An integrated machine learning framework for effective prediction of cardiovascular Disease

Dungala seeta Rama lakshmi[1], Dr.G.A.V. Rama Chandra Rao[2]

#1M.tech Specialization:-Computer Science and Engineering, Department:-CSE

,Lendi Institute of Engineering and Technology ,Vizag-Vizianagaram Road,Jonnada, Denkada Mandal,Vizianagaram Dist - 535005 (A.P)

Email- seeta.dungala@gmail.com

#2Associate Professor (CSE),Lendi Institute of Engineering and Technology,Vizag-Vizianagaram Road,Jonnada, Denkada Mandal,Vizianagaram Dist - 535005 (A.P)

ABSTRACT_ Cardiovascular disorders are regarded as the most dangerous conditions, having the greatest fatality rate worldwide. They have become exceedingly common over time and are now overstretching national healthcare systems. High blood pressure, family history, stress, age, gender, cholesterol, BMI, and an unhealthy lifestyle are all key risk factors for cardiovascular disease. Researchers have proposed numerous ways for early diagnosis based on these criteria. However, due to the intrinsic criticality and life-threatening hazards of cardiovascular disorders, the accuracy of offered procedures and approaches need specific modifications.

A Malc add framework is proposed in this study for the effective and precise prediction of cardiovascular disorders. The methodology, in particular, addresses missing values and data imbalances first. As a result, the Feature Importance approach is used to choose features. Finally, for more accurate prediction, an ensemble of Logistic Regression, SVM, Random Forest, Decision Tree, and KNN classifiers is presented. Finally, the comparative study shows that MaLCaDD predictions are more accurate than current state-of-the-art techniques. As a result, MaLCaDD is highly trustworthy and can be used in the actual world for the early detection of cardiovascular disorders.

1.INTRODUCTION

The current era's hectic pace leads to an unhealthy lifestyle that generates anxiety and despair to cope with these conditions, people tend to engage in excessive smoking, drinking, and drug use. All of them are the root causes of many severe diseases, such as cardiovascular disease and cancer According to the World Health

Organization (WHO), cardiovascular diseases (CVDs) are the leading cause of death worldwide. CVDs account for over 31% of all fatalities worldwide Early detection of these disorders is critical so that precautionary actions can be implemented before something terrible occurs.

Cardiovascular Diseases (CVDs) are a group of conditions that affect the heart or blood vessels. Coronary Heart Disease, Stroke/Transient Ischemic Attack (TIA/ MiniStroke), Peripheral Artery Disease, and Aortic Disease are the four major kinds of CVDs The actual origin of CVDs is still unknown; however, some risk factors for these diseases include high blood pressure, smoking, diabetes, body mass index (BMI), cholesterol, age, family history, and so on. These parameters vary from person to person. Age, gender, stress, and an unhealthy lifestyle are other key factors that contribute to CVDs The main challenge is to accurately predict these diseases in time so that mortality rates can be reduced through effective medication.

2.LITERATURE SURVEY

2.1 A. L. Bui, T. B. Horwich, and G. C. Funaro, ‘‘Epidemiology and risk profile of heart failure,’’Nature Rev. Cardio., vol. 8, no. 1, p. 30, 2011.

Heart failure (HF) is a foremost public fitness trouble with a modern-day occurrence of over 5.8 million inthe USA and over 23 million worldwide.1,2 Every 12 months in the USA, extra than 550,000 folks are recognized with HF for the first time, and there is a lifetime chance of one in 5 of growing this syndrome.1,3 A analysis of HF includes big hazard of morbidity and mortality, notwithstanding advances in management. Over 2.4 million sufferers who are hospitalized have HF as a principal or secondary diagnosis, and almost 300,000 deaths yearly are immediately attributable to HF.

From the Nineteen Seventies to 1990s, a dramatic extend in the occurrence of HF and wide variety of HF hospitalizations used to be observed,4–6 and an epidemic was once declared.7,8 Most of the HF burden is borne via people aged ≥65 years, who account for greater than 80% of the deaths and time honored instances in the USA and Europe.6,9 The developing incidence of HF may mirror growing incidence, an growing old population, enhancements in the remedy of acute cardiovascular disorder and HF, or a mixture of these factors. Promising proof from country wide databases as properly as community-based cohorts, such as these based totally in Framingham and Olmsted

County,3,10–16 suggests that the incidence of HF appearsto be stabilizing, if now not decreasing, for women, and that the size of survival in sufferers with HF is increasing.

Such tendencies might also have resulted from demographic shifts, adjustments in the occurrence of danger factors, or enhancements in the availability and software of HF treatments.17,18 Furthermore, cognizance of and grasp for HF and preserved left ventricular ejection fraction (LVEF) is increasing. HF and preserved LVEF now represents >50% of HF instances and can have effects as terrible as these related with HF and decreased LVEF, however it does now not but have a validated wonderful administration strategy.19– 21 In this Review, we describe the epidemiology of HF, highlighting tendencies in general prevalence, incidence, and mortality of HF as a total and in subgroups. We additionally spotlight how recognized hazard elements affect each incidence and severity of HF and talk about the have an impact on of HF on the utilization of fitness services.

2.2 M. Park and Y.-H. Kim, ‘‘Model for predicting cardiovascular disease: Insights from a Koreancardiovascular risk model,’’ Pulse, vol. 3, no. 2, pp. 153–157, 2015, doi: 10.1159/000438683.

Between Western and Asian populations, the profile and prevalence of risk factors for cardiovascular disease (CVD) differ. For the primary prevention of CVD in asymptomatic people, the guidelines advocateindividualised interventions based on risk stratification based on CVD risk models. Current risk models for predicting CVD in Asian populations, on the other hand, are restricted. A CVD risk model for predicting global cardiovascular risk was constructed in a recent research of a large cohort of asymptomatic Korean individuals, and it performed well in predicting cardiovascular events. This strategy could be effective in the primary prevention of CVD in both East Asians and Koreans.

3.PROPOSED SYSTEM

A MaLCaDD (Machine Learning based Cardiovascular Disease Diagnosis) framework is presented in this article. MaLCaDD intends to deal with missing values and data that isn't in balance in order to improveoverall accuracy.

In this proposed system we are applying feature selection machine learning techniques they are KNN, SVM,Logistic Regression, MLP, Gaussian Naive Bayes, RNN-CNN, Random Forest Decision tree algorithms.

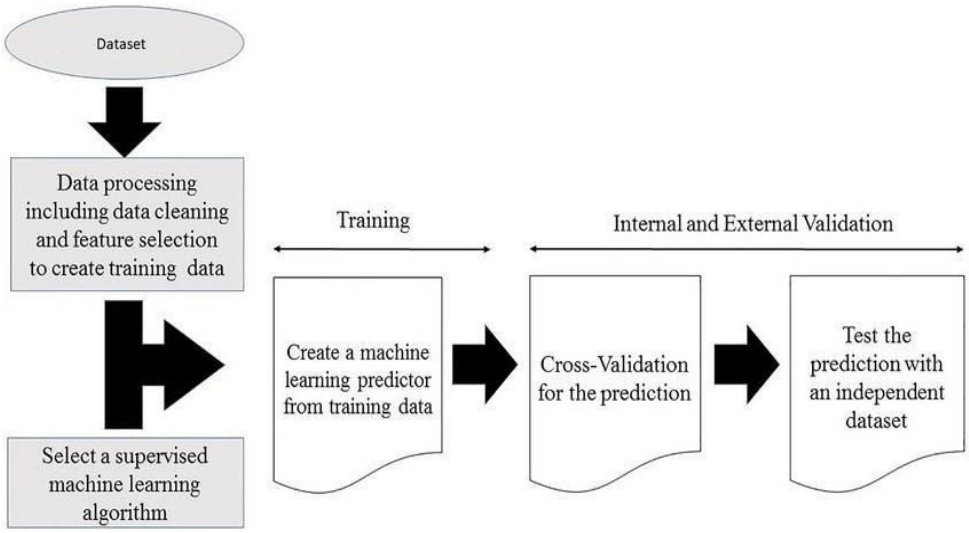


Fig 1:Architecture

3.1 IMPLEMENTATION

Data Collection: Gather a comprehensive dataset that includes relevant features and target variables related to cardiovascular diseases. This data can include medical records, patient demographics, lifestyle factors, medical test results, and historical health data.

Data Preprocessing: Clean and preprocess the collected data to ensure its quality and compatibility with machine learning algorithms. This step may involve handling missing values, outliers, and data normalization.

Feature Selection: Identify the most informative features from the dataset to improve prediction accuracy and reduce computational complexity. Consider using techniques like statistical analysis, correlation analysis, or feature importance scores from machine learning models.

Feature Engineering: Transform and engineer new features from the existing dataset to capture complex relationships and improve prediction performance. This process may involve creating new variables, aggregating data, or encoding categorical variables.

Model Selection: Choose appropriate machine learning algorithms based on the problem statement and available data. Common algorithms for cardiovascular disease prediction using various algorithms

Model Training: Split the preprocessed data into training and testing sets. Train the

selected machine learning models using the training set and tune their hyperparameters to optimize their performance. Techniques like cross-validation can be employed for robust evaluation and parameter selection.

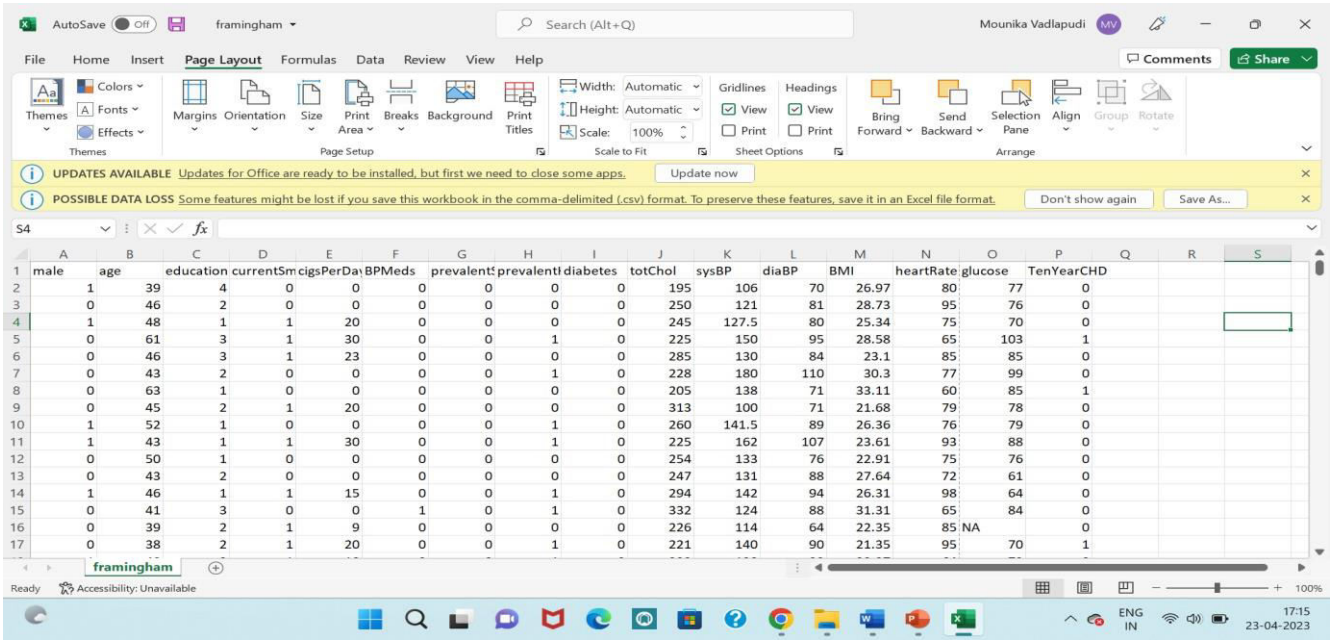
Model Evaluation: Evaluate the trained models using the testing set to assess their predictive performance. Common evaluation metrics for classification tasks in cardiovascular disease prediction include accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC).

Model Ensemble (Optional): Combine multiple models to leverage their individual strengths and improve overall prediction performance. Ensemble techniques such as bagging, boosting, or stacking can be applied to create a robust and accurate predictive model.

Deployment: Once a satisfactory model is obtained, deploy it in a production environment where it can be used for real-time prediction of cardiovascular diseases. This may involve creating an application or integrating the model into an existing healthcare system.

Monitoring and Maintenance: Continuously monitor the performance of the deployed model and update it as new data becomes available or the model's performance deteriorates. Regular maintenance ensures the model remains accurate and effective over time.

3.2 ABOUT DATASET



male	age	education	currentSm	cigsPerDa	BPMeds	prevalent	prevalentI	diabetes	totChol	sysBP	diaBP	BMI	heartRate	glucose	TenYearCHD
1	39	4	0	0	0	0	0	0	195	106	70	26.97	80	77	0
0	46	2	0	0	0	0	0	0	250	121	81	28.73	95	76	0
1	48	1	1	20	0	0	0	0	245	127.5	80	25.34	75	70	0
0	61	3	1	30	0	0	1	0	225	150	95	28.58	65	103	1
0	46	3	1	23	0	0	0	0	285	130	84	23.1	85	85	0
0	43	2	0	0	0	0	1	0	228	180	110	30.3	77	99	0
0	63	1	0	0	0	0	0	0	205	138	71	33.11	60	85	1
0	45	2	1	20	0	0	0	0	313	100	71	21.68	79	78	0
1	52	1	0	0	0	0	1	0	260	141.5	89	26.36	76	79	0
1	43	1	1	30	0	0	1	0	225	162	107	23.61	93	88	0
0	50	1	0	0	0	0	0	0	254	133	76	22.91	75	76	0
0	43	2	0	0	0	0	0	0	247	131	88	27.64	72	61	0
1	46	1	1	15	0	0	1	0	294	142	94	26.31	98	64	0
0	41	3	0	0	1	0	1	0	332	124	88	31.31	65	84	0
0	39	2	1	9	0	0	0	0	226	114	64	22.35	85	NA	0
0	38	2	1	20	0	0	1	0	221	140	90	21.35	95	70	1

FIG 2:DATASET INFORMATION

In above dataset first line contains section names and different lines contains values as 0 or 1 and in the event that patient is under Cardiovascular illness, its segment worth will be 1 else 0 and in last section contains class mark as 0 or 1 where 0 methods patient record is typical and 1 method patient record contains Cardiovascular disease. In this dataset in excess of 7000 lines are there and 16 sections are accessible .we are applying feature selection machine learning techniques they are KNN, SVM, Logistic Regression, MLP, Gaussian Naive Bayes, RNN-CNN, Random Forest Decision tree algorithms.

calculations and due to upgrade highlights Arbitrary timberland forecast exactness can be increment.

3.2.1 Attributes of dataset

The dataset contains qualities and number qualities which are circulated in a record (heart.csv) whose connection is gives toward the finish of the paper in the part of information accessibility. The attributes of the dataset used (risk factors for heart attack) are discussed below. The behavioural and attributes information of the entire dataset is provided.

Age in years:

This is a profoundly vital gamble factor for the event of coronary failures on the grounds that the gambleof getting coronary episodes can twofold as age increments. In grown-ups, the greasy streaks demonstrative of coronary course illness begins to create and it is demonstrated that over 80% instances of cardiovascular failures because of coronary illness are in patients matured 65 or above.16

Gender:

It has been demonstrated that men have a higher risk of heart attack than women under the age of 50.17 However, there is some debate regarding whether women and men have the same risk of heart attack aftermenopause. Women with diabetes are more likely to have a heart attack.

Chest pain (CP):

Angina is the condition in which the heart muscle does not receive sufficient oxygenated blood. Angina can cause squeezing or high pressure in the chest as well as discomfort in the shoulder, jaw, back, or neck.Indigestion can also occur in angina. The aggravation can be felt in the hands. Stable angina, pectoris angina, unstable angina, Prinz metal angina, and microvascular angina are all types of

angina.

BP in units of tbsps:

High blood pressure can affect the arteries. This can occur for a variety of reasons, including obesity, high sugar, and unbalanced cholesterol. which can make the risks worse. Cholesterol (chol):

Corridors again can get impacted because of imbalanced or awful cholesterol. Low-density lipoprotein cholesterol, in particular, is more likely to narrow the arteries. Blood fat, or triglycerides, with high cholesterol levels, which can also raise the risk of heart attacks, is another cause. Thus, it is prudent to keep up with great cholesterol to bring down the gamble of a coronary episode.

.

Blood sugar during a fast (fbs):

High glucose can turn into a reason for a cardiovascular failure. It could happen because the pancreas produces less hormone or the body doesn't respond to insulin.

Pulse (thalach):

According to research25, an increase of ten beats per minute (bpm) increases the likelihood of cardiac death by twenty percent, which is consistent with the correlation between an increase in blood pressure andan increase in the risk of heart disease. The same holds true for the 10 mm Hg increase in blood pressure

4.RESULTS AND DISCUSSION

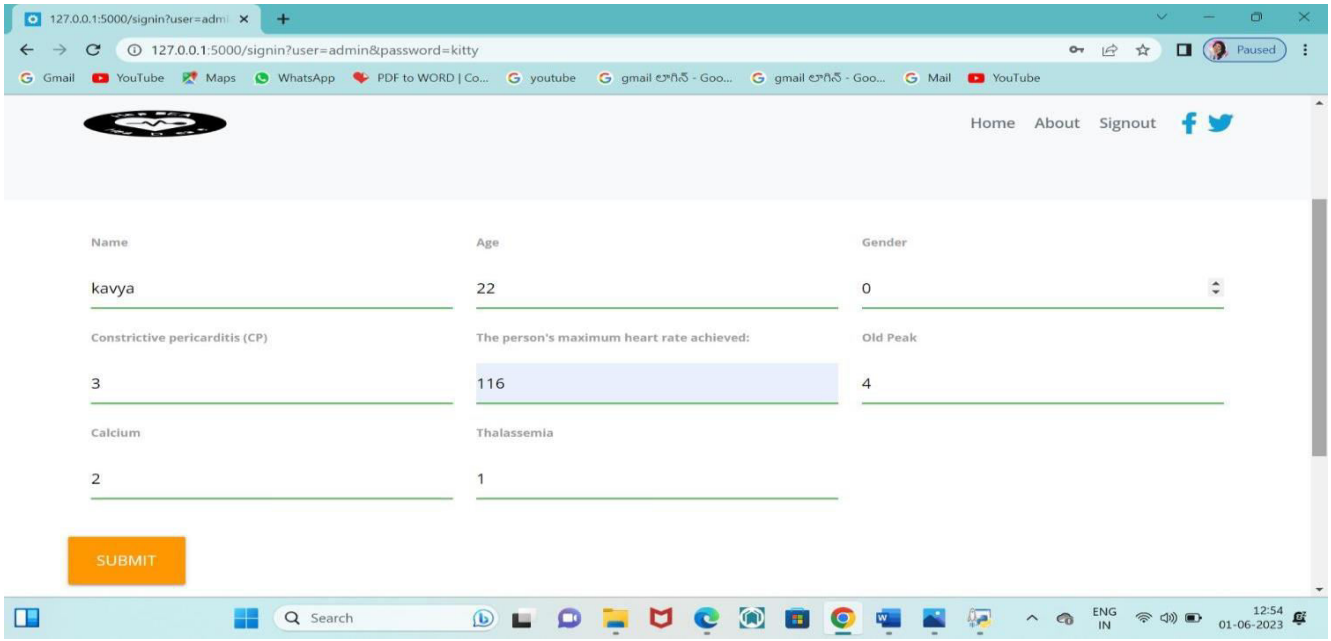


Fig 3:Once the heart disease prediction page is open, we must enter the values based on the providedinformation before clicking the submit button.

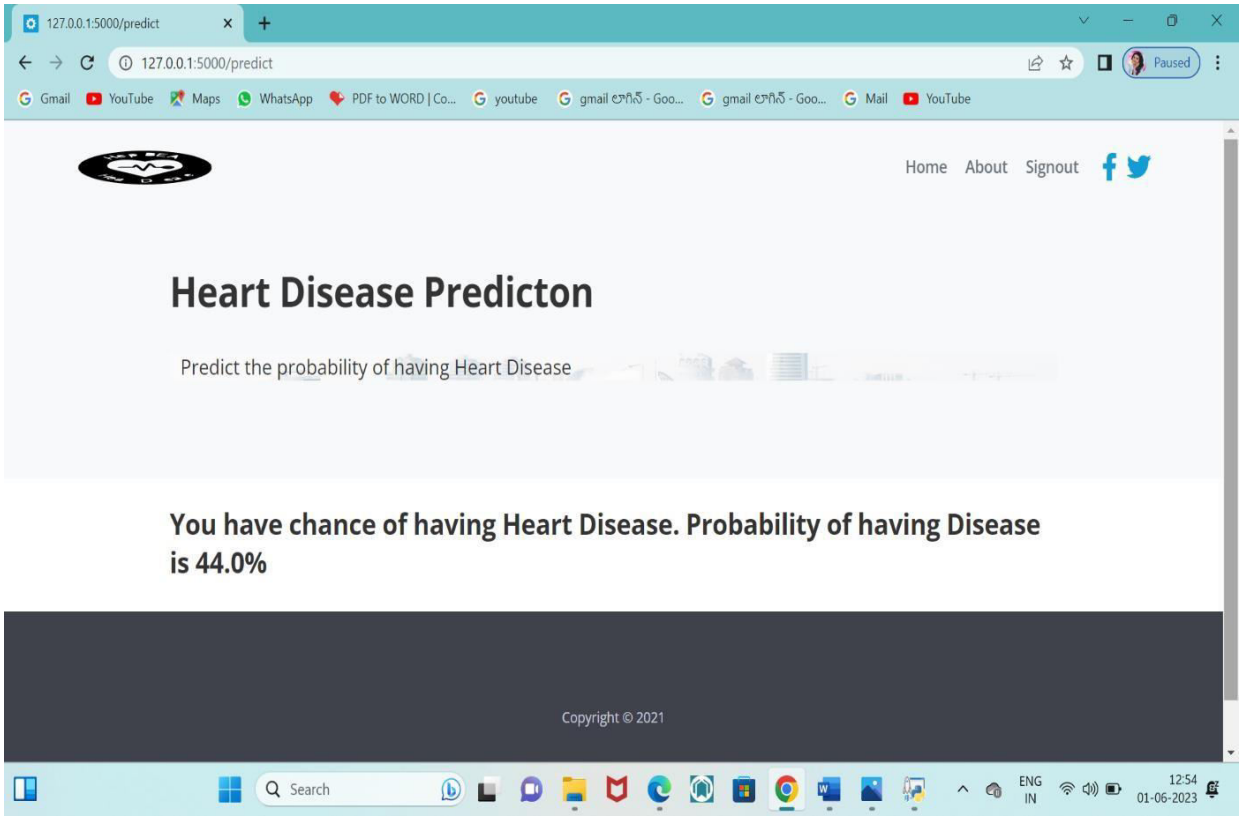


Fig 4:If the patient has a condition, it indicates that you have a chance of developing heart disease. Once the details have been submitted, we will receive the results to determine whether or not the patient has cardiovascular disease. Some likelihood of having sickness

5.CONCLUSION

This paper introduces the MaLCaDD platform for the early prediction and diagnosis of cardiovascular illnesses. The framework is divided into four primary phases, the first of which deals with missing value management using the mean replacement technique. In the second phase, the data imbalance problem is addressed using the Synthetic Minority Oversampling Technique (SMOTE). Thefeature significance technique is used to pick features in the third phase. Finally, for enhanced prediction, KNN Classifier,Logistic Regression,Decision Tree Classifier,Ensemble of LR and Decision Tree,Random Forest,SVM is proposed. The MaLCaDD implementation is written in Python and is publicly available at the GitHub repository. The comparison analysis demonstrates that MaLCaDD surpasses state-of-the-art research by improving accuracy with a smaller collection of characteristics. As a result, MaLCaDD is highly trustworthy and may be used effectively in the actual

world for the early detection of cardiovascular disorders.

REFERENCES

[1] Organization Mondale de la santo. 2017). Diseases of the heart (CVDs) Online]. Available: https://www.who.int/wellbeing points/cardiovascular diseases

[2] E. J. Benjamin et al., " Statistics on heart disease and stroke—update for 2019: The American Heart Association's report, Circulation, vol. 139, no. 10, pp. Mar. 2019, e56–e528, Doi: 10.1161/CIR.0000000000000659.

[3] Korea Statistics 2018). Statistics on Causes of Death in 2018 [Online]. Available:

[4] http://kostat.go.kr/portal/eng/pressR eleases/8/index.board?bmode=read&bSeq =&aSeq=378787 2017).Diseases of the heart (CVDs) Online]. Available:

[5] P. Greenland, J. S. Alpert, G. A. Beller, E. J. Benjamin, M. J. Bud off, Z. A. Fayad, E. Foster, M. A. Hlatky, J. M. Hodgson, F. G. Kushner, M. S. Lauer, L. J. Shaw, S. C. Smith, A. J. Taylor, W. S. Weintraub, and N. A report on practice guidelines from the American Heart Association/American College of Cardiology Foundation, Circulation, vol. 122, no. 25, pp. Dec. 2010, e584–e636, doi:10.1161/CIR.0b013e3182051b4c.

[6] J. Perk and colleagues, " 2012 version of the European guidelines for clinical prevention of cardiovascular disease: The European association for cardiovascular prevention & rehabilitation (EACPR) made a special contribution to the development of the fifth joint task force on cardiovascular disease prevention in clinical practice, which was formed by representatives of nine societies and invited experts. Heart J., vol. 33, no. 13, pp. 1635-1701, Jul. 2012, doi: 10.1093/eurheartj/ehs092.

[7] "Model for predicting cardiovascular disease:" by G.-M. Park and Y.-H. Kim A Korean cardiovascular risk model's insights, Pulse, vol. 3, no. 2, pp. 153–157, 2015, doi: 10.1159/000438683.

[8] G. J. Njie, K. K. Proia, A. B. Thota, R. K. C. Finnie, D. P. Hopkins, S. M. Banks, D. B. Callahan, N. P. Pronk, K. J. Rask, D. T. Lackland, and T. E. Kottke, "Clinical choice emotionally supportive networks and anticipation," Amer. Preventive Medicine, vol. 49, no. 5, pp. Nov. 2015, pp. 784–795, doi: 10.1016/j.amepre.2015.04.006.

[9] "Comparative impact of guidelines, clinical data, and decision support on prescribing decisions:" by V. Sintchenko, E.

Coiera, J. R. Iredell, and G. L. Gilbert J. Amer describes it as "an interactive Webssssss experiment with simulated cases." Med. Inform. Assoc., vol. 11, no. 1, pp. Jan. 2004, pp. 71–77, doi:10.1197/jamia. M1166.

AUTHOR PROFILES