery calcification scores. Arthritis & rheumatology (Hoboken, NJ), 67(2), 381.

[4] Warner, F., Dhruva, S. S., Ross, J. S., Dey, P., Murugiah, K., & Krumholz, H. M. (2018). Accurate estimation of cardiovascular risk in a non-diabetic adult: detecting and correcting the error in the reported Framingham Risk Score for the Systolic Blood Pressure Intervention Trial population. BMJ open, 8(7), e021685.

[5] Abeles, R. D., Mullish, B. H., Forlano, R., Kimhofer, T., Adler, M., Tzallas, A., ... & Thursz, M. R. (2019). Derivation and validation of a cardiovascular risk score for prediction of major acute cardiovascular events in non-alcoholic fatty liver disease; the importance of an elevated mean platelet volume. Alimentary pharmacology & therapeutics, 49(8), 1077-1085.

[6] Ancheta, I. B., Battie, C. A., Volgman, A. S., Ancheta, C. V., & Palaniappan, L. (2017). Cardiovascular Disease Risk Score: Results from the Filipino–American Women Cardiovascular Study. Journal of racial and ethnic health disparities, 4(1), 25-34.

[7] Krikke, M., Hoogeveen, R. C., Hoepelman, A. I. M., Visseren, F. L. J., & Arends, J. E. (2016). Cardiovascular risk prediction in HIV-infected patients: comparing the Framingham, atherosclerotic cardiovascular disease risk score (ASCVD), Systematic Coronary Risk Evaluation for the Netherlands (SCORE-NL) and Data Collection on Adverse Events of Anti-HIV Drugs (D: A: D) risk prediction models. HIV medicine, 17(4), 289-297.

[8] Khanna, N. N., Jamthikar, A. D., Gupta, D., Araki, T., Piga, M., Saba, L., ... & Gupta, A. (2019). Effect of carotid image-based phenotypes on cardiovascular risk calculator: AECRS1. 0. Medical & biological engineering & computing, 57(7), 1553-1566.

[9] Chia, Y. C., Gray, S. Y. W., Ching, S. M., Lim, H. M., & Chinna, K. (2015). Validation of the Framingham general cardiovascular risk score in a multiethnic Asian population: a retrospective cohort study. BMJ open, 5(5).

[10] Chung, C. P., Oeser, A., Avalos, I., Raggi, P., & Stein, C. M. (2006). Cardiovascular risk scores and the presence of subclinical coronary artery atherosclerosis in women with systemic lupus erythematosus. Lupus, 15(9), 562-569.

[11] Meddeb, M., Chaudhry, K. N., & Dickfeld, T. (2019). ECG PATTERNS OF INITIATION OF VENTRICULAR FIBRILLATION: ANALYSIS OF 100 ELECTROPHYSIOLOGY STUDIES. Journal of the American College of Cardiology, 73(9 Supplement 1), 375.

[12] Fokkenrood, S., Leijdekkers, P., & Gay, V. (2007, June). Ventricular tachycardia/fibrillation detection algorithm for 24/7 personal wireless heart monitoring. In International conference on smart homes and health telematics (pp. 110-120). Springer, Berlin, Heidelberg.

[13] Brown CG, Dzwonczyk R. Signal analysis of the human electrocardiogram during ventricular fibrillation: frequency and amplitude parameters as predictors of successful countershock. Ann Emerg Med. 1996 Feb;27(2):184-8. doi:10.1016/s0196-0644(96)70346-3. PMID: 8629749.

[14] Neurauter A, Eftestøl T, Kramer-Johansen J, et al. Prediction of countershock success using single features from multiple ventricular fibrillation frequency bands and feature combinations using neural networks. Resuscitation. 2007; 73(2):253–263. doi:10.1016/j.resuscitation.2006.10.

[15] Lederer W, Rheinberger K, Lischke V, Amann A. Analyse von Flimmersignalen zur Abschätzung der Defibrillierbarkeit beim Kammerflimmern [Analysis of ventricular fibrillation signals for the evaluation of defibrillation success in the treatment of ventricular fibrillation]. Anasthesiol Intensivmed Notfallmed Schmerzther. 2003;38(12):787–794. doi:10.1055/s-2003-45401