

RESEARCH ARTICLE

Leveraging Regression Analysis to Predict Overlapping Symptoms of Cardiovascular Diseases

SARA GHORASHI¹, KHUNSA REHMAN², ANAM RIAZ³,
HEND KHALID ALKAHTANI⁴, (Member, IEEE), AHMED H. SAMAK⁵,
IVAN CHERREZ-OJEDA⁶, AND AMNA PARVEEN⁷

¹Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia

²Services Institute of Medical Sciences, Lahore 54000, Pakistan

³Department of General Medicine, Rawalpindi Medical University, Rawalpindi 46000, Pakistan

⁴Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia

⁵Faculty of Science, Menofia University, Shibeen El-Kom 32511, Egypt

⁶Allergy and Pulmonology, Espiritu Santo University, Samborondón 0901-952, Ecuador

⁷College of Pharmacy, Gachon University, Medical Campus, Yeonsu-gu, Incheon 21936, South Korea

Corresponding author: Amna Parveen (amnaparvin@gmail.com)

This work was supported by the Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, through Researchers Supporting under Project PNURSP2023R384.

ABSTRACT In medical informatics, deep learning-based models are being used to predict and diagnose cardiovascular diseases (CVDs). These models can detect clinical signs, recognize phenotypes, and pick treatment methods for complicated illnesses. One approach to predicting CVDs is to collect a large dataset of patient medical records and use it to train a deep learning model. This study investigated CVDs for early prediction using deep learning-based regression analysis on a dataset of 2621 medical records from UAE hospitals, including age, symptoms, and CVD information. We propose a long short-term memory-based deep neural network for early prediction of CVDs by leveraging the regression analysis. It can be seen that the accuracy level of the diseases increased when they were simulated in pairs of one disease with another due to the overlapping symptoms. The study's results suggest that coronary heart disease has been predicted with an 71.5% accuracy level, with 84.4% overlapping with Dyspnea; when accuracy measured with a combination of three conditions the accuracy was 86.7%, Dyspnea, Chest Pain, and Cyanosis, it has been increased up to 88.9%. Weakness, Fatigue, and Emptysis showed a value of 89.8%. In our proposed work, the combinations were Dyspnea, Chest Pain, Cyanosis, Weakness and Fatigue, Emptysis, and discomfort pressure in the chest have shown the ideal value of accuracy measured up to 90.6%, and with Fever, the accuracy is 91%. We show the effectiveness of our proposed method on several evaluation benchmarks.

INDEX TERMS Regression analysis, cardiovascular diseases, deep learning, disease prediction.

I. INTRODUCTION

Cardiovascular diseases (CVDs) are a group of disorders that affect the heart and blood vessels [1]. They are among the leading causes of death and disability worldwide. Some common types of CVDs include coronary artery disease, heart failure, stroke, and hypertension (high blood pressure). There

The associate editor coordinating the review of this manuscript and approving it for publication was Bing Li¹.

are many types of CVDs, and they can have a wide range of symptoms. Some CVDs, such as coronary artery disease and heart failure, can cause chest pain or discomfort, shortness of breath, fatigue, and irregular heartbeats [2]. Other CVDs, such as stroke and aneurysms, can cause weakness or numbness on one side of the body, difficulty speaking or understanding speech, vision loss, and severe headache. Different CVDs may have overlapping symptoms in some cases, making it difficult to distinguish between them based

on symptoms alone. For example, chest pain or discomfort can be a symptom of both coronary artery disease and heart failure.

Similarly, shortness of breath and fatigue can be symptoms of both heart failure and hypertension (high blood pressure). To accurately diagnose and treat CVDs, it is important to have a thorough medical evaluation, which may include a physical exam, laboratory tests, and imaging studies. In some cases, additional specialized tests may be necessary to confirm a diagnosis and determine the best course of treatment [3].

A heart or blood vessel disease is a cardiovascular disease (arteries and veins). Heart disease has overtaken infectious illnesses as the leading cause of mortality and disability globally for those over 65 years old [4]. It is now considered a second epidemic' in many nations due to the alarmingly high and steadily growing prevalence of the condition. Early identification of cardiovascular illness may help decrease the death rate. Echocardiography is one method of diagnosing cardiac conditions. Echocardiography, or echo, is an unpaintable test to make images of the heart on sound waves. The test provides information on the heart's size and shape and how effectively the heart chambers or valves function. Echo may also identify cardiac abnormalities in toddlers and babies [5].

The incidents and mortality rates due to CVDs have been increasing in recent years, and how to diagnose and prevent such diseases has become a challenging situation. Acute forms of a cardiovascular catastrophe, such as thrombosis, and atherosclerosis plaque formation, involve 15-20 years. Such long periods of silent symptoms have made it difficult to diagnose early screening healthcare physicians [6]. Most cases, such as ischemic cardiomyopathy and acute coronary syndrome, reach progressive phases of the condition at the first appointment. Moreover, though AS plaques have been discovered, it is a struggle to address whether or not lesions need involvement or weakness in clinical practice [7]. Meanwhile, some of the CVDs may share similar characteristic symptoms like chest pain, irregular heartbeat, shortness of breath, pain in the neck, jaw, throat, upper abdomen, or back, and numbness, which increases the chances of physicians' diagnosis misleading.

Such post-determined symptoms and silenced indications for CVDs can be estimated and assessed using deep-learning neural networks. Deep Belief Networks, Medical Image Segmentation [7], [8], and Bayesian Networks can use the sizeable clinical dataset to predict risk factors' impacts. In this research, the researcher will develop a predictive model based on machine learning algorithms to assist and precisely predict early symptoms of heart diseases in the United Arab Emirates (UAE).

In recent years, deep learning has played a significant role in detecting CVDs [9]. Deep learning algorithms are a type of artificial neural network that can learn and recognize patterns in data, making them well-suited for analyzing large and complex datasets commonly found in medical imaging and diagnostic testing. Deep learning has been instrumental in

detecting and diagnosing heart disease from medical images such as echocardiograms, angiograms, and MRI scans. Deep learning algorithms can learn to identify specific patterns and features in these images that may indicate the presence of heart disease, including abnormalities in heart structure, blood flow, and tissue texture [10].

In addition to image analysis, deep learning has also been applied to other types of cardiovascular data, such as electrocardiograms (ECGs) and patient health records. For example, deep learning models have been used to analyze ECGs and predict the risk of arrhythmia and other cardiac events. They have also been used to analyze patient health records and predict the risk of cardiovascular disease based on factors such as age, sex, family history, and lifestyle factors. Overall, the use of deep learning in cardiovascular disease detection and diagnosis is still in its early stages, but the results are promising. As more data becomes available and the technology improves, deep learning is likely to become an increasingly important tool in the fight against heart disease [11].

Deep learning uses artificial neural networks to learn patterns and make predictions from large datasets. It has shown great promise in healthcare, especially in image analysis and natural language processing [12], [13]. One advantage of deep learning is its ability to learn features from raw data, making it useful in cases where manual feature engineering is complicated. Deep learning models can handle large amounts of data and scale well to new tasks, making them well-suited for real-world healthcare applications [14], [15]. One of the methods to predict CVD is using regression analysis using deep learning-based feature extraction from data patterns [16]. In another work [17], the authors propose to use machine learning algorithms for chronic disease outbreak prediction in disease-frequent communities. They also applied the prediction models to hospital data collected from China between 2013-2015. The authors employed latent component model to rebuild missing data from central China hospital medical records. Second, statistical information might identify the region's significant chronic ailments. CNN algorithm selects characteristics for unstructured text. Finally, they provide a CNN-based multimodal disease risk prediction method for structured and unstructured data.

Regression analysis predicts continuous outcomes from predictor variables. Regression models may predict CVD risk using age, blood pressure, cholesterol, and other risk factors [18]. Linear, logistic, and non-linear regression models might be utilized. The data and prediction task determine the model. The regression model should be thoroughly assessed for accuracy and reliability. Divide the dataset into training, validation, and test sets and test the model on the test set. Cross-validation also improves model performance estimates. Regression models may assist in identifying high-risk CVD patients and advise preventive and therapy [19].

This work aims to create a deep learning-based regression analysis framework for early CVD symptom prediction. As the literature has made clear, deep neural networks may

construct algorithms that anticipate and extract outputs based on training subjects using large datasets. We concentrate on the influence of overlapping symptoms in CVDs, the use of the UAE clinical dataset as a predictor of risk variables, and the addition of the literature to enhance the diagnostic methods for CVDs. For predicting the overlapping symptoms of CVDs, we use simple linear regression (SLR) and multiple linear regression (MLR) analyses.

The developed deep learning-based framework's results suggest that coronary heart disease has been predicted with an 71.5% accuracy level, with 84.4% overlapping with Dyspnea; when accuracy measured with a combination of three conditions the accuracy was 86.7%, Dyspnea, Chest Pain, and Cyanosis, it has been increased up to 88.9%. Weakness, Fatigue, and Emptysis showed a value of 89.8%. In our proposed work, the combinations were Dyspnea, Chest Pain, Cyanosis, Weakness and Fatigue, Emptysis, and discomfort pressure in the chest have shown the ideal value of accuracy measured up to 90.6%, and with Fever, the accuracy is 91%. We show the effectiveness of our proposed method on several evaluation benchmarks.

The rest of this work is organized as follows. Section II reviews the literature about CVD and deep learning models. The methodological description of this paper is available in Section III. We have presented the experimental evaluation in Section IV. A discussion related to the results of the proposed study is available in Section V. Finally, we conclude this work in Section VI.

II. LITERATURE REVIEW

A. MEDICAL BACKGROUND

The category of illnesses connected with the heart and blood vessels is cardiovascular disease (CVD). This involves blood vessel restriction that brings blood to the heart, such as coronary heart disease, Cerebrovascular disorder [20] in the head and peripheral arterial disease in the limbs. Second, cardiac and heart valve tissue damage induced by rheumatic fever, such as rheumatic heart attack and heart muscle failure, pumps more blood into the blood vessels.

The rising incidence of CVD places a heavy financial burden on public and private healthcare systems [21]. CVD is a burden on the economy, and the leading cause of preventable death worldwide [22]. A patient experiencing symptoms or a high risk of heart disease may undergo a battery of diagnostic procedures, including an electrocardiogram (ECG) and a stress test, to determine the best course of treatment. The diagnostic and therapeutic costs associated with this approach are quite high due to its high complexity and high demands on time and resources.

Troubles with the normal pace or rhythm of one's heartbeat constitute arrhythmia. It might be an indication of a rapid, sluggish, or irregular heartbeat. Whenever the heart beats quicker than usual, tachycardia is termed. If the heart beats too slowly, bradycardia is termed. Atrium fibrillation is the most frequent form of arrhythmia, which produces irregular and rapid heartbeat [23].

Noale et al. [24] focused their research on the epidemiology of CVDs among individuals aged 65 years and older. They have examined risk factors associated with the most common causes of disability and mortality in this age group, including hypertension, lipids and lipoproteins, smoking, diabetes, physical inactivity, obesity, and risk of acute stroke.

Athanasίου et al. [25] studied the use of CVD risk prediction models. These models utilize various techniques to improve human understanding of the reasoning process behind CVD risk assessment, increase transparency, and foster trust in the results. One such technique is the XGBoost algorithm, which has gained popularity for its computational speed, generalization capabilities, and high predictive accuracy. They also explored how these techniques could be used to understand and improve existing CVD risk calculators, such as Framingham, SCORE, and DECODE. Ultimately, the team proposed an innovative approach to CVD risk prediction, utilizing the XGBoost and SHAP methods to create an explainable model.

Charlton et al. [26] discussed the basics of wearable photoplethysmography (PPG) and its analysis. With the increasing popularity of smart wearables equipped with PPG sensors, there is an opportunity to monitor cardiovascular health in a convenient and non-invasive way. The PPG signal, which measures arterial blood volume, is explored along with various design configurations. The study highlights the potential applications of wearable PPG devices, such as detecting atrial fibrillation, identifying obstructive sleep apnea, monitoring the spread of infectious diseases, and assessing mental stress. Overall, the authors emphasize the exciting opportunity that wearable PPG devices present for daily monitoring of cardiovascular health.

Chieng and Kistler [27] reviewed the impact of habitual coffee and tea consumption on the prevention of CVD. Their findings suggest that moderate coffee/caffeine intake, equivalent to 2-3 cups daily, can positively affect metabolic syndrome. Additionally, coffee consumption has been linked to a decreased risk of coronary heart disease. Results from the study indicate a 9% reduction in the risk of hypertension with daily consumption of 7 cups of coffee. Moreover, coffee and tea consumption may benefit CVDs like arrhythmia, heart failure, stroke, and cardiovascular mortality.

The triglyceride-glucose (TyG) index, which has been suggested as a different biomarker for insulin resistance (IR), was reviewed by Tao et al. [28]. Reduced insulin sensitivity and response, or IR, is a CVD predictor. Combining lipid- and glucose-related components makes up the TyG index. According to studies, the TyG index is a trustworthy and practical instrument for determining IR and a predictor of obesity, hypertension, and other conditions. When evaluating IR in people with and without diabetes, the TyG index is superior to the widely used HOMA-IR. The TyG index, according to the authors, may be optimized for risk classification and CVD outcome prediction.

Battineni et al. [29] proposed a study to evaluate the effectiveness of telemedicine systems in providing personalized care for preventing CVD and reducing the need for hospital visits. The use of telemedicine platforms can facilitate regular checkups for CVD patients, and the concept of precision medicine is applicable throughout all stages of CVD. Telemedicine involves securely transmitting clinical information and data using various forms such as voice, text, or images. This technology allows patients and healthcare professionals to have phone or video consultations, providing significant benefits. Telecardiology is the application of telemedicine in cardiovascular medicine. The study highlights the crucial role of telemedicine in delivering personalized care to CVD patients.

Bays et al. [30] published a summary document on the various risk factors associated with CVD. The survey identified ten key risk factors, which include unhealthy dietary habits, physical inactivity, dyslipidemia, hyperglycemia, high blood pressure, obesity, demographics (such as older age, race/ethnicity, and sex differences), thrombosis/smoking, kidney dysfunction, and genetics/familial hypercholesterolemia. It is common for CVD patients to have multiple risk factors, requiring a multifaceted approach to prevention. The document also highlights the importance of physical activity, dietary control, and intermittent fasting in reducing the risk of CVD. Overall, the study emphasizes the significance of identifying and addressing various risk factors associated with CVD to improve prevention strategies.

Dickson et al. [31] conducted a study exploring self-care among older workers with CVD and the impact of work-related factors. The study is based on the middle-range theory of self-care, which defines self-care as a natural decision-making process. Work-related factors, such as job stress resulting from poor work organization, can negatively affect the health of workers with CVD by increasing physiological demands. The study examines several factors that can reduce the risk of CVD in the workplace, including job control, workplace support, work-life balance, and organizational justice. A multi-level intervention is needed to promote sustainable self-care among older workers with CVD within the workplace and mitigate barriers to self-care.

Depression diagnosis also relies on somatic symptoms that coincide with the overlapping symptoms of certain medical conditions. Ellis et al. [32] studied disease signs correlated with depressive symptoms. The author's studies reviewed reported interviews with 46 of 61 qualifying populations residing among older adults with advanced disorders and a significant number of somatic depressive symptoms [33]. Participants replied to an interactive query regarding emotions and legal questions about depressive symptoms. In Hsu et al. [34] studies, The overlapping general populace is very frequent with gastroesophageal reflux (GERD) and dyspepsia, although the relationship between each is little known. 107 participants had "epigastric pain or burning" overlaps, and 761 did not develop these symptoms. Subjects of

GERD-D showed more serious GERD signs and were more frequently linked to GERD alone (IBS) syndrome [34].

Overlapping cardiovascular problems, including palpitations, tightness in the chest, and shortness of breath that arise in healthy individuals, like stress, suggests doctors and their patients have a tough time assigning their causal or associated function to mental well-being. Primary care clinicians and cardiologists prioritize the management of signs and risk conditions, allowing less opportunity to cope with thoughts and emotions [35].

Though requirements for the formal diagnosis of generalized anxiety disorder are not fulfilled, general concern over everyday activities has been found to encourage and accelerate coronary disease in psychological Pain, including anguish and tension [36]. The impact can also be progressive, indicating that further episodes of anxiety, cold, and fatigue, despite the normal coronary arteries or palpitations or heart beating, may be correlated with a greater incidence of cardiovascular disorders, as well as with the combination of symptoms or indicators of CVDs, such as chest pain act as overlapping symptoms of CVDs [37].

B. MODELING METHODS

Deep learning is a technique for constructing high-dimensional predictors in input-output models by reducing high-dimensional data. Specifically, deep learning seeks to select, research, and acquire information from abundant data, creating a concise pattern capable of explaining enigmatic trends or relationships and addressing challenging problems [38].

Yang et al. [39] conducted a study using electronic health records where they leveraged logistic regression analysis for the CVDs. They developed prediction models using machine learning-based models such as random Forest, Ada Boost, Bagged tree, and multivariate regression models.

The work by Swathy and Saruladha [3] explores the potential of AI and Data Mining in predicting CVD. The study compares various models used for CVD prediction, including classification, data mining, machine learning, and deep learning. Authors identified four categories of CVD and proposed a prediction system consisting of data collection, preprocessing, classification, and model evaluation. They also compared datasets with prediction and deployment models, using tools such as WEKA, TANGARA, and MATLAB for cardiovascular disease.

Ghosh et al. [40] efficient data collecting, preprocessing, and transformation model are correct. The research selects features using Relief, Least Absolute Shrinkage, and Selection Operator methods. For training, the authors created hybrid classifiers including Decision Tree Bagging Method, Random Forest Bagging Method (RFBM), AdaBoost, K-Nearest Neighbors, and Gradient Boosting. The research calculated Accuracy, Error Rate, and False Positive/Negative Rate using machine learning techniques. The research used Decision Tree, K-Nearest Neighbors,

AdaBoost, Gradient Boosting, Random Forest, and hybrid classifiers. RFBM and Relief feature choices gave the model the maximum accuracy of $98.05\pm\%$.

Al-Absi et al. [41] performed case-control research to diagnose CVD using retinal pictures and DXA data. Their deep learning-based solution employed DXA scans and retinal fundus pictures to concentrate on the retinal image center. This non-invasive, fast approach diagnosed CVD with approximately 75% accuracy, according to the research. The dataset was only tested on Qatari people; therefore, its effectiveness in different demographics and locales may vary. A unique DL-based model incorporating retinal pictures and DXA scans to differentiate CVD from control groups achieved 78.3% accuracy.

Kumar et al. [42] examined how machine learning classifiers may predict cardiovascular disease from patient complaints. The study cleaned and analyzed a dataset using SVM, Logistic Regression, and KNN classifiers. The Random Forest classifier method yielded $85.71\pm\%$ accuracy and 0.8675 ROC AUC. Confusion matrices assessed classifier accuracy.

Abdalrada et al. [43] created and tested a two-stage machine learning model to predict DM and CVD. In the first step, a Multivariate Adaptive Regression Splines (MARS) model using logistic regression (LR) and Evimp functions (EVF) identified significant shared risk variables based on voting criteria. The correlation matrix eliminated duplicate risk variables. DM and CVD prediction models were created using a Classification and Regression (CART) method in the second step. The first step extracted common risk variables from the dataset and utilized machine learning to forecast comorbidities. Clinicians may benefit from the authors' ML model's excellent accuracy, sensitivity, and specificity.

Kee et al. [44] reviewed machine learning-based cardiovascular disease (CVD) prediction models for type 2 diabetics. Scopus and WoS were searched systematically. Logistic regression predicted early CVD in diabetic patients, although machine learning is more flexible and variable. Unlike regression models, machine learning algorithms use many algorithms to get the best match. NN, SVM, DT, and k-NN machine learning algorithms have been used to create prediction models. The suggested models were tested using C-value, sensitivity, accuracy, precision, and area under the curve.

Venkat et al. [45] conducted a study using AI/ML techniques to analyze RNA-seq data from CVD patients. They aimed to identify genes associated with heart failure (HF), atrial fibrillation (AF), and other CVDs, and accurately predict disease using the data. To achieve their research objectives, they developed a new approach called "Findable, Accessible, Intelligent, and Reproducible." They proposed a model that included clinical data analysis and the implementation of AI/ML for predictive analysis. The AI/ML methodology was structured into input, data preparation, methodology, data analysis, outcome, and validation.

TABLE 1. Dataset collection information.

Hospital Name	Type
Ahalia Hospital	Private
Al Corniche Hospital	SEHA
Al Gharbia Hospitals	SEHA
Al Mafraq Hospital	SEHA
Al Noor Hospital	Private
Al Raha Hospital	Private
Al Rahba Hospital	SEHA
Al Salama Hospital	Private
American European Medical Center	Private
The British Clinic	Private
Burjeel Hospital	Private
Cleveland Clinic	Private
Dar Al Shifaa	Private
Exeter Medical Center	Private
Franco-Emirien Hospital	Private
Gulf Diagnostic Center	Private
Ibn Nafees Medical Center	Private
LLH Hospital	Private
New Medical Centre (NMC)	Private
Silkor Laser Hair Removal	Private
Universal Hospital	Private
Welcare Hospitals	Private
American Hospital Dubai	Private
Aster Hospital	Private
Belhoul European Hospital	Private
Belhoul Speciality Hospital	Private

Wang et al. [46] presented a methodology for high-dimensional; pattern regression for medical images using machine learning techniques. In their proposed work, the preprocessed training images are taken and sub-sampled. After that, feature extraction and selection are applied to the medical images data, and then the relevance vector regression is applied. Their proposed framework is built upon regression using a relevance vector machine strategy. This work is closely related to our approach in this paper.

In another work by Mudassir et al. [47], the authors proposed applying classification and regression problems to bitcoins data. In their proposed framework, a feature selection is applied to the data after collecting and pre-processing the data. Afterward, the models are trained for classification and regression, where predictions on the data are made using trained data.

III. MATERIALS AND METHODS

A. DATA

Due to the complexity of the research and the need to include as many participants as possible, a quantitative approach is used in this situation. In the case of convenience sampling, the likelihood of the variable being chosen from the survey is not understood because there can be no inferential statistics concerning the population. Convenience sampling would be used for our test study, which is the kind of unlikely sample. The total number of models used in the present study is 2621 entries from different UAE hospitals such as Al Gharbia Hospitals, Al Mafraq Hospital, Al Noor Hospital, Al Raha Hospital, and Dar Al Shifaa. A complete list of the hospitals is available in Table 1.

To evaluate the study's objectives, the population for sampling will consist of UAE hospitals, and a dataset Consisting

TABLE 2. Dataset description.

Variable name		Variable description	Designation
Basic Information	id	Patient ID	[0, 1]
	gender	Age	[0, 1]
	age	Age years	[0, 1]
	cgh	Cough	[-1, 0.5, 1]
Symptom	emp	Emptysis/Expectoration	[-1, 0.5, 1]
	dys	Dyspnea	[-1, 0.5, 1]
	gas	Gasping/Panting	[-1, 0.5, 1]
	nau	Nausea/vomiting	[-1, 0.5, 1]
	dizz	Dizziness	[-1, 0.5, 1]
	head	Headache	[-1, 0.5, 1]
	fev	Fever	[-1, 0.5, 1]
	cyan	Cyanosis	[-1, 0.5, 1]
	pal	Palpitations	[-1, 0.5, 1]
	weak	Weakness or fatigue	[-1, 0.5, 1]
	press	Discomfort, pressure, heaviness in the chest	[-1, 0.5, 1]
		Chest pain (precordial)	
	cp	region, posterior sternal)	[-1, 0.5, 1]
	ell	Edema of lower limb	[-1, 0.5, 1]
Inducement and history	kfoi	Known factors of	[-1,0,1]
		inducement	
	URI	Upper respiratory infection	[-1, 0.5, 1]
		Chronic pulmonary	
	cpd_his	diseased history	[-1, 0.5, 1]
	Hypertension family history		
hyp_his	Coronary heart diseases	[-1, 0.5, 1]	
	CHD_his	family history	[-1, 0.5, 1]
Physical sign and assistant examination	bp	Blood pressure	[0,1]
	cve	Cervical venous engorgement	[-1, 0.5, 1]
	bar_che	Barrel chest	[-1, 0.5, 1]
	sound_b	Sound of breath in pulmonary	[-1, 0.5, 1]
	Moi	Moist rales	[-1, 0.5, 1]
	sonant	Boundary of heart sonant	[-1, 0.5, 1]
	sm	Systolic murmur	[-1, 0.5, 1]
	dm	Diastolic murmur	[-1, 0.5, 1]
	s_thri	Systolic thrill	[-1, 0.5, 1]
	SHS	Second heart sound	[-1, 0.5, 1]
	LK_tend	Liver and kidney tenderness	[-1, 0.5, 1]
	Hyp_note	Hyperresonant note	[-1, 0.5, 1]
		Eminence in	
	APR	precordial region	[-1, 0.5, 1]
		Cardiac rhythm in	
	CRO	good order or not	[-1, 0.5, 1]
	P_wave	Pulmonic P-wave	[-1, 0.5, 1]
Neck_vr	Neck venous return	[0,1]	
C_enlar	Cardiac enlargement	[-1, 0.5, 1]	
ECG	Electrocardiogram	[-1, 0.5, 1]	
STT_alter	ST-T alteration	[-1, 0.5, 1]	

of CVD risk assessment tools in UAE by Oulhaj A [48] is used. Further descriptions of the dataset from where data was collected are given in Table 1. In our research study, the research study population or participants were selected from the UAE Hospitals. The medical record from the hospitals was used as a dataset from the UAE hospitals.

Data description: There are three types of input features:

- 1) Objective: facts-based information
- 2) Examination: medical investigation-based results
- 3) Subjective: patients-oriented information.

Following the clinic, pathological variables will be chosen for sampling the dataset, extracting experimentally on many patient cases. Table 2 shows the variables to be used in conducting this study, categorized into four groups, i.e., Basic Information, Symptoms, Inducement and Medical History, Physical signs and Assistant Examination. Variables selected for the study are given in the following table, which includes the symptoms and their abbreviations used in the software, which are calculated on 3 level scale, -1 equals none, 0.5 equals normal, and 1 equals higher intensity of symptoms.

Descriptive studies assess age, race, and medical history descriptive data, including frequencies, numbers, mediums (M), medium and regular difference (SD). The remaining work included utilizing a prediction model and a multilayer perceptron with a modern Google framework named TensorFlow and Python. We use simple linear regression (SLR) and multiple linear regression (MLR) for the analysis of our study.

B. METHODOLOGY

The architecture of our proposed method is illustrated in Figure 1. We collected the dataset from the UAE hospitals, on which we applied the data preprocessing. A preliminary data analysis is run to get the visual data insights from the datasets. A feature selection is done using principle component analysis (PCA) [49]. Our long short-term memory (LSTM) network is built on deep learning and trained on the UAE's hospital data. Finally, we use regression analysis to forecast the degree of overlap in the intensity of the symptoms.

The proposed framework consists of several modules. Initially, the health data from UAE hospitals is used to train our LSTM-based deep neural network (DNN) [50]. PCA is applied to select the most relevant features from the data, and SLR and MLR are applied for CVD predictions. We also evaluate our model using mean absolute error and compare different modules on CVD data.

1) FEATURE SELECTION

Feature selection using principal component analysis (PCA) [49] is a valuable method for analyzing CVD's data. PCA is a dimensionality reduction technique that identifies essential features in a dataset and reduces the number of features while preserving vital information.

In the context of CVD data, PCA is used to identify the most relevant clinical measurements that are associated with the development of CVD. These measurements include age, gender, blood pressure, cholesterol levels, smoking history, and medical history. To perform PCA-based feature selection on CVD data, we employed the following steps:

- Before applying PCA, we preprocessed the data by normalizing the features to have zero mean and unit variance. This step is critical to ensure that all features have equal weight in the PCA analysis.
- Next, we performed PCA on the preprocessed data. PCA identified the principal components and linear combinations of the original features that capture the most significant variations in the data.
- Once we have identified the principal components to retain, we use the loadings of each principal component to determine which original features are most important. Features with high loadings on a given principal component are more important for explaining the variation in that component.
- We selected a subset of features based on their importance as determined by the PCA analysis. This subset

can be used for subsequent analysis or modeling of CVD risk.

In general, the PCA-based feature selection is used for analyzing CVD data and identifying the most important features for predicting CVD risk. It helped to reduce the number of features and improve the interpretability of the data, which can be particularly useful in clinical settings where simplicity and ease of use are essential. However, ensuring that the selected feature subset is representative of the underlying biology of CVD and that the PCA analysis is carefully validated to ensure its reliability is crucial.

2) LSTM FOR CVD

LSTM RNNs eliminate gradient disappearance in regular RNNs. LSTMs represent sequential data with long-term dependencies well. The input gate, forget gate, cell state, and output gate comprise an LSTM cell. The forget gate removes previous cell state information, whereas the input gate adds fresh information. The cell state acts as a memory for the LSTM, and the output gate controls how much of it is sent to the next time step.

The equations for computing the values of the input gate, forget gate, cell state, and output gate at time step t are as follows:

Input gate:

$$i_t = \text{sigmoid}(W_i \times [h_{t-1}, x_t] + b_i) \quad (1)$$

Forget gate:

$$f_t = \text{sigmoid}(W_f \times [h_{t-1}, x_t] + b_f) \quad (2)$$

Cell state:

$$c_t = f_t \times c_{t-1} + i_t \times \tanh(W_c \times [h_{t-1}, x_t] + b_c) \quad (3)$$

Output gate:

$$o_t = \text{sigmoid}(W_o \times [h_{t-1}, x_t] + b_o) \quad (4)$$

$$h_t = o_t * \tanh(c_t) \quad (5)$$

where i_t , f_t , and o_t are the input, forget, and output gates at time step t , respectively. c_t is the cell state at time step t . h_t is the hidden state at time step t . x_t is the input at time step t . W_i , W_f , W_c , and W_o are weight matrices for the input, forget, cell, and output gates, respectively. b_i , b_f , b_c , and b_o are bias vectors for the input, forget, cell, and output gates. sigmoid is the sigmoid activation function. tanh is the hyperbolic tangent activation function.

The LSTM architecture uses a set of equations to compute the values of the input gate, forget gate, cell state, and output gate at each time step. These equations enable the LSTM to learn the temporal dependencies between patients' medical records, such as their age, symptoms, and medical history. The LSTM is trained on a large dataset of patient records using a regression analysis approach, where the input is a sequence of medical records, and the output predicts the likelihood of developing cardiovascular disease (CVD). The LSTM can be trained on various patient attributes and clinical

data, such as age, gender, blood pressure, cholesterol levels, smoking history, and medical history, to predict the likelihood of CVD.

We use the trained LSTM to predict new patients' likelihood of developing CVD by taking in the patient's medical records sequence. By leveraging the temporal dependencies in patient health records, LSTM-based deep neural networks can potentially identify patients at high risk of developing CVD and intervene with preventative measures before the disease progresses. However, it is crucial to ensure that the data used to train the model is representative of the patient population and that the model is thoroughly validated before deployment in a clinical setting.

3) SIMPLE LINEAR REGRESSION (SLR)

Simple Linear Regression is a statistical method used to study the relationship between a dependent variable and one independent variable. It is called "simple" because it only involves one independent variable. The method is used to predict the value of the dependent variable based on the value of the independent variable.

The equation for a simple linear regression model is:

$$y = b_0 + b_1 \times x \quad (6)$$

where y is the dependent variable, x is the independent variable, b_0 is the y-intercept, and b_1 is the slope of the line. The goal of simple linear regression is to find the values of b_0 and b_1 that best fit the data.

SLR models are fitted using least squares. Least squares minimizes the squared disparities between anticipated and actual dependent variable values. "Regression line" is the best fit. SLR helps predict and comprehend two variables. It's limited when the variables' relationships are increasingly complicated. Multivariate linear regression may be better in such circumstances.

4) MULTIPLE LINEAR REGRESSION (MLR)

Multiple Linear Regression examines dependent-independent relationships. MLR may assess a patient's symptoms, risk factors, and other data to uncover patterns and associations that can be utilized to predict CVDs. An MLR model may include age, blood pressure, cholesterol, smoking history, and family history of cardiovascular disease to predict heart attack and stroke risk.

The equation for an MLR is as follows.

$$y = b_0 + b_1 \times x_1 + b_2 \times x_2 + \dots + b_n \times x_n \quad (7)$$

where y is the dependent variable (e.g. risk of cardiovascular disease), x_1, x_2, \dots, x_n are the independent variables (e.g. age, blood pressure, etc.), and $b_0, b_1, b_2, \dots, b_n$ are the coefficients of the model that are estimated from the data.

MLR is a powerful tool for predicting CVDs, as it allows for the consideration of multiple variables and their interactions. However, it is important to note that multiple linear regression assumes that the relationship between the independent and dependent variables is linear, and that there is

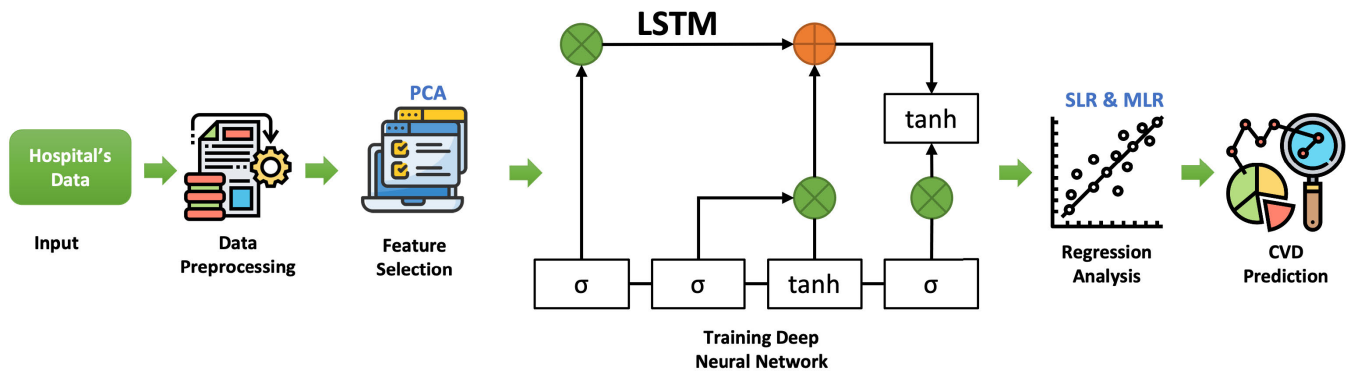


FIGURE 1. The proposed framework of our study, in which we input the hospital's data and apply the data preprocessing to it. After that, feature selection using the PCA is applied. We then train the LSTM and apply regression analysis on the data to predict CVDs.

TABLE 3. SLR findings of our study.

Predictor	R-Square	Standardized Beta	Significance
EMPTY	0.643	0.802	0.000
DYSP	0.699	0.836	0.000
FIVE	0.149	-0.386	0.000
HEADACHE	0.029	0.171	0.000
CYAN	0.406	-0.637	0.000
WFAT	0.054	0.232	0.002
DPCH	0.299	0.547	0.000
CHEP	0.715	0.846	0.000

no multicollinearity among the independent variable and also require large sample size. In cases where the relationship is non-linear or when the sample size is small, other methods such as polynomial regression or decision tree analysis may be more appropriate.

IV. EXPERIMENTAL EVALUATION

A. MODEL AND ASSUMPTIONS

First, each of the eight potential predictors that cause data leaks should be seen as a simple linear regression (SLR). After that, the researcher performed multiple linear regressions (MLR) to obtain usability variables predictions. This model will evaluate and identify the usability criteria. In both regression analyses, the SPSS software was used (version 25.0). SLR findings are shown in Table 3. As per the questionnaire, we have evaluated data in the below terms: Chest Pain, Cyanosis, Emptytysis, Fever, Fatigue, Discomfort, Pressure in the Chest, Weakness, and how users rated them in terms of Heart issues. These are abbreviated CHEP as Chest Pain, dyspepsia as DYSP, Fever defined as FEVE, Headache as HEADACHE, Cyanosis defined as CYAN, Weakness and Fatigue as WFAT, fulfillment as DPCH, and the definition of success are CHEP.

Following is the formulation of SLR:

$$\bar{Y} = 0.347 + (0.846 * CHEP) \quad (8)$$

Disease prediction is that Chest Pain has been predicted with an 84.6% accuracy level.

TABLE 4. Findings of MLR-based models.

Sr. No.	Predictors	R-Square
1	CHEP	0.715
2	CHEP, DYSP	0.844
3	DYSP, CYAN, and CHEP	0.867
4	DYSP, CYAN, CHEP and WFAT	0.889
5	EMPT, DYSP, CYAN, WFAT and CHEP	0.898
6	EMPT, DYSP, CYAN, WFAT, DPCH and CHEP	0.906
7	EMPT, DYSP, FEVE, CYAN, WFAT, DPCH and CHEP	0.91

MLR Analysis of Symptoms: The dataset included 2621 regression instances in which the model was a predictor (eight different usability factors). Shortening the gap between predictor responses is the driving force behind the multi-linear method used in stepping regression.

This method incorporates the dependent variable with various independent variables, such as two factors IV, three factors IV, four factors IV, and so on, to create a more complex model. Combining all 0.000 statistically significant predictors, the R^2 value is computed. The tables below in the next section provide the MLR results for different combinations of predictors. A better model has a higher R^2 value.

After applying MLR, the researcher got findings for all combinations of predictor variables. A secure approach where data breaches are impossible has been proposed as a result of this. Increase the number of variables gradually until the optimal model is obtained. Table 4 lists the most effective regression models that produced significant findings.

In the MLR model with seven predictors, the larger value of the R-Square is 0.911, as shown at Sr. No. 7. The R-Square, on the other hand, shows just a 0.001 rise when the eight variables are included in the regression model, but the R-Square shows a higher change in prior models. As a result, the final model has seven variables excluding headaches.

R-Square: The percentage of the dependent variable's variation that can be predicted from the independent variable is expressed statistically by the R-square (R^2). It serves as a gauge for how strongly the two variables are related. Higher numbers suggest a stronger association. R-squared values vary from 0 to 1. R-squared is the square of the Pearson correlation coefficient between the dataset's observed responses and the model-predicted responses in a linear regression

model. It may be used to assess the model’s quality of fit. Following is the formulation of R^2 .

$$R^2 = \frac{V_{n=1}^N \hat{y}_n}{V_{n=1}^N y_n} \tag{9}$$

where $V_{n=1}^N \hat{y}_n$ is the residual variance and y is the outcome.

B. COMPARISON

Using elements refined by Delphi methodologies, an optimal hybrid model with eight usability aspects was determined in this study, and they are as follows: The independent variables/factors in the hybrid model that influenced the dependent variable’s rating regressively were empyasis, learnability, fever, cyanosis, weakness or fatigue, chest discomfort pressure, and chest pain. The next section compares our suggested RUF (Rating of Usability Factor) to the HYBRID model. An eight-factor Hybrid Model Rating, a RUF Model, and the MLR and SLR techniques are used to finish the RUF Model.

Users’ ratings of the hybrid model were shown to be non-significantly predicted by the model’s R^2 value, indicating that it is responsible for 71.5% of the variation in user ratings R. Following is the regression formulation:

$$\bar{Y} = 0.347 + (0.846 * CHEP) \tag{10}$$

The predicted value of user rating (dependent variable) is denoted by \bar{Y} , whereas the rating provided by the hybrid model is represented by OR (independent variable). Regarding forecasting user rating apps’ usability characteristics, the RUF Model (based on 7 elements) performs better than other models, according to the coefficient of determination. According to the seven usability criteria, RUF explained 91% of the variance ($R^2 = 0.910$), whereas Hybrid explained 91% of the variance ($R^2 = 0.911$).

C. MODEL ASSESSMENT AND VALIDATION

In addition to summarizing k-fold models with R-square values, this part gives a thorough explanation of how to evaluate model values using K-fold cross-validation using PRED(X) and MMRE.

1) MODEL ASSESSMENT

To attain the remarkable precision of the forecast model, de facto average prediction accuracy criteria are employed. MMRE and PRED(x) are frequently employed for accuracy verification. in the written word. PRED value is more than or equal to 0.75 (0.25). When the MMRE value exceeds 0.25, it will be discarded and treated as PRED (X). The MMRE value was discovered to be 0.0515.

2) VALIDATION METHOD

The method of model validation known as K-fold cross-validation is not comprehensive. The effectiveness of the projected model on an independent data set is used in this method. K-fold cross-validation is a technique for assessing how well a machine learning model performs on a dataset.

TABLE 5. Summary of K-fold cross-validation.

Folds	R^2	MMRE
1	0.914	0.1523
2	0.915	0.0705
3	0.908	0.0486
4	0.905	0.0471
5	0.911	0.0509
6	0.914	0.0451
7	0.909	0.0408
8	0.91	0.0558

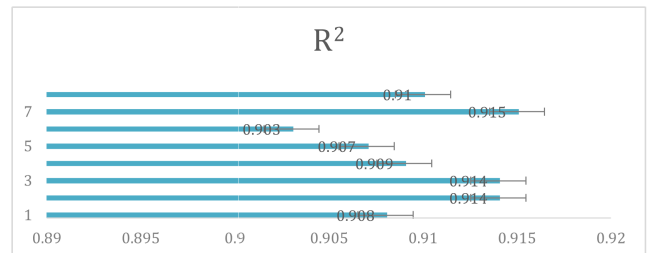


FIGURE 2. Predicted R-square.

The dataset is randomly divided into “k” folds or subsets, and the model is trained and evaluated k times, each time. The model’s effectiveness on the dataset is then estimated using the average performance over all k iterations.

K-fold cross-validation might be used to assess how well a machine learning model-like a regression model-performed on a dataset of patient medical records when it came to forecasting CVDs. It is feasible to obtain a more reliable assessment of the model’s performance and prevent overfitting to the training set by utilizing k-fold cross-validation. Choosing an acceptable value for k is crucial when employing k-fold cross-validation. K=10 is a popular choice, however depending on the situation, K=5 or K=20 may also be utilized.

Depending on the quantity of the data, the dataset is divided into the same number of folds. By k-fold cross-validation, data calibration and validation double their data points. All folds’ data values are split into 10 predetermined categories. Each category has a total of 21 times throughout each fold, ranging from 1 to 21 and 22 to 42. K = 1 to K = 10 for fold.

Each fold of the data is shown to be significant even after being divided into eight equal parts, demonstrating the accuracy and suitability of our dataset. We found that every single one of its values in the K fold forecast are significant values. If the value after training is 0.05, we observed that all six predictors are substantial. PRED (0.25) is computed in each fold, it should be noticed. Table 5 contains a validation summary and each fold’s R^2 MMRE values. All MMRE and R^2 values have validated the suggested model. Figure 2 shows the results of a final prediction made using the data and R^2 .

D. HYBRID MODEL

When the MLR is applied to all eight components, the R-Square only rises by 0.1 percent, but the but condition is no

longer crucial. The sixth model is the greatest choice since the half-breed design does not provide a practical deep-learning procedure to safeguard a patient from heart failure.

We discover and demonstrate that all of the K-FOLDS are 0-21 by using MRE to the remaining 21-147 K-folds with Constants of 0-21 and finishing the comparable with irregular data. In this way, we were able to isolate the data. Just one of the 336 MRE esteems had a Pred(X) of 0.25 or less, yet MMREs fit every one of them, demonstrating the accuracy of our data. Our data show that our ordered mean relative error (MMRE) values are 0.1523, 0.0705, 0.0486, 0.0471, 0.0509, 0.0451, 0.0408 and 0.0558; some values are less than 0.05; we used grand MRRE and discovered 0.05, indicating that our dataset is free of mean relative error. Moreover, MRE values greater than 0.25 are regarded as Pred(X) values indicating that the Rating of Usability Factor (RUF) model is correct and ideal.

E. MODEL ACCURACY

We discuss the accuracy measure metric we consider for CVD prediction. We also present results without and with time-series approaches. In the end, we describe the quantization technique results in terms of accuracy and size reduction for CVD data of UAE hospitals.

Our framework's prediction accuracy is measured by Mean Absolute Error (MAE). MAE measures the average absolute difference between expected and actual values. MAE ignores outliers since it doesn't punish huge mistakes. Lower values improve model performance. The following equation calculates MAE. 11.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i| \quad (11)$$

We evaluate our model based on the aforementioned metric for the prediction of overlapping symptoms on unseen test data. We extracted most latest samples from the dataset to make test data. The test data was not used during the training and validation of the model.

Figure 3 shows the MAE values for comparing single-input and multiple-input-time models following the time-series approach. The single-input-time dense model showed better accuracy than the single-input-time linear model, it can be because of additional layers and neurons. However, multi-input-time dense and convolutional-based model results were less accurate than single-input-time models. On the other hand, the LSTM model exhibited the highest accuracy compared to the above. This can be attributed to the LSTM memory capability of retaining information.

Figure 4 compares single-shot models that predicted multiple time steps. The accuracy gradually increases, going from linear to more advanced models. As compared to single-output models, the accuracy has moderately decreased. This is as expected since the number of time steps ahead in the predictions has increased.

Table 6 shows a performance comparison regarding MSE on test data. It can be observed that in this particular problem,

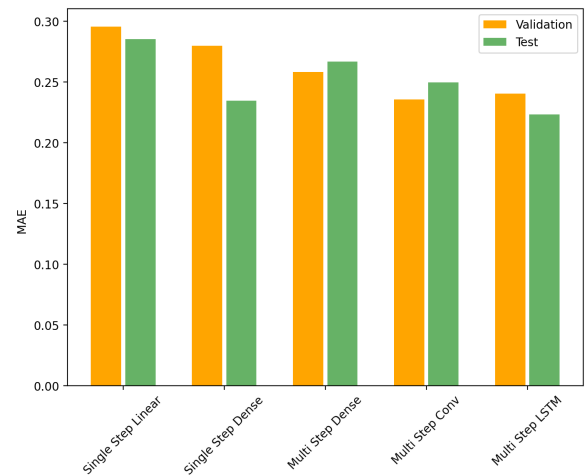


FIGURE 3. Performance of models based on single and multi-input time steps.

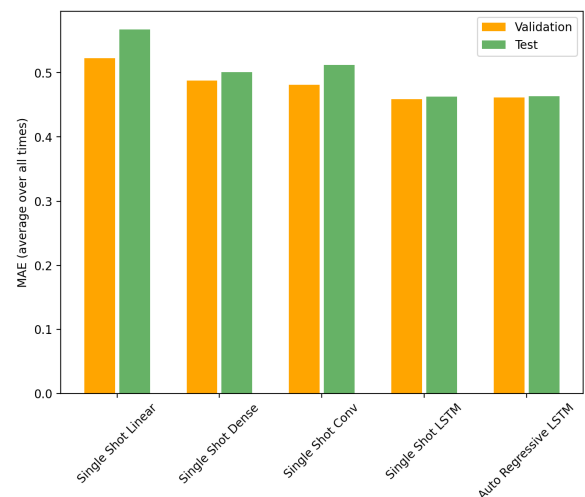


FIGURE 4. Comparison of single shot and autoregressive models for CVD prediction.

TABLE 6. Models accuracy before and after compression.

Model	Loss/MSE
DNN model	0.2219
DNN model with LSTM	0.2214
DNN model for Healthcare	0.2220
DNN model with LSTM for CVD	0.2484

quantization has almost no effect on accuracy in terms of MSE.

V. DISCUSSION

We provide a brief analysis of our work related to CVD carried out from deep learning-based regression techniques and which have been rated by early model prediction as above in terms of overlapping and disease prediction with accuracy level. According to the questionnaire provided against our factors/predictors, most of the individual users of these applications in their regular daily existences have made many

mistakes from a security point of view. The specialist surveyed how data could be secured from breaching.

The Hybrid ideal model was used with SLR representing up to 8 elements: (i) CHEP (ii) DPCH (iii) DYSP (iv) FEVE (v) Headache (vi) EMPT (vii) WFAT (viii) CYAN. All of the traits were determined to be significant and to have strong R-squared values. In the mixed model, EMPT accounts for 64 percent of the change, while DYSP accounts for 69.9 percent, FEVE for 14%, HEADACHE for 2%, CYAN for 40%, WFAT for 5%, DPCH for 29%, and CHEP for 71.5% protect our sensitive characteristics from overlapping symptoms. Deep learning users who had a heart attack or heart failure determined that the most effective overlap symptoms factors are EMPT, DDYSP, CYAN, and CHEP. The scientist then made the decision to build this using multiple linear regression by making sets of these components dependent on every feasible configuration of a secure modeling model. The formation of twofold pairings resulted in the discovery of 21 out of 29 sets as important, with R-Square values in the good range and sig 0.000 R-Squares. Once again, DYSP and CHEP have the greatest R-Squares (0.844), which account for 84.4 percent of the difference in results and show which aspects are most crucial for user safety.

Crossover uses MLR. DYSP, CHEP, and CYAN explain 86.7% of model variance. CYAN increased the model's R Square. Four and five regressions utilized MLR. R-Square is 0.889 in the fourth matched relapse, expanding 88.9% of the equation. The fifth R-Square is 0.898 and explains 0.9 percent more than the fourth matched relapse. WFAT and Productive improved model explainability.

Furthermore, it has been shown that the number of essential blends decreases when the matching approach is enlarged. For example, one set of EMPT, DYSP, FEVE, CYAN, WFAT, DPCH, and CHEP is essential in the seventh combined relapse. As a result, the model's R-Square is enlarged to 0.910, making it more straightforward than the previous one.

In conclusion, regression analysis is a powerful tool that can be leveraged to predict overlapping symptoms of CVDs. By analyzing data on patient symptoms, risk factors, and disease outcomes, regression models can identify patterns and relationships that can be used to accurately predict which patients are at risk for specific CVDs. This can help doctors and healthcare providers to more quickly and effectively diagnose and treat these conditions, ultimately improving patient outcomes and reducing the burden on the healthcare system.

VI. CONCLUSION

With more than 8000 fatalities annually, CVDs are a significant cause of mortality and disability in the United Arab Emirates. Since CVD may exhibit a wide variety of symptoms, it can be challenging to get a precise diagnosis. Researchers have created a prediction model that uses regression analysis to foretell overlapping symptoms of CVDs to solve this problem.

To identify individuals at risk of CVDs, our suggested model combines self-reported symptoms, lifestyle variables,

and test data. Regression analysis is used to find trends in the data that may be utilized to forecast the risk that a patient would develop CVD. Additionally, the model may be used to evaluate the relative hazards connected to various lifestyle choices, such as smoking and obesity.

The proposed model in this work is reliable in predicting the development of CVDs in patients with comparable symptoms after being validated using a large dataset of patients with known CVDs. The model may also be used to identify individuals who need closer monitoring since they have a greater risk of acquiring CVDs. Our suggested prediction model is a crucial tool for bettering the diagnosis and treatment of CVDs and can potentially lower mortality and morbidity associated with CVDs.

ACKNOWLEDGMENT

This work was supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R384), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

REFERENCES

- [1] L. D. Weberling, D. Lossnitzer, N. Frey, and F. André, "Coronary computed tomography vs. cardiac magnetic resonance imaging in the evaluation of coronary artery disease," *Diagnostics*, vol. 13, no. 1, p. 125, Dec. 2022.
- [2] M. C. Badescu, O. V. Badulescu, L. I. Butnariu, I. B. Bojan, M. C. Vladeanu, N. Dima, C. E. Vlad, L. G. Foia, M. Ciocoiu, and C. Rezus, "Cardiovascular risk factors in patients with congenital hemophilia: A focus on hypertension," *Diagnostics*, vol. 12, no. 12, p. 2937, Nov. 2022.
- [3] M. Swathy and K. Saruladha, "A comparative study of classification and prediction of cardio-vascular diseases (CVD) using machine learning and deep learning techniques," *ICT Exp.*, vol. 8, no. 1, pp. 109–116, Mar. 2022.
- [4] M. A. Eid, K. Mehta, J. A. Barnes, Z. Wanken, J. A. Columbo, D. H. Stone, P. Goodney, and M. M. Smith, "The global burden of peripheral artery disease," *J. Vascular Surg.*, vol. 77, no. 4, pp. 1119–1126, Apr. 2023.
- [5] N. Culliford-Semmens, E. Tilton, N. Wilson, J. Stirling, R. Doughty, T. Gentles, B. Peat, E. Dimalapang, and R. Webb, "Echocardiography for latent rheumatic heart disease in first degree relatives of children with acute rheumatic fever: Implications for active case finding in family members," *eClinicalMedicine*, vol. 37, Jul. 2021, Art. no. 100935.
- [6] R. O. Escárcega, M. J. Lipinski, M. García-Carrasco, C. Mendoza-Pinto, J. L. Galvez-Romero, and R. Cervera, "Inflammation and atherosclerosis: Cardiovascular evaluation in patients with autoimmune diseases," *Autoimmunity Rev.*, vol. 17, no. 7, pp. 703–708, Jul. 2018.
- [7] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert, "Deep learning for cardiac image segmentation: A review," *Frontiers Cardiovascular Med.*, vol. 7, p. 25, Mar. 2020.
- [8] Y. Ma, M. U. Hassan, D. Niu, and L. Wang, "The segmentation of glandular cavity based on K-means and mathematical morphology," in *Proc. 4th Int. Conf. Syst. Informat. (ICSAI)*, Nov. 2017, pp. 1287–1291.
- [9] V. Chang, V. R. Bhavani, A. Q. Xu, and M. Hossain, "An artificial intelligence model for heart disease detection using machine learning algorithms," *Healthcare Anal.*, vol. 2, Nov. 2022, Art. no. 100016.
- [10] P. Mathur, S. Srivastava, X. Xu, and J. L. Mehta, "Artificial intelligence, machine learning, and cardiovascular disease," *Clin. Med. Insights, Cardiol.*, vol. 14, Jan. 2020, Art. no. 117954682092740.
- [11] S. Hong, Y. Zhou, J. Shang, C. Xiao, and J. Sun, "Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review," *Comput. Biol. Med.*, vol. 122, Jul. 2020, Art. no. 103801.
- [12] S. Suganyadevi, V. Seethalakshmi, and K. Balasamy, "A review on deep learning in medical image analysis," *Int. J. Multimedia Inf. Retr.*, vol. 11, no. 1, pp. 19–38, Mar. 2022.
- [13] M. U. Hassan, S. Alaliyat, and I. A. Hameed, "Image generation models from scene graphs and layouts: A comparative analysis," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 35, no. 5, May 2023, Art. no. 101543.

- [14] O. Ghorbanzadeh, T. Blaschke, K. Gholamnia, S. Meena, D. Tiede, and J. Aryal, "Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection," *Remote Sens.*, vol. 11, no. 2, p. 196, Jan. 2019.
- [15] Y. Ma, J. Zhang, D. Niu, M. U. Hassan, and X. Zhao, "An unsupervised approach for 3D medical image registration," in *Proc. 2019 7th Int. Conf. Inf. Technol., IoT Smart City*, 2019, pp. 259–263.
- [16] K. K. L. Wong, G. Fortino, and D. Abbott, "Deep learning-based cardiovascular image diagnosis: A promising challenge," *Future Gener. Comput. Syst.*, vol. 110, pp. 802–811, Sep. 2020.
- [17] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," *IEEE Access*, vol. 5, pp. 8869–8879, 2017.
- [18] K. Matsushita, N. Ding, M. Kou, X. Hu, M. Chen, Y. Gao, Y. Honda, D. Zhao, D. Dowdy, Y. Mok, J. Ishigami, and L. J. Appel, "The relationship of COVID-19 severity with cardiovascular disease and its traditional risk factors: A systematic review and meta-analysis," *Global Heart*, vol. 15, no. 1, p. 64, Sep. 2020.
- [19] K. C. Siontis, P. A. Noseworthy, Z. I. Attia, and P. A. Friedman, "Artificial intelligence-enhanced electrocardiography in cardiovascular disease management," *Nature Rev. Cardiol.*, vol. 18, no. 7, pp. 465–478, Jul. 2021.
- [20] M. Nii, K. Maeda, H. Wakabayashi, S. Nishioka, and A. Tanaka, "Nutritional improvement and energy intake are associated with functional recovery in patients after cerebrovascular disorders," *J. Stroke Cerebrovascular Diseases*, vol. 25, no. 1, pp. 57–62, Jan. 2016.
- [21] M. M. H. Abdullah, J. Hughes, and S. Grafenauer, "Legume intake is associated with potential savings in coronary heart disease-related health care costs in Australia," *Nutrients*, vol. 14, no. 14, p. 2912, Jul. 2022.
- [22] M. Ullah, A. Wahab, S. U. Khan, U. Zaman, K. U. Rehman, S. Hamayun, M. Naeem, H. Ali, T. Riaz, S. Saeed, A. M. Alsuhaibani, and M. S. Refat, "Stent as a novel technology for coronary artery disease and their clinical manifestation," *Current Problems Cardiol.*, vol. 48, no. 1, Jan. 2023, Art. no. 101415.
- [23] C. Gutierrez and D. G. Blanchard, "Diagnosis and treatment of atrial fibrillation," *Amer. Family Physician*, vol. 94, no. 6, pp. 442–452, 2016.
- [24] M. Noale, F. Limongi, and S. Maggi, "Epidemiology of cardiovascular diseases in the elderly," *Frailty and Cardiovascular Diseases: Research into an Elderly Population*, 2020, pp. 29–38.
- [25] M. Athanasiou, K. Srintzeri, K. Zarkogianni, A. C. Thanopoulou, and K. S. Nikita, "An explainable XGBoost-based approach towards assessing the risk of cardiovascular disease in patients with type 2 diabetes mellitus," in *Proc. IEEE 20th Int. Conf. Bioinf. Bioengineering (BIBE)*, Oct. 2020, pp. 859–864.
- [26] P. H. Charlton, P. A. Kyriacou, J. Mant, V. Marozas, P. Chowienzyk, and J. Alastruey, "Wearable photoplethysmography for cardiovascular monitoring," *Proc. IEEE*, vol. 110, no. 3, pp. 355–381, Mar. 2022.
- [27] D. Chieng and P. M. Kistler, "Coffee and tea on cardiovascular disease (CVD) prevention," *Trends Cardiovascular Med.*, vol. 32, no. 7, pp. 399–405, Oct. 2022.
- [28] L.-C. Tao, J.-N. Xu, T.-T. Wang, F. Hua, and J.-J. Li, "Triglyceride-glucose index as a marker in cardiovascular diseases: Landscape and limitations," *Cardiovascular Diabetol.*, vol. 21, no. 1, pp. 1–17, Dec. 2022.
- [29] G. Battineni, G. G. Sagaró, N. Chintalapudi, and F. Amenta, "The benefits of telemedicine in personalized prevention of cardiovascular diseases (CVD): A systematic review," *J. Personalized Med.*, vol. 11, no. 7, p. 658, Jul. 2021.
- [30] H. E. Bays, P. R. Taub, E. Epstein, E. D. Michos, R. A. Ferraro, A. L. Bailey, H. M. Kelli, K. C. Ferdinand, M. R. Echols, H. Weintraub, and J. Bostrom, "Ten things to know about ten cardiovascular disease risk factors," *Amer. J. Preventive Cardiol.*, vol. 5, Mar. 2021, Art. no. 100149.
- [31] V. V. Dickson, J. Jun, and G. D. Melkus, "A mixed methods study describing the self-care practices in an older working population with cardiovascular disease (CVD): Balancing work, life and health," *Heart Lung*, vol. 50, no. 3, pp. 447–454, May 2021.
- [32] G. K. Ellis, J. A. Robinson, and G. B. Crawford, "When symptoms of disease overlap with symptoms of depression," *Austral. Family Physician*, vol. 35, no. 8, pp. 647–649, 2006.
- [33] N. F. A. Alhadeethy, A. M. Zeki, and A. Shah, "Deep learning model for predicting and detecting overlapping symptoms of cardiovascular diseases in hospitals of UAE," *Turkish J. Comput. Math. Educ. (TURCOMAT)*, vol. 12, no. 14, pp. 5212–5224, 2021.
- [34] C.-S. Hsu, S.-H. Wen, J.-S. Hung, T.-T. Liu, C.-H. Yi, W.-Y. Lei, F. Pace, and C.-L. Chen, "Overlap of dyspepsia in patients with gastroesophageal reflux disease: Impact of clinical, metabolic, and psychosocial characteristics," *Digestive Diseases Sci.*, vol. 62, no. 4, pp. 994–1001, Apr. 2017.
- [35] A. Chaddha, E. A. Robinson, E. Kline-Rogers, T. Alexandris-Souphis, and M. Rubenfire, "Mental health and cardiovascular disease," *Amer. J. Med.*, vol. 129, no. 11, pp. 1145–1148, 2016.
- [36] G. M. Goodwin, "Depression and associated physical diseases and symptoms," *Dialogues Clin. Neurosci.*, vol. 8, no. 2, pp. 259–265, Jun. 2006.
- [37] A. Daoulah, A. Elfarnawany, T. Al Garni, A. S. Hersi, M. Alshehri, W. Almahmeed, N. Yousif, R. Abuelatta, A. Alasmari, and N. E. Elsheikh-Mohamed, "Outcomes of myocardial revascularization in diabetic patients with left main coronary artery disease: A multicenter observational study from three Gulf countries," *Cardiovascular Revascularization Med.*, vol. 46, pp. 52–61, Jan. 2023.
- [38] M. U. Hassan, D. Niu, M. Zhang, and X. Zhao, "Asymmetric hashing based on generative adversarial network," *Multimedia Tools Appl.*, vol. 82, no. 1, pp. 389–405, Jan. 2023.
- [39] L. Yang, H. Wu, X. Jin, P. Zheng, S. Hu, X. Xu, W. Yu, and J. Yan, "Study of cardiovascular disease prediction model based on random forest in Eastern China," *Sci. Rep.*, vol. 10, no. 1, p. 5245, Mar. 2020.
- [40] P. Ghosh, S. Azam, M. Jonkman, A. Karim, F. M. J. M. Shamrat, E. Ignatious, S. Shultana, A. R. Beeravolu, and F. De Boer, "Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques," *IEEE Access*, vol. 9, pp. 19304–19326, 2021.
- [41] H. R. H. Al-Absi, M. T. Islam, M. A. Refaee, M. E. H. Chowdhury, and T. Alam, "Cardiovascular disease diagnosis from DXA scan and retinal images using deep learning," *Sensors*, vol. 22, no. 12, p. 4310, Jun. 2022.
- [42] N. K. Kumar, G. S. Sindhu, D. K. Prashanthi, and A. S. Sulthana, "Analysis and prediction of cardio vascular disease using machine learning classifiers," in *Proc. 6th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Mar. 2020, pp. 15–21.
- [43] A. S. Abdalrada, J. Abawajy, T. Al-Quraishi, and S. M. S. Islam, "Machine learning models for prediction of co-occurrence of diabetes and cardiovascular diseases: A retrospective cohort study," *J. Diabetes Metabolic Disorders*, vol. 21, no. 1, pp. 251–261, Jan. 2022.
- [44] O. T. Kee, H. Harun, N. Mustafa, N. A. A. Murad, S. F. Chin, R. Jaafar, and N. Abdullah, "Cardiovascular complications in a diabetes prediction model using machine learning: A systematic review," *Cardiovascular Diabetol.*, vol. 22, no. 1, pp. 1–10, Jan. 2023.
- [45] V. Venkat, H. Abdelhalim, W. DeGroat, S. Zeeshan, and Z. Ahmed, "Investigating genes associated with heart failure, atrial fibrillation, and other cardiovascular diseases, and predicting disease using machine learning techniques for translational research and precision medicine," *Genomics*, vol. 115, no. 2, Mar. 2023, Art. no. 110584.
- [46] Y. Wang, Y. Fan, P. Bhatt, and C. Davatzikos, "High-dimensional pattern regression using machine learning: From medical images to continuous clinical variables," *NeuroImage*, vol. 50, no. 4, pp. 1519–1535, May 2010.
- [47] M. Mudassir, S. Bennbaia, D. Unal, and M. Hammoudeh, "Time-series forecasting of Bitcoin prices using high-dimensional features: A machine learning approach," *Neural Comput. Appl.*, pp. 1–15, Jul. 2020.
- [48] A. Oulhaj, S. Bakir, F. Aziz, A. Suliman, W. Almahmeed, H. Sourij, and A. Shehab, "Agreement between cardiovascular disease risk assessment tools: An application to the United Arab Emirates population," *PLoS ONE*, vol. 15, no. 1, Jan. 2020, Art. no. e0228031.
- [49] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433–459, 2010.
- [50] T. Santosh, D. Ramesh, and D. Reddy, "LSTM based prediction of malaria abundances using big data," *Comput. Biol. Med.*, vol. 124, Sep. 2020, Art. no. 103859.



SARA GHORASHI received the B.Sc. and M.Sc. degrees in computer science from the University of Khartoum, and the Ph.D. degree in computer science from the Sudan University of Science and Technology, in 2016. She is currently an Assistant Professor with the Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia. Her research interests include artificial intelligence, machine learning, data mining, network security, and other related topics.



KHUNSA REHMAN received the M.B.B.S. degree from the Services Institute of Medical Sciences, Lahore, Pakistan, in 2014. She did a house job in medicine surgery, ENT, and GYNAE OBS, in 2015. She cleared FCPS Part 1, in August 2022. From 2015 to 2018, she was a Women Medical Officer in the basic health unit of Bhuman Shah. She was with Mansoorah Hospital, from 2018 to 2020. Currently, she is a Women Medical Officer with Social Security Hospital,

Lahore. Her research interests include the diagnosis, prevention, and treatment of diseases that includes biological and medical treatment techniques using the study of plasma sources and fabrication of micro or nanostructured surfaces.

ANAM RIAZ received the M.B.B.S. degree from Rawalpindi Medical University, Rawalpindi, Pakistan. She is currently a Women Medical Officer with Government Headquarters Hospital, Hafizabad, Pakistan. Her research interests include medical imaging, cardiovascular disease prediction, and deep learning.

HEND KHALID ALKAHTANI (Member, IEEE) received the B.Sc. degree and the M.Sc. degree in information management from The George Washington University, USA, in 1992 and 1993, respectively, and the Ph.D. degree in information security from Loughborough University, in 2018. She is currently an Assistant Professor with the College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Saudi Arabia. She has 23 years of work experience as a lecturer, a computer center president, and a statistic center president in faculty colleges. She received an award from the SIDF Academy: Leading Creative Transformation in Critical Time Program, Stanford University, Center for Professional Development. She has two conference publications in which she received a certificate as the best publication presented.



AHMED H. SAMAK received the M.Sc. and Ph.D. degrees in computer science from Menofia University, Egypt, in 2004 and 2011, respectively. He is currently an Assistant Professor with the University of Bisha, Saudi Arabia. He has published several papers in ISI journals and international conferences. His current research interests include computer vision, steganography, image processing, and the IoT.



IVAN CHERREZ-OJEDA received the M.D., M.Sc., and Ph.D. degrees. He is currently a Professor in allergy, immunology and pulmonary medicine with the Department of Allergy, Immunology and Pulmonology, the Medical and Research Director of the Respiralab Research Group, and a Research Professor with Espiritu Santo University (UEES), Grupo Hospitalario Clínica Kennedy Sección Delta, Ecuador. He has a soundtrack of publishing around medical domains

in several high-level venues.



AMNA PARVEEN received the Master of Pharmacology degree from Government College University Faisalabad, Faisalabad, Pakistan, in 2013, and the Ph.D. degree in pharmacognosy from Chung-Ang University, South Korea, in 2016.

She was as a Lecturer with the College of Pharmacy, Government College University Faisalabad (2009–2013). After completing her Ph.D. degree, she started her career as a Researcher with the College of Pharmacy, Gachon University, South Korea, where she is currently an Assistant Professor. She is an expert in the field of natural product chemistry, particularly in the discovery of bioactive constituents from medicinal plants against pharmacological conditions. She has published several research papers in different national and international journals. Her research interests include isolating and identifying new constituents from traditionally used medicinal plants that may have potential against various pharmacological conditions, such as cancer, diabetes, aging, and neurodegenerative diseases. She also has an ongoing interest in identifying bioactive constituents from natural resources for aging and life span extension.

...