Deep learning Approach for Cardiovascular Disease Risk Stratification and Survival Analysis on a Canadian Cohort

Mrinalini Bhagawati[1], Sudip Paul[1], Laura Mantella[2], Amer M. Johri[3], John R. Laird[4], Inder M. Singh[5], Rajesh Singh[6], Deepak Garg[7], Mostafa M Fouda[8], Narendra N. Khanna[9], Riccardo Cau[10], Ajith Abraham[11], Mostafa Al-Maini[12], Esma R. Isenovic[13], Aditya M. Sharma[14], Jose Fernandes E. Fernandes[15], Seemant Chaturvedi[16], Mannudeep K. Karla[17], Andrew Nicolaides[18], Luca Saba[10], Jasjit S. Suri[5, 8, 19]*

[1]Department of Biomedical Engineering, North-Eastern Hill University, Shillong, INDIA.
[2]Division of Cardiology, Department of Medicine, University of Toronto, Toronto, Canada.
[3]Division of Cardiology, Department of Medicine, Queen's University, Kingston, Canada.
[4]Heart and Vascular Institute, Adventist Health St. Helena, St Helena, CA, 94574, USA.
[5]Stroke Diagnostic and Monitoring Division, AtheroPoint™, Roseville, CA, 95661, USA.
[6]Division of Research and Innovation, UTI, Uttaranchal University, Dehradun, INDIA.
[7]School of Cowereter Science and Artificial Intelligence, SR University, Warangal, Telangana, 506371, INDIA.
[8]Department of ECE, Idaho State University, Pocatello, ID, 83209, USA.
[9]Cardiology Department, Apollo Hospitals, New Delhi, INDIA.
[10]Department of Radiology, Azienda Ospedaliero Universitaria, Cagliari, 40138, ITALY.
[11]Bennett University, Gr. Noida, INDIA
[12]Allergy, Clinical Immunology and Rheumatology Institute, Toronto, ON, CANADA.
[13]Department of Radiobiology and Molecular Genetics, National Institute of The Republic of Serbia, University of Belgrade, Belgrade, 11001, SERBIA.
[14]Division of Cardiovascular Medicine, University of Virginia, Charlottesville, VA 22904, USA.
[15]Department of Vascular Surgery, University of Lisbon, Lisbon, Portugal.
[16]Department of Neurology & Stroke Program, University of Maryland, Baltimore, MD, USA.
[17]Department of Radiology, Massachusetts General Hospital, Boston, MA.
[18]Vascular Screening and Diagnostic Centre, University of Nicosia Medical School, Nicosia, Cyprus.
[19]Department of CE, Graphic Era Deemed to be University, 248002 Dehradun, INDIA.

*Corresponding Author:
Dr. Jasjit S. Suri, PhD., MBA, FIEEE[a], FAIMBE[b], FAIUM[c], FSVM[d], FAPVS[e] , FAAAI[f]
[a]Fellow, Institute of Electrical and Electronics Engineering
[b]Fellow, American Institute of Medical and Biological Engineering
[c]Fellow, American Insitute of Ultrasound in Medicine
[d]Fellow, Society of Vascular Medicine
[e]Fellow, Asia Pacific Vascular Society
[f]Fellow, Asia Pacific Association of Artificial Intelligence
Stroke Monitoring and Diagnostic Division
AtheroPoint™, Roseville, CA 95661, USA
Phone: (916)-749-5628
Email: jasjit.suri@atheropoint.com

Deep learning Approach for Cardiovascular Disease Risk Stratification and Survival Analysis on a Canadian Cohort

Abstract

Background: The quantification of carotid plaque has been routinely used to predict cardiovascular risk in cardiovascular disease (CVD) and coronary artery disease (CAD). To determine how well carotid plaque features predict the likelihood of CAD and cardiovascular (CV) events using deep learning (DL) and compare against the machine learning (ML) paradigm.

Methodology: The participants in this study consisted of 459 individuals who had undergone coronary angiography, contrast-enhanced ultrasonography, and focused carotid B-mode ultrasound. Each patient was tracked for thirty days. The measurements on these patients consisted of maximum plaque height (MPH), total plaque area (TPA), carotid intima-media thickness (cIMT), and intraplaque neovascularization (IPN). CAD risk and CV event stratification were performed by applying eight types of DL-based models. Univariate and multivariate analysis was also conducted to predict the most significant risk predictors. The DL's model effectiveness was evaluated by the area-under-the-curve measurement while the CV event prediction was evaluated using the Cox proportional hazard model (CPHM) and compared against the DL-based concordance index (c-index).

Results: IPN showed a substantial ability to predict CV events (p<0.0001). The best DL system improved by **21%** (0.929 *vs*. 0.762) over the best ML system. DL-based CV event prediction showed a **~17%** increase in DL-based c-index compared to the CPHM (0.86 *vs*. 0.73).

Conclusions: CAD and CV incidents were linked to IPN and carotid imaging characteristics. For survival analysis and CAD prediction, the DL-based system performs superior to ML-based models.

Keywords: Cardiovascular risk prediction, focused carotid ultrasound, intraplaque neovascularization, deep learning, receiver operating characteristics, survival analysis.

1. Introduction

Cardiovascular disease (CVD) is the leading cause of mortality globally [1]. The mortality is increasing and getting affected due to different comorbidities such as hypertension, diabetes, stress, and lifestyle [2, 3]. There is an urgent need for early and non-invasive detection of CVD risk. A valuable screening method for cardiovascular events (CVE) is vascular ultrasonography which is an excellent diagnostic non-invasive tool to assess atherosclerotic plaque [4, 5]. The surrogate markers derived from the carotid artery are primarily maximum plaque height (MPH) [6], total plaque area (TPA) [7], and carotid intima-media thickness (cIMT) [8-10]. Intraplaque neovascularization (IPN), a measure of plaque instability and development, has recently been shown to be a major and reliable predictor of CVE and coronary artery disease (CAD) [11]. The previous methods to study the relationship between carotid ultrasound-based image phenotype (CUSIP) and CAD primarily used standard regression tools, and these tools were mainly linear [12, 13]. As a result, it was an oversimplification when dealing with patterns which are non-linear between the risk predictors and CAD outcomes. One of the short coming of the regression-based techniques is its inability to handle large cohort sizes. As a result, they are unable to accurately predict CVD events. The improved systems with the integration of artificial intelligence (AI) were not previously used that leads to slower and less accurate results [14, 15].

It has been recently shown that artificial intelligence (AI) techniques are superior to traditional statistical-based models [16]. AI has proven useful in a number of medical imaging applications such as dermatology [17, 18], ophthalmology [19], cardiology [20], radiology [16, 21], and recently its use in carotid ultrasonography for CVD risk assessment [22-25]. These AI models can learn the relationship between the risk predictors and the CVD endpoints using the training paradigm in an offline mode thereby generating the trained models. These trained models are then applied to transform the test risk predictors to predict the CVD events in an online mode.

3

Deep learning (DL) has started to emerge in several areas of medicine [26]. Our study hypothesize that DL-based algorithms predict CVD risk more accurately when compared to machine learning (ML)-based models [27-29]. Since every component for an input sequence receives signals from dual sources such as present and past, thus, bidirectional models are stronger than the unidirectional models [30, 31]. Further, DL models have added advantages of automated feature extractions and works with the multiple layered architecture for fast feature extraction [26, 31, 32]. The authors have benchmarked the DL-based bidirectional and unidirectional models against the existing three ML-models namely gradient boost (GB), random forest (RF), and support vector machine (SVM) [33]. The authors have also performed survival analysis and univariate and multivariate analysis. The proposed DL-based CVD risk online prediction system named as AtheroEdge 3.0$_{DL}$ is shown in Figure 1. The system shows the fusion of cardiovascular risk factors, image phenotypes, and intraplaque neovascularization (IPN), which is then fed to the trained model for CVD risk prediction yielding the multi-risk granular risk (given in four colours).

The layout of the proposed study includes: The methodology is presented in section 2, while the results are presented in Section 3. The discussions are evaluated in section 4. Finally, conclusions are presented in section 5.
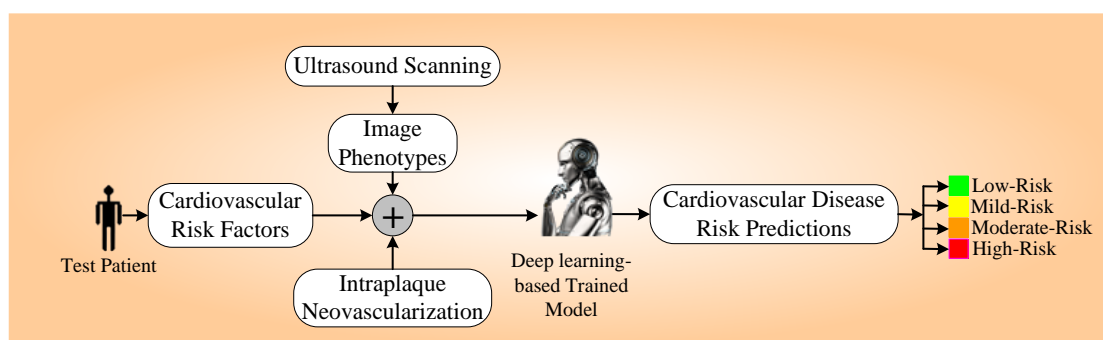

Figure 1. Global architecture for DL-based CVD prediction system (AtheroEdge 3.0$_{DL}$).

2. Methods

Study population

For this study, 459 consecutive patients (age >18 years) were recruited between December 2016 and June 2018 from the Cardiac Catheterization Laboratory at Kingston General Hospital in

Ontario, Canada. Of the 160 patients presenting with ACS, 13 had UA. All participants who were stable at the time of admission (n=299) were followed for 30 days. Regarding the 33% of patients referred for MI, the MI was within 7 days of presentation to the Catheterization laboratory. Patients with remote MI (>7 days) were excluded from the study. Further, out of 459 patients, 139 patients received stents. The inclusion criteria were (a) age $\geq$ 18 years; (b) referred for clinically indicated angiography for CAD assessment; and (c) the absence of clinical contraindication to angiography [11]. The patients with choric total occlusion (CTO), severe calcification, multi-vessel disease, left main disease, and acute kidney failure etc. were included in the study. The different types of CV events that occurred during the 30 days follow-up period were described further in Appendix B. The criteria chosen for elimination were patients who had coronary artery bypass graft surgery, carotid endarterectomy, percutaneous coronary intervention ($\geq$ 1 week), perflutren allergy, known or suspected cardiac shunt, transient ischemic attack ($\geq$ 1 week) or history of myocardial infarction, and stroke. Every participant generated baseline data, which included vital signs and medical and surgical histories. Ahead of angiography, cholesterol readings could not be collected because of staffing and scheduling issues. For every subject, written informed consent was taken. This study fully complies with the ethical criteria outlined in the 1975 Declaration of Helsinki along with interventional review Board (IRB) approval by Teaching Hospital Research Ethics Board in association with Queen's University Health Sciences, Queens, Ontario, Canada. A batch of sample data is been further demonstrated in Appendix D.

Coronary artery angiography and multiclass scoring

The modality adopted for the gold standard was coronary angiograms and these were acquired using the GE system 2000 (GE Healthcare) [34]. The clinical scientist has rated the angiograms based on their experiences [35, 36]. Four types of scores were created from high-risk to low-risk and the following criteria were used. In the case of the severe CAD disease (score of 3), the criteria adopted was 70% stenosis in the right coronary artery (RCA), left anterior descending (LAD), coronary originating from circumflex (LCx) or 50% blockage in the left

main coronary artery. In the case of the moderate-risk, the stenosis of 50-69% in the RCA, LAD, and LCx resulted in a score of 2. In the case of the mild-risk, the score considered was 1 having a blockage of 20-49%. For a score of 0, it was considered as no or minimal-risk, having the narrowing between 0-19%. The identified substantial CAD (>= 50% stenosis) had created a binary outcome [37, 38].

B-mode carotid ultrasound focused imaging

The acquisition of B-mode focused ultrasound of the carotid artery was conducted using GE Vivid E9 Ultrasound System that consisted of linear array transducer (9L-D) having a central frequency of 2.4-10 MHz. Within 24 hours following coronary angiography, an expert imaging technician conducted an ultrasound scan. The static frames of the carotid artery acquired was saved in digital image and communications in medicine (DICOM) format. The EchoPath software by GE healthcare was used for image phenotype measurement namely TPA, MPH, and cIMT. The MPH image phenotype is defined as the distant parting of the interfaces of lumen-intima media with respect to the two sides of the neck. The manual measurement of the same is done by capillary function in the bulb area or internal carotid artery. Another image phenotype is plaque area which is defined in the carotid arteries and bifurcations aligning 1 cm proximally in the both sides. Lastly, the area plaque regions are averaged for obtaining the TPA [39].

Contrast bubble carotid ultrasound imaging

Another important biomarker that depicts the plaque instability is IPN, and has been identified using contrast-enhanced carotid ultrasonography (CEUS). The criteria for IPN identification of swift migration of microbubbles contrast gained from the adventitial wall towards the plaque core formed in the carotid artery. It was then categorised into three categories: zero for no microbubbles visible in the carotid plaque, one for minimal microbubbles present but restricted to the adventitial plaque wall, and two for microbubbles present throughout the plaque. The mean IPN scores were given a single average score for every patient for each of the neck sides.

Our most recent study showed that IPN had a strong predictive value for both major CAD and upcoming CV occurrences [40].

Cardiovascular events follow-up for survival analysis

A 30 days follow-up period was fixed and each patient were observed for any kind of CV event occurrence during the follow-up periods. The different CV events are (i) coronary revascularization; (ii) stroke; (iii) cardiac death; (iv) heart failure; and (v) nonfatal myocardial infraction. So as to reduce the bias, the patients having coronary angiography in the past were not included in the follow-up study. To avoid evaluating planned revascularization therapy, patients who have undergone revascularization of coronary in between seven days after the first angiography are excluded from the subject list.

Deep learning framework for cardiovascular disease prediction

Nearly all fields of medicine have shown the role of DL for disease evaluation [41-43]. The main reason is its ability to extract a superior set of features by leveraging the large number of layers in the neural network [3, 43, 44]. Thus, we can characterize the prediction of CAD in DL framework given the risk predictors and the gold standard. The data was pre-processed before partitioning into training and testing data sets. The study implemented the normalization, scaling, min-max, scaler as part of the quality control on the data sets by using the Sklearn library in Python. The built in function from Sklearn library was used to normalize each feature by scaling the data in the range of 0-1. During implementation, the minimum and maximum values are computed and each feature value is then normalized between 0 and 1 [45-47]. We only apply the feature extraction and feature selection protocols to ML system, unlike in DL system which is self capable [45-47]. The cross-validation protocol is the most popular paradigm in the AI framework [41]. In this paradigm the data taken from the cohort (459 samples/subjects) is divided into the two subsets namely, training sets and validation/test sets (see Figure 2). During the offline mode, the training dataset is used for generating the training model that utilizes the 24 risk factors and the corresponding ground truth. This trained system is used to transform the validation/test data set to predict the risk classes leading to risk

7

stratification [48-50]. Additionally, this work used fivefold (K-5) cross-validation methods (80% training, 10% validation, and 10% testing) to examine the performance of the CV event prediction system and CAD. The training and testing system is referred as offline and online respectively as the training is done only once by using the training data set whereas the testing system is used for risk prediction.

In order to accomplish substantial CAD prediction in this study, we employed the following unidirectional deep learning (UniDL) recurrent neural network (RNN), gated recurrent unit (GRU), and long short-term memory (LSTM)) Generative Adversarial Network (GAN) and a bidirectional deep learning (BiDL) (bidirectional GAN (BiGAN), bidirectional RNN (BiRNN); bidirectional LSTM (BiLSTM); bidirectional GRU (BiGRU)) classifiers. Further, the architecture of the above mentioned UniDL and BiDL systems are described in the Appendix C. Generally, the GANs are used for data generation and data augmentation [51], however we have used GAN as a classifier. The GAN is trained as a usual model, and then it is used for CVD risk stratification using cross-validation. It should be noted that the training database did not include the test samples. The results or the outcomes of the DL system were evaluated using metrics such as accuracy, and area-under-the-curve (AUC) along with its p-value, specificity, and sensitivity. The endpoint was not a component of the "online test system"; rather, it was just utilized for the performance evaluation. As part of the benchmarking strategy, we compared our DL-based system against the ML-based (GB, RF, and SVM) paradigm attempted by Johri *et al.* [33]. In their study, the ML-based system is compared against the univariate analysis with IPN as an independent risk predictor.

The rationale behind using RNN, LSTM, GRU, and GAN
The atherosclerosis is an ethnicity-based disease that involves the blocking of arteries caused by the deposition of plaque, fats, and cholesterol on the walls of the arteries. The data used in this study involves the pool of patients belonging to the same age group i.e., elderly people, having the same ethnicity i.e., typically Caucasian, and having some form of coronary artery disease (CAD). The patients in consideration have similar characteristics with very subtle

8

differences between them. These patients are exposed to different risk classes based on their specific medical conditions. Since all these patients are sampled from the same statistical distribution and appear for medical examination in a stipulated window of time, we can therefore arrange these patients in increasing order of risk of coronary artery disease. This new distribution can be assumed to represent the progression of atherosclerosis at discrete time intervals for a single patient based on the risk at the various stages of the disease, keeping in mind the primary assumption that all patients are somewhat identical in all other attributes [30].

Deep Learning models, specifically the sequence models like RNN are well known for their ability to effectively extract temporal features and high-level semantics from sequential as well as discrete data. This is the primary intuition behind using RNN for the classification of the risk of experiencing atherosclerosis and comparing it against traditional ML-based classification models. Thus LSTM, RNN, GRU and GAN are ideal solutions for such representations [52].

Survival analysis: Cox proportional hazard model vs. Deep learning

The Univariate and multivariate logistic regression was performed and the odds ratios (ORs) with 95% confidence interval (CI) were computed to determine the risk predictors signifcantly associated with the endpoint. The OR represented the difference between the risk factor and the ground truth in both univariate and the multivariate analysis [33]. This was computed based on one year increase. DL-based survival analysis was performed using the c-index and benchmarked with the Cox proportional hazard model (CPHM). The survival and cumulative hazard curves were further plotted using DL models.

3. Results

Study inclusion and exclusion criteria

The study population for coronary angiographic screening had a cohort size of 1211 subjects. Three stages of participant exclusion occurred (Figure 3): (i) 610 participants in the initial phase satisfied the inclusion and exclusion requirements. Consequently, the study's remaining 601 individuals were not included. 110 patients were dropped from the second phase mostly due to

9

consent withdrawal, coronary angiography cancellation, and lack of IV access. Thus, only 500 people had the scanning done. 41 individuals were eliminated following picture processing because they did not have carotid artery plaque. For this study, 459 people in total were chosen.
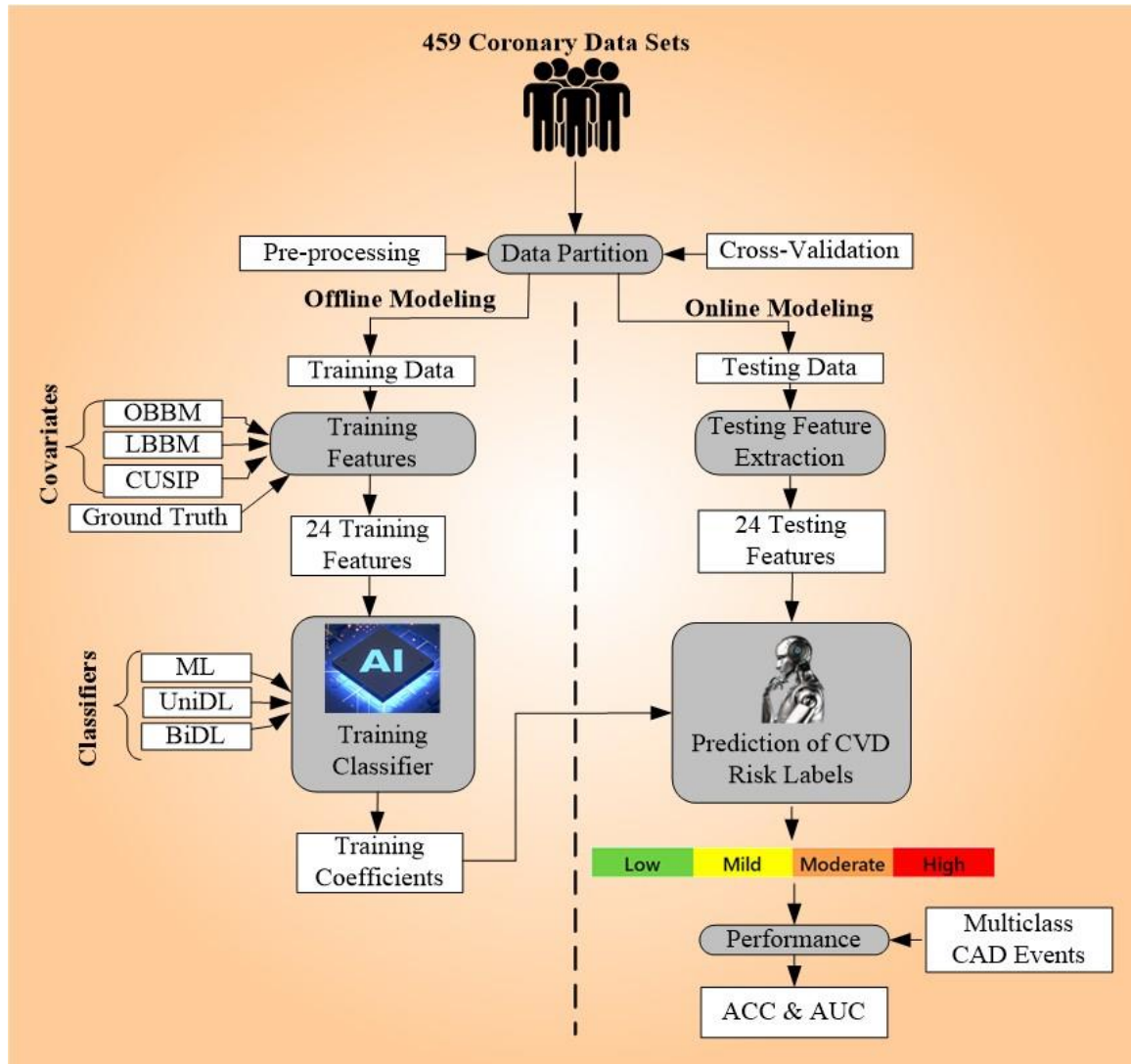

Figure 2: The architecture of an AtheroEdge 3.0$_{DL}$, a DL-based coronary artery disease prediction system. OBBM: Office-based biomarkers; LBBM: Laboratory-based biomarkers; CUSIP: Carotid ultrasound image phenotypes; ML: Machine learning (Gradient Boost, Random Forest, Support Vector Machine); DL: Deep learning (Recurrent Neural Network, Gated Recurrent Unit, Long Short Term Memory, Generative Adversarial Network, Bidirectional RNN, Bidirectional GRU, Bidirectional LSTM, Bidirectional GAN); AUC: Area-under-the-curve; ACC: Accuracy. Note that pre-processing is a quality control step for both ML and DL paradigms.
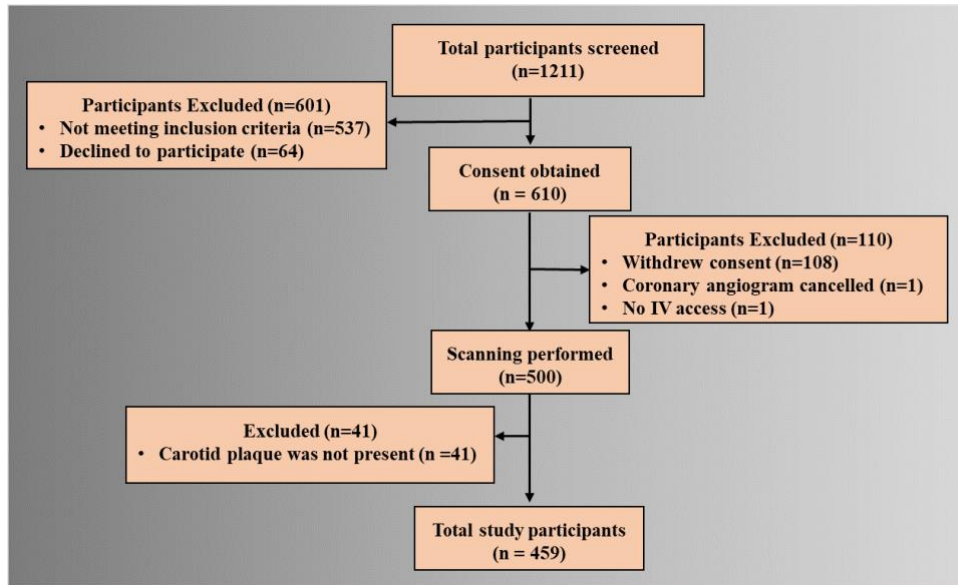
10

Figure 3. Flowchart for patient selection.

Baseline Characteristics

The p-values were calculated using the Paired T-Test [53, 54] and ANOVA test [30, 55, 56] based on continuous values or categorical values. The main risk factors was less than 0.02 were age, gender, hyperlipidemia, diabetes mellitus, and smoking history (Table 1). Their corresponding p-values were 0.022, <0.0001, 0.003, 0.048, and 0.02, respectively. Our observation showed that there was a significant difference in subjects who were taking medications namely, beta-blockers (p=0.002), anti-platelet/anti-coagulants (p=0.001), and statins (p=0.023). A significant association was found between reasons for referral (positive stress test and myocardial infarction) and CVD (p<0.05). As far as plaque characteristics for the carotid ultrasound, our observation showed a significant difference for the following biomarkers namely, TPA (p<0.0001), MPH (p<0.0001), cIMT (p<0.0001), and IPN (p<0.0001) (see Table 1).

Table 1. Baseline characteristics.

SN	Risk Factor	Overall (n=459)	p-value
1	Age	65.11±10.3	0.022
2	Male	326 (71.0%)	<0.001
3	Hypertension	317 (69.1%)	0.126
4	Diabetes T2D	78.34±18.2	0.048
5	Hyperlipidemia	269 (58.6%)	0.003
6	BMI	30.10±5.9	0.334
7	eGFR	78.34±18.2	0.122

11

8	Smoking Hx	309 (67.3%)	0.02
9	Family History	297 (64.7%)	0.341
10	Statins	250 (54.5%)	0.023
11	ACE Inhibitors	179 (39.0%)	0.731
12	ARBs Angiotensis	44 (9.6%)	0.163
13	Beta-Blockers	221 (48.1%)	0.002
14	Calcium Channel Blockers	87 (19.0%)	0.935
15	Anti-Platelet/Anti-Coagulants	339 (73.9%)	0.001
16	Diuretics	92 (20.0%)	0.561
17	cIMT	0.76±0.2	<0.001
18	MPH	2.85±1.2	<0.0001
19	TPA	51.78±44.0	<0.0001
20	IPN Score	1.26±0.8	<0.0001
21	Referral: MI, n (%)	154 (33.6%)	<0.0001
22	Referral: Chest pain, n (%)	214 (46.6%)	0.713
23	Referrel: +Stress Test, n (%)	145 (31.6%)	0.002
24	Referral: Shortness of Breath	121 (26.4%)	0.108

Unidirectional vs. bidirectional deep learning-based cardiovascular disease prediction

The uniDL results are shown in Table 2, while the biDL results in Table 3. As seen in the Table 2 GAN model outperforms the rest of the three unidirectional models (RNN, GRU, and LSTM). This can be seen in the accuracy column. The same behaviour is observed for all the evaluation metrics, namely sensitivity, specificity, and area-under-the-curve. Table 3 depicts the behaviour of biDL classification models. Our observation showed that BiGAN still outperforms the rest of the three bidirectional models (BiRNN, BiLSTM, and BiGRU). This can be clearly seen in the accuracy column. The rest of the performance attributes demonstrate the same behaviour as the uniDL models. Note carefully that there was a **~5%** improvement in the AUC of biDL systems when compared to the uniDL system.

Table 2. Performance evaluation parameters for unidirectional DL systems.

SN	Models	Accuracy	Sensitivity	Specificity	AUC	p-value
1	RNN	80.10	82.88	84.96	0.886	<0.002
2	GRU	81.18	83.17	85.32	0.895	<0.002
3	LSTM	82.88	84.37	87.67	0.910	<0.002
4	GAN	83.01	86.34	91.13	0.916	<0.002

Table 3. Performance evaluation parameters for bidirectional DL systems.

SN	Models	Accuracy	Sensitivity	Specificity	AUC	p-value
1	BiRNN	83.22	84.34	82.16	0.896	<0.002
2	BiGRU	84.42	85.18	83.17	0.908	<0.002
3	BiLSTM	85.32	86.12	85.29	0.916	<0.002

| 4 | BiGAN | 86.62 | 88.12 | 91.78 | 0.929 | <0.002 |

DL vs. ML performance for cardiovascular disease risk stratification

Figure 4 demonstrates the ROC analysis for the best DL (BiGAN) and best ML (Gradient boost). Our observation shows that there was an improvement of ~21% in AUC between DL *vs.* ML. Our observations demonstrated better CAD prediction in DL framework (AtheroEdge 3.0_{DL}) compared to ML paradigm. This behaviour was mimicked when comparing ML-based algorithms against the conventional models or calculators. DeLong test is also been performed and compared with the obtained results (see Appendix E).
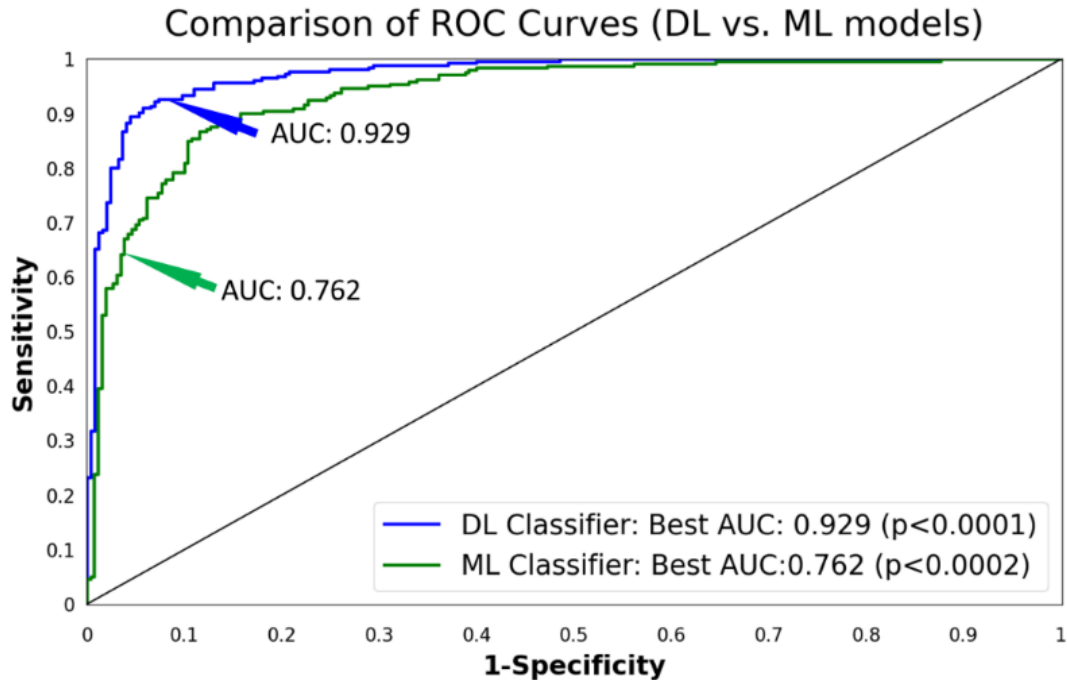

Figure 4. Comparison of ROC curves between DL *vs*. ML systems.

Univariate and Multivariate Analysis

Table 4 indicates the results for univariate logistic regression for various risk factors. The odds ratios (ORs) for the risk factors were age (OR=1.05, p=0.022), male (OR=2.26, p<0.001), hypertension (OR=1.47, p=0.126), diabetes T2D (OR=1.67, p=0.048), hyperlipidemia (OR=1.90, p=0.003), BMI (OR=0.98, p=0.334), eGFR (OR=0.96, p=0.122), smoking history (OR=1.65, p=0.02), family history (OR=1.34, p=0.341). The other group of significant risk factors are medication use such as statins (OR=1.60, p=0.023), ACE Inhibitors (OR=1.02, p=0.731), ARBs angiotensin (OR=1.75, p=0.163), beta-blockers (OR=1.87, p=0.002), calcium

13

channel blockers (OR=0.97, p=0.935), anti-platelet/anti-coagulants (OR=2.06, p=0.001), deuretics (OR=0.88, p=0.561). The CUSIP covariates are MPH (OR=2.35, p<0.0001), TPA (OR=1.95, p<0.0001), cIMT (OR=1.92, p<0.0001), IPN score (OR=2.43, p<0.0001).

Table 4. Odds ratios for univariate and multivariate analysis.

SN	Risk Factors	Univariate Analysis		Multivariate Analysis	
		OR	p-value	OR	p-value
1	Age	1.05	0.022	1.02	0.005
2	Male	2.26	<0.001	2.26	0.001
3	Hypertension	1.47	0.126	-	-
4	Diabetes T2D	1.67	0.048	1.43	0.017
5	Hyperlipidemia	1.90	0.003	1.42	0.14
6	BMI	0.98	0.334	-	-
7	eGFR	-	0.122	-	-
8	Smoking Hx	1.65	0.02	1.65	0.02
9	Family History	1.34	0.341	-	-
10	Statins	1.60	0.023	0.99	0.94
11	ACE Inhibitors	1.02	0.731	-	-
12	ARBs Angiotensis	1.75	0.163	-	-
13	Beta-Blockers	1.87	0.002	1.67	0.036
14	Calcium Channel Blockers	0.97	0.935	-	-
15	Anti-Platelet/Anti-Coagulants	2.06	0.001	1.63	0.042
16	Diuretics	0.88	0.561	-	-
17	MPH	2.35	<0.0001	1.88	<0.0001
18	TPA	1.95	<0.0001	1.7	0.007
19	cIMT	1.92	<0.0001	1.60	0.005
20	IPN Score	2.43	<0.0001	2.45	<0.0001

SN: Serial number; OR: odds ratios; BMI: Body mass index; eGFR: estimated glomerular filtration rate; ARB: Angiotensin receptor blockers; ACE: Angiotensin-coverting enzyme; IPN: intraplque neovascularization; MPH; maximum plaque height; cIMT: carotid intima-media thickness.

The main findings were as follows: (i) Clinical Biomarkers using univarate and multivariate analysis: The univariate analysis resulted several significant risk factors whose p-values were less than 0.05. They were namely, age (OR: 1.05, p-value: 0.022), sex (OR: 2.26: p-value: <0.001), diabetes T2D (OR:1.67, p-value: 0.048), hyperlipidemia (OR:1.9, p-value: 0.003), smoking history (OR: 1.65, p-value: 0.02), statins (OR:1.60, p-value: 0.023), beta-blockers (OR: 1.87, p-value: 0.002), anti-platelet/anti-coagulants (OR: 2.06, p-value: 0.001) taken from the biomarker clusters such as OBBM, LBBM, and MedUSE. When looking at the multivariate analysis, we see that age, gender, and diabetes T2D showed OR>1.0 and having the p-value <0.005. Gender had the least p-value (OR: 2.26, p-value: 0.001), while age (OR:1.02, p-value:

0.05) and diabetes T2D had an OR of 1.43 with p-value: 0.017. (ii) <u>MedUSE biomarkers using univariate and multivariate analysis</u>: In univarate analysis, we observed that there were three set of MedUSE feature set which had strong OR with p-value close to 0.001, thus influencing the CVD risk stratification. Note that for the multivariate analysis statins did not influence strongly unlike in univariate analysis. Beta-blockers and Anti-Platelet/Anti-Coagulants had OR>1.0, while p-values were 0.036 ad 0.042, respectively. In general, the p-values in univariate was lower compared to multivariate analysis. (iii) <u>Radiomics Features manual methods</u>: Further, all the carotid ultrasound image phenotypes (CUSIP) risk factors namely, MPH, TPA, cIMT showed of highest significance. While in our current study, the CUSIP were manually derived, but powerful automated system AtheroEdge™ (AtheroPoint™, Roseville, CA, USA) has been developed for measurements (see cIMT [57, 58], TPA [7, 59, 60], and MPH [6]). IPN score has shown to be powerful risk predictor for CVD risk stratification when using ML paradigm [22, 33, 35]. (iv) <u>Comparison between univariate vs. multivariate analysis using radiomics features</u>: Note that OR for MPH, TPA, cIMT, and IPN score were 2.35, 1.95, 1.92, and 2.43, having p-value <0.0001, <0.0001, <0.0001, and <0.0001, respectively. Note that OR was greater than 1.0 and having a p-value, clearly demonstrating the power of the radiomics features. These risk predictors (cIMT, MPH, TPA, and IPN) that showed strong contributors for CVD risk stratification were further considered for the multivariate analysis. Table 4 shows that carotid image phenotypes showed the same behavior when comparing against the multivariate analysis. The OR during multivariate analysis for MPH, TPA, cIMT, and IPN score was 1.88, 1.70, 1.60, and 2.45, respectively and having the p-values <0.0001, 0.007, 0.005, <0.0001, respectively.

Survival Analysis

An improvement of 17.8% was observed in c-index obtained from the survival analysis experiment when using the DL-based algorithm (0.96) over the traditional CPHM for hazard (0.73). The results from the survival analysis were shown in Figure 5. Among all the 24 risk factors, IPN was of highest importance for the prediction of CV events as it has the first ranked

15

in the metric. The first 20 risk factors having more ability for the prediction of CV events is displayed in Figure 6. Note that for the survival and cumulative hazard curves, since the number of patients are 459 representing 459 curves, we use the standardized method for representation [33, 61-63], where the curves are clubbed together.
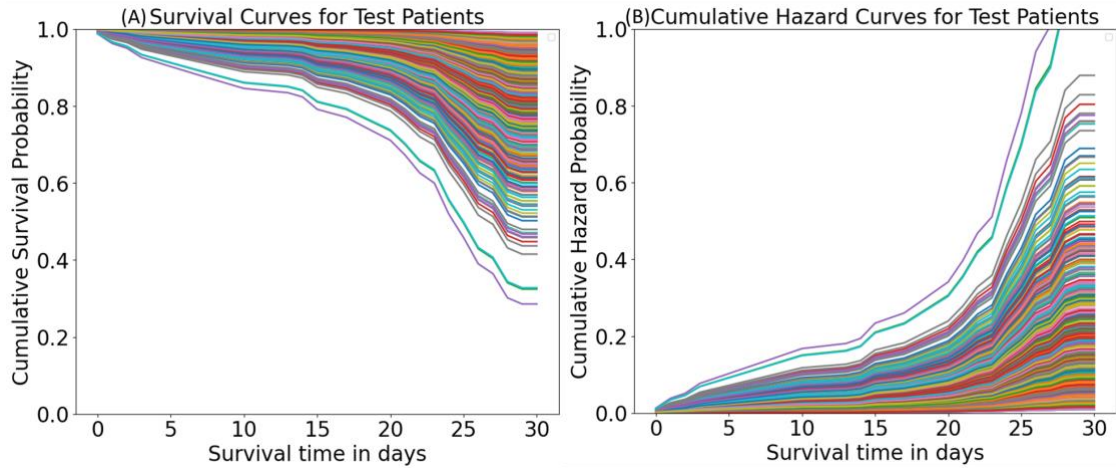

Figure 5. Plots for survival analysis using DL algorithms for 459 subjects (A) Survival curves; (B) Cumulative hazard curves.
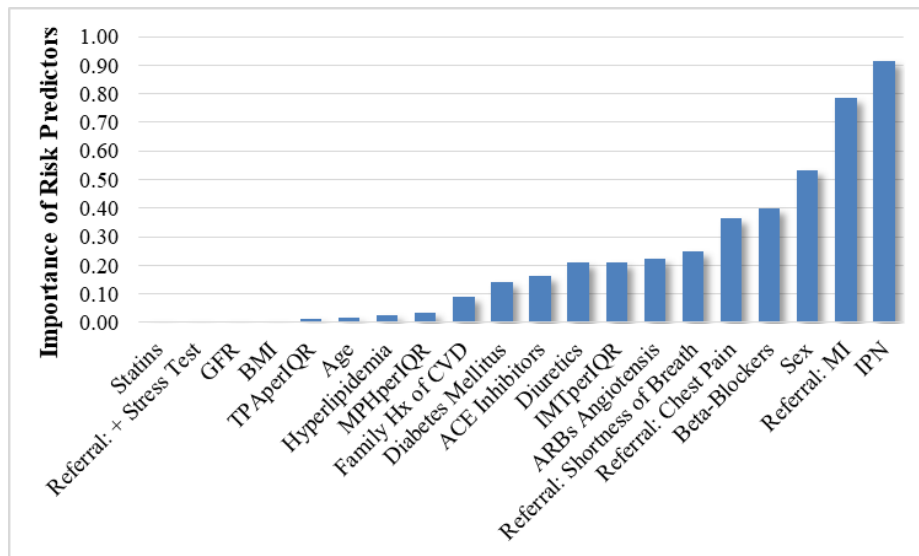

Figure 6. DL-based importance of risk predictors.

4. Discussion

This is the first study of its kind that uses deep learning for CVD risk stratification that combines CAD angiographic scores as the gold standard along with risk factors such as carotid ultrasonography plaque features, IPN, and traditional clinical risk factors. Our study

16

demonstrated a ~21% improvement in AUC for cardiovascular event (CVE) prediction in DL framework when compared against ML paradigm.

According to this study, the bidirectional DL systems performed better for CVD risk stratification than unidirectional DL models. Improved performance is due to automated feature extraction offered by DL systems. IPN was found to be the most significant risk predictor of CV events for survival analysis. It is evident from the baseline features that older people in this cohort have an increased chance of acquiring severe CVD."

Benchmarking deep learning against machine learning and conventional paradigms

Table 5 displays the benchmarking table containing eight studies with 10 attributes proposed for DL-based CVD risk stratification. The attributes are author, number of features (NOF), ML strategy, DL strategy, number of patients/images, cross-validation (CV) protocol, survival analysis (SA), and results in column C8. The first study Unnikrishnan *et al.* [64] (R1) have used the Australian datasets with 2406 patients having nine risk factors for each patient. They have opted for K-5 protocol for Framingham Risk Score (FRS) with linear regression classifiers and conventional cardiovascular risk calculator (CCVRC). The authors showed that the AUC for CCVRC using 10-year risk assessment was 0.57 unlike 0.71 for the ML-based calculators. In another study by Jamthikar *et al.* [35] (R2), a comparison was implemented between conventional calculators namely The systematic coronary risk evaluation (SCORE), FRS, and Extreme atherosclerotic cardiovascular disease (ASCVD) against ML calculators namely RF, GB, and SVM. This study used 500 subjects containing 39 risk factors. The authors showed a drastic difference between CCVRC and ML models in terms of AUC which came out to be 0.50 and 0.95, respectively along with p-value<0.0001. The third study (R3) is by Alaa *et al.* [14] that demonstrated the performance of ML classifiers against FRS system for a UK dataset with 423,604 patients having 473 covariates by following a five-year follow-up strategy. The results were better for ML classifiers (namely, RF, AdaBoost, gradient boosted machine (GBM), and SVM) when compared to CCVRC, having AUC of 0.774 (for ML) against 0.724 (CCVRC).

The CVD risk stratification using carotid scans has been gearing up. The fourth study (R4) by Zhou *et al.* [65] showed the plaque segmentation for a multi-ethnic database by using the UNet++ model. Based on the cross-validation protocol, the training and testing datasets consisted of 33, 33, 34 (3 datasets) and 44 images, respectively. As the data size was low in size, the system failed to provide proper justification for the clinical settings. The group even lacked the benchmarking of the system with other state-of-the-art systems. In another study Jain *et al.* [66] (R5) the hybrid DL models namely SegNet-UNet and SegNet-UNet+ was introduced for plaque segmentation. The segment of artery considered was internal carotid artery (ICA). The rotation transform augmentation technique was applied for the enhancement of the datasets. The drawback of the system proposed were biased to racial, data selection, and source of the data obtained. The Jain *et al.* [44] (R6) used two sets of multiethnic (Japan and Hong Kong) CCA datasets for conducting their experiments. They proposed the hybrid deep learning (HDL) architecture for the unseen dataset to avoid the different types of biases. However, the proposed system encountered bias occurred in virtue of validation as the authors lacked to perform validation of the proposed system in reference to any other systems available in the market for CVD risk prediction. The last study considered was Jain *et al.* [67] (R7), where the authors have compared their HDL and solo deep learning (SDL) models with the commercially available systems proposed by AtheroPoint™ LLC, CA, USA (AtheroEdge™ 2.0). The results showed a plaque area errors for HDL (8 mm^2), SDL (9.9 mm^2), and for conventional models (9.6 mm^2) for the image datasets.

Our system AtheroEdge 3.0_{DL} is the first of its kind that uses DL models that uses coronary angiography as gold standard and contrast-enhanced ultrasonography, and focused carotid B-mode ultrasound as part of the risk predictors. Our system consisted of four unidirectional and four bidirectional models, where bidirectional models do better than unidirectional models. Further, AtheroEdge 3.0_{DL} showed an improvement of 21% over previous ML-based solution for CVD risk stratification on the same Canadian cohort. As

18

regards survival analysis, AtheroEdge 3.0$_{DL}$ 17% improvement over CPHM, while none of the studies used survival analysis or hazard curves.

A Special Note on unidirectional and bidirectional deep learning

Our study presents the advantages of using the DL paradigm for stratification of CVD risk having the ability of feature extraction by PCA pooling mechanisms and SMOTE. Our group has conducted a comparative study of outcomes from the DL and ML-based models. We have compared unidirectional and bidirectional systems for deep learning systems. Using the same demographics data for ML and DL systems, the accuracies obtained for various opted (a) bidirectional DL models were (i) BiGAN: 86.62%, (ii) BiLSTM: 85.32%, (iii) BiGRU: 84.42%, (iv) BiRNN: 83.22%; (b) unidirectional DL systems were (i) GAN: 83.01%, (ii) LSTM: 82.88%, (iii) GRU: 81.18%, (iv) RNN: 80.10%; (c) ML systems were (i) GB: 72.00%, (ii) RF: 70.62%, (iii) SVM: 68.12%. The biDL models showed a higher performance against the unidirectional or ML-based models because the biDL models can handle the nonlinearities better between risk factors and the ground truth. This helps enable capture the complex relations between the two. The four diverse groups of risk factors namely CUSIP, LBBM, OBBM, and MedUSE were combined for the first time under the DL-based models. All 24 risk factors were grouped into these clusters based on the common features retained by the risk factors. Another aspect that we have approached is the effect of different feature clusters on the prediction of CVD risk. Further, the performance of the DL systems was enhanced by the addition of radiomics-based features (CUSIP) clusters to the earlier available clusters (namely, LBBM, OBBM, and MedUSE).

A Special Note on the final network for CVD risk stratification protocol

The group of neural networks namely, RNN, GRU, and LSTM is based on the finding of Rumelhart *et al*. [68]. These are very impactful architectures which are used to approximate nonlinear unknown dynamical systems [69]. The challenges with these systems are (i) more complex optimization objectives; (ii) disappearance of the gradient problem that directly impacts the model stability during the time of training [70]. The Figure 7 shows the final

network diagram of the RNN system which comprises of a single RNN unit with ReLU activation along with the four dense layer parallelized together. The three dense layer have 64, 32, and 8 nodes respectively. The output layer have softmax activation with four nodes in it. The model is trained end-to-end to predict the class of atherosclerotic risk to which a patient belongs to the input features.
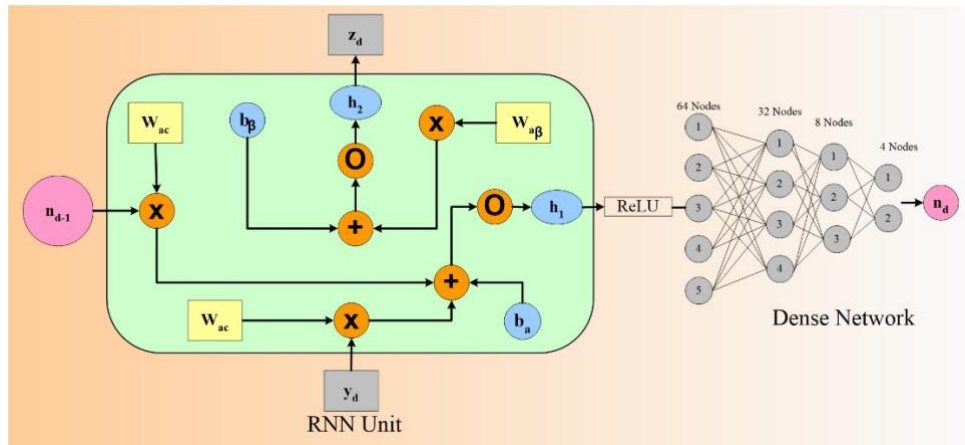

Figure 7. RNN with the network.

A special note on GAN as a classifier

While GANs is typically utilized for data generation and data augmentation, however we have adopted GAN for CVD risk stratification using cross-validation protocols. Following rationale points are presented in its support: (i) *Data Augmentation:* Synthetic samples can be added to the training dataset using GANs. GANs can help classification models perform better by producing more realistic data, particularly in situations when the training dataset is short or unbalanced. Improved generalization and robustness of classifiers can result from augmented data [51, 71-75]; (ii) *Feature Learning*: Although GANs are not designed explicitly for classification, the generator and discriminator networks in GANs learn useful feature representations of the input data. These learned features can be extracted from the discriminator or intermediate layers of the generator and used as input features for classification models. This process, known as feature extraction or transfer learning, can enhance the performance of classifiers, particularly when labeled data is limited; (iii) *Semi-supervised Learning*: There are variants of GANs, such as Semi-supervised GANs (SGANs) that combine generative and discriminative objectives. In semi-supervised learning settings, where only a small portion of

20

the data is labeled, GANs can be used to generate additional unlabeled data samples. These generated samples, along with the labeled data, can then be used to train a classifier in a semi-supervised manner, potentially improving classification performance; (iv) *Domain Adaptation*: GANs can be employed for domain adaptation, where a classifier trained on data from one domain is adapted to perform well on a different domain. By generating synthetic samples in the target domain, GANs can help bridge the domain gap and improve the performance of classifiers when labeled data in the target domain is scarce. (v) *Data Fusion for Analytics That Preserve Privacy*: Medical data is frequently delicate and governed by stringent privacy laws. Synthetic patient records that conceal individual-level information while maintaining the statistical characteristics of the real data may be created using GANs. With this method, risk prediction and other analytics activities may be carried out on synthetic data by academics and healthcare practitioners without jeopardizing patient privacy. It is feasible to create synthetic data that maintains the underlying distribution of actual patient data while guaranteeing anonymity by training GANs on huge, anonymised datasets; (vi) *Enhanced Generalization*: By exposing predictive models to a greater variety of situations and patient characteristic alterations, data generated by GANs can aid in improving the models' capacity for generalization.

Strengths, Weaknesses, and Extensions

The uniDL and biDL-based paradigm were proposed for the first time. The proposed system AtheroEdge 3.0_{DL} used coronary angiography, contrast-enhanced ultrasonography, and focused carotid B-mode ultrasound. The DL-based models possess deep layers which when trained with optimized values of batch sizes, epochs, learning rate, yields generalized training model. This brings superior DL results compared to ML results. Another advantage offered by AtheroEdge 3.0_{DL} is the ability to fuse OBBM, LBBM with radiomics-based features offered by CUSIP, IPN, and MedUSE features. The survival analysis was conducted using DL-based paradigm unlike conventional CPHM. The DL provides flexibility to increase the cohort size for retraining.

The weakness and their mitigation of the system are as follows: (i) <u>Retaining</u>: With the change in the input data, the models must undergo retraining. This change includes additions of risk predictors or change in the sample size (number of patients). Even though one requires retraining from stretch, there are alternatives for mitigation such as uses of transfer learning models [28, 45]; (ii) <u>Training Model sizes</u>: AI models are always liable for large training sizes. This is due to large number of parameters. With the increase in the risk predictors and patient cohort, the AI models are further strain in terms of training model sizes. Even though the model sizes can increase, there are mitigation strategies namely, pruning of AI algorithms [61, 76]; (iii) <u>AI Bias</u>: Bias is always a challenge in the AI due to multiple reasons such as bias in input datasets, bias in algorithm, and the bias in the prediction system. Even though bias exists in different stages and phases of AI system design, there are mitigation strategies by which we can spot and eliminate AI bias, namely by algorithms such as Butterfly, ROBINS, and PROBAST methods [26, 32, 77]; (iv) <u>Optimization during training</u>: Every AI-based training algorithm requires optimization of parameters using algorithms such as ADAM, gradient search, etc. One alternative to mitigating this is by reducing the noise in the original data this includes outliers' removal, normalization. In part of the mitigation we can also use percentage dropout during the layering process in DL [22, 43, 66]; (v) <u>Over fitting</u>: AI algorithms several time demonstrate very high accuracies giving an indication of over fitting. Another cause of over fitting is du rot small sample sizes of the input data. Even though over fitting can become a challenge in the training model design one can mitigate by introducing dropouts, augmentation of the data using SMOTE or ADASYN [30]; (vi) <u>Hardware Constrains</u>: Typically, AI systems running on CPU cluster consumes more time. This is based on number of epochs, learning rate, batch sizes, and data sizes. Even though these factors influence the performance time they can be mitigated using GPU cluster [78].

As part of the extensions, one can extend the base AI models like unidirectional and bidirectional models by attention-based and transformer-based models for CVD risk stratification. The AtheroEdge 3.0$_{DL}$ can be extended by fusion of classifiers, which comes

under the class of hybrid DL systems [45, 79, 80]. Scientific validation using cross-modality can be a possible extension [81]. The different types of plaque namely asymptomatic *vs*. symptomatic plaque obtained by grayscale modality can be added as a risk factor for CVD risk stratification. The wall segmentation methods can be improved by applying advanced DL techniques for CUSIP measurements. As part of the future work, big data concept can be opted for obtaining more accurate results for the CVD risk predictions [82].

5. Conclusion

This study presented a novel approach for CVD risk prediction using deep learning when using coronary angiography, contrast-enhanced ultrasonography, and focused carotid B-mode ultrasound. The study presented eight kinds of DL models and compared against three kinds of ML models. The main contributing risk predictors were image phenotype and intra-plaque neovascularization. There was a ~21% improvement in AUC for the DL-based (AtheroEdge 3.0_{DL}) systems when compared to ML-based algorithms. The DL-based system showed superior performance with a ~17% increase in CV event prediction compared to Cox proportional hazard models for survival analysis. This is the pilot study towards preventive, precision, and personalized medicine.

Disclosures:

Dr. Jasjit S. Suri is with AtheroPoint™ LLC, Roseville, CA, USA, dedicated in the area of cardiovascular and stroke imaging.

Dr. Aditya M. Sharma: Institutional research funding: Boston Scientific Corporation and Vascular Medcure. Speakers bureau: Boston Scientific Corporation.

Table 5. Benchmarking Table.

C0	C1	C2	C3	C4	C5	C6	C7	C8
SN	Authors	NOF	ML Model	DL Model	#Patients/#Images	CV	SA	Results
R1	Unnikrishnan *et al.* [64]	9	SVM	✗	2.4 K	K-5	✗	ML AUC:0.71; CCVRC AUC:0.57
R2	Jamthikar *et al.* [35]	39	SVM, RF, XGBoost	✗	500	K-10	✓	ML AUC:0.95; CCVRC AUC: 0.50
R3	Alaa *et al.* [14]	473	SVM, GBM, RF, AdaBoost	✗	423.6 K	K-10	✗	ML AUC: 0.724; CCVRC AUC: 0.774
R4	Zhou *et al.* [65]	✗	✗	UNet++	144/510 497/638	K-5	✗	DSC: 83.3-85.7%; TPA error: 5.55±4.34 mm^2
R5	Jain *et al.* [66]	✗	✗	UNet, UNet+, SegNet-UNet, SegNet, SegNetUNet+	97/970	K-5	✗	UNet AUC: 0.91; UNet+ AUC: 0.911; SegNet-UNet AUC: 0.908; SegNet AUC: 0.905, and SegNetUNet+ AUC: 0.898 (CE-loss models) and 0.883, 0.889, 0.905, 0.889, and 0.907 (DSC-loss models), respectively. PA error 3.49 mm^2 for SDL; 4.21 mm^2 for HDL
R6	Jain *et al.* [44]	24	✗	UNet	379, 300	K-10	✗	Unseen FoM: 70.96 and 91.14 Seen FoM: 97.57, 88.89, and 99.14
R7	Jain *et al.* [67]	24	✗	UNet, SegNet-UNet, AtheroEdge 2.0 (AtheroPoint™)	379	K-10	✗	UNet AUC: 0.93, SegNet-UNet AUC: 0.94; AtheroEdge 2.0 AUC: 0.95, respectively. SDL PA error: 9.9mm^2; HDL PA error: 8 mm^2; and AtheroEdge: 2.0 mm^2; PA error: 9.6 mm^2
R8	Proposed method	24	SVM, RF, XGBoost	RNN, GRU, GAN, LSTM, BiRNN, BiGAN, BiGRU, BiLSTM	459	K-5	✓	ML ACC: 72.00%; RNN ACC: 80.10%; GRU ACC: 81.18%; LSTM ACC: 82.88%; GAN ACC: 83.01%; BiRNN ACC: 83.22%; BiGRU ACC: 84.42%; BiLSTM ACC: 85.32%; BiGAN ACC: 86.62%.

SN: Serial Number; NOF: Number of features; ML: Machine learning; DL: Deep learning; CV: Cross-validation, SA: Survival Analysis; RSF: Random survival forest; GBM: Gradient boosting method; RF: Random forest; SVM: Support vector machine; LR: Logistic regression; XGBoost: Extreme gradient boosting method; GAN: Generative adversarial network; LSTM: Long short term memory; RNN: Recurrent neural network; GRU: Gated recurrent unit; BiRNN: Bidirectional RNN; BiLSTM: Bidirectional LSTM; BiGAN: Bidirectional GAN; BiGRU: Bidirectional GRU; AUC: Area-under-the-curve; ACC: Accurracy.

References

1. Kaptoge S., Pennells L., De Bacquer D., Cooney M.T., Kavousi M., Stevens G., Riley L.M., Savin S., Khan T., and Altay S., *World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions.* The Lancet global health, 2019. **7**(10): p. e1332-e1345.

2. Suri J.S., Agarwal S., Gupta S.K., Puvvula A., Biswas M., Saba L., Bit A., Tandel G.S., Agarwal M., and Patrick A., *A narrative review on characterization of acute respiratory distress syndrome in COVID-19-infected lungs using artificial intelligence.* Computers in Biology Medicine, 2021. **130**: p. 104210.

3. Jamthikar A.D., Puvvula A., Gupta D., Johri A.M., Nambi V., Khanna N.N., Saba L., Mavrogeni S., Laird J.R., and Pareek G., *Cardiovascular disease and stroke risk assessment in patients with chronic kidney disease using integration of estimated glomerular filtration rate, ultrasonic image phenotypes, and artificial intelligence: a narrative review.* International Angiology: a Journal of the International Union of Angiology, 2020. **40**(2): p. 150-164.

4. Suri J.S., Kathuria C., and Molinari F., *Atherosclerosis disease management.* 2010: Springer Science & Business Media.

5. Saba L., Jamthikar A., Gupta D., Khanna N.N., Viskovic K., Suri H.S., Gupta A., Mavrogeni S., Turk M., and Laird J.R., *Global perspective on carotid intima-media thickness and plaque: should the current measurement guidelines be revisited?* International Angiology, 2019. **38**(6): p. 451-465.

6. Johri A.M., Lajkosz K.A., Grubic N., Islam S., Li T.Y., Simpson C.S., Ewart P., Suri J.S., and Hétu M.-F., *Maximum plaque height in carotid ultrasound predicts cardiovascular disease outcomes: a population-based validation study of the American society of echocardiography's grade II–III plaque characterization and protocol.* The International Journal of Cardiovascular Imaging, 2021. **37**(5): p. 1601-1610.

7. Puvvula A., Jamthikar A.D., Gupta D., Khanna N.N., Porcu M., Saba L., Viskovic K., Ajuluchukwu J.N., Gupta A., and Mavrogeni S., *Morphological carotid plaque area is associated with glomerular filtration rate: A study of south asian indian patients with diabetes and chronic kidney disease.* Angiology, 2020. **71**(6): p. 520-535.

8. Amato M., Montorsi P., Ravani A., Oldani E., Galli S., Ravagnani P.M., Tremoli E., and Baldassarre D., *Carotid intima-media thickness by B-mode ultrasound as surrogate of coronary atherosclerosis: correlation with quantitative coronary angiography and coronary intravascular ultrasound findings.* European heart journal, 2007. **28**(17): p. 2094-2101.

9. Bots M.L., *Carotid intima-media thickness as a surrogate marker for cardiovascular disease in intervention studies.* Current medical research opinion, 2006. **22**(11): p. 2181-2190.

10. Spence J.D., *Ultrasound measurement of carotid plaque as a surrogate outcome for coronary artery disease.* The American journal of cardiology, 2002. **89**(4): p. 10-15.

11. Mantella L.E., Colledanchise K.N., Hetu M.-F., Feinstein S.B., Abunassar J., and Johri A.M., *Carotid intraplaque neovascularization predicts coronary artery disease and cardiovascular events.* European Heart Journal-Cardiovascular Imaging, 2019. **20**(11): p. 1239-1247.

12. Goldstein B.A., Navar A.M., and Carter R.E., *Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges.* European heart journal, 2017. **38**(23): p. 1805-1814.

13. Jamthikar A.D., Gupta D., Saba L., Khanna N.N., Viskovic K., Mavrogeni S., Laird J.R., Sattar N., Johri A.M., and Pareek G., *Artificial intelligence framework for predictive cardiovascular and stroke risk assessment models: A narrative review of integrated approaches using carotid ultrasound.* Computers in Biology Medicine, 2020. **126**: p. 104043.

14. Alaa A.M., Bolton T., Di Angelantonio E., Rudd J.H., and Van der Schaar M., *Cardiovascular disease risk prediction using automated machine learning: A*

25

prospective study of 423,604 UK Biobank participants. PloS one, 2019. **14**(5): p. e0213653.

15. Weng S.F., Reps J., Kai J., Garibaldi J.M., and Qureshi N., *Can machine-learning improve cardiovascular risk prediction using routine clinical data?* PloS one, 2017. **12**(4): p. e0174944.

16. Biswas M., Kuppili V., Saba L., Edla D.R., Suri H.S., Cuadrado-Godia E., Laird J.R., Marinhoe R.T., Sanches J.M., and Nicolaides A., *State-of-the-art review on deep learning in medical imaging.* Frontiers in Bioscience-Landmark, 2019. **24**(3): p. 380-406.

17. Sanga P., Singh J., Dubey A.K., Khanna N.N., Laird J.R., Faa G., Singh I.M., Tsoulfas G., Kalra M.K., and Teji J.S., *DermAI 1.0: A Robust, Generalized, and Novel Attention-Enabled Ensemble-Based Transfer Learning Paradigm for Multiclass Classification of Skin Lesion Images.* Diagnostics, 2023. **13**(19): p. 3159.

18. Shrivastava V.K., Londhe N.D., Sonawane R.S., and Suri J.S., *Reliable and accurate psoriasis disease classification in dermatology images using comprehensive feature space in machine learning paradigm.* Expert Systems with Applications, 2015. **42**(15-16): p. 6184-6195.

19. Fritzsche K., Can A., Shen H., Tsai C., Turner J., Tanenbuam H., Stewart C., Roysam B., Suri J., and Laxminarayan S., *Automated model based segmentation, tracing and analysis of retinal vasculature from digital fundus images.* State-of-The-Art Angiography, Applications Plaque Imaging Using MR, CT, Ultrasound X-rays, 2003. **29**: p. 225-298.

20. Saba L., Sanfilippo R., Sannia S., Anzidei M., Montisci R., Mallarini G., and Suri J.S., *Association between carotid artery plaque volume, composition, and ulceration: a retrospective assessment with MDCT.* American Journal of Roentgenology, 2012. **199**(1): p. 151-156.

21. Saba L., Biswas M., Kuppili V., Godia E.C., Suri H.S., Edla D.R., Omerzu T., Laird J.R., Khanna N.N., and Mavrogeni S., *The present and future of deep learning in radiology.* European journal of radiology, 2019. **114**: p. 14-24.

22. Jamthikar A., Gupta D., Khanna N.N., Saba L., Araki T., Viskovic K., Suri H.S., Gupta A., Mavrogeni S., and Turk M., *A low-cost machine learning-based cardiovascular/stroke risk assessment system: integration of conventional factors with image phenotypes.* Cardiovascular diagnosis therapy, 2019. **9**(5): p. 420.

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

24. Jamthikar A., Gupta D., Saba L., Khanna N.N., Araki T., Viskovic K., Mavrogeni S., Laird J.R., Pareek G., and Miner M., *Cardiovascular/stroke risk predictive calculators: a comparison between statistical and machine learning models.* Cardiovascular diagnosis therapy, 2020. **10**(4): p. 919.

25. Jamthikar A., Gupta D., Khanna N.N., Saba L., Laird J.R., and Suri J.S., *Cardiovascular/stroke risk prevention: A new machine learning framework integrating carotid ultrasound image-based phenotypes and its harmonics with conventional risk factors.* Indian heart journal, 2020. **72**(4): p. 258-264.

26. Bhagawati M., Paul S., Agarwal S., Protogeron A., Sfikakis P.P., Kitas G.D., Khanna N.N., Ruzsa Z., Sharma A.M., and Tomazu O., *Cardiovascular disease/stroke risk stratification in deep learning framework: a review.* Cardiovascular Diagnosis Therapy, 2023. **13**(3): p. 557.

27. Saxena S., Jena B., Mohapatra B., Gupta N., Kalra M., Scartozzi M., Saba L., and Suri J.S., *Fused deep learning paradigm for the prediction of o6-methylguanine-DNA methyltransferase genotype in glioblastoma patients: A neuro-oncological investigation.* Computers in Biology Medicine, 2023. **153**: p. 106492.

28. Sanagala S.S., Nicolaides A., Gupta S.K., Koppula V.K., Saba L., Agarwal S., Johri A.M., Kalra M.S., and Suri J.S., *Ten fast transfer learning models for carotid*

ultrasound plaque tissue characterization in augmentation framework embedded with heatmaps for stroke risk stratification. Diagnostics, 2021. **11**(11): p. 2109.

29. Saba L., Sanagala S.S., Gupta S.K., Koppula V.K., Johri A.M., Sharma A.M., Kolluri R., Bhatt D.L., Nicolaides A., and Suri J.S., *Ultrasound-based internal carotid artery plaque characterization using deep learning paradigm on a supercomputer: a cardiovascular disease/stroke risk assessment system.* The International Journal of Cardiovascular Imaging, 2021. **37**: p. 1511-1528.

30. Johri A.M., Singh K.V., Mantella L.E., Saba L., Sharma A., Laird J.R., Utkarsh K., Singh I.M., Gupta S., and Kalra M.S., *Deep learning artificial intelligence framework for multiclass coronary artery disease prediction using combination of conventional risk factors, carotid ultrasound, and intraplaque neovascularization.* Computers in Biology Medicine, 2022. **150**: p. 106018.

31. Khanna N.N., Maindarkar M.A., Viswanathan V., Puvvula A., Paul S., Bhagawati M., Ahluwalia P., Ruzsa Z., Sharma A., and Kolluri R., *Cardiovascular/Stroke Risk Stratification in Diabetic Foot Infection Patients Using Deep Learning-Based Artificial Intelligence: An Investigative Study.* Journal of Clinical Medicine, 2022. **11**(22): p. 6844.

32. Suri J.S., Bhagawati M., Paul S., Protogerou A.D., Sfikakis P.P., Kitas G.D., Khanna N.N., Ruzsa Z., Sharma A.M., and Saxena S., *A powerful paradigm for cardiovascular risk stratification using multiclass, multi-label, and ensemble-based machine learning paradigms: A narrative review.* Diagnostics, 2022. **12**(3): p. 722.

33. Johri A.M., Mantella L.E., Jamthikar A.D., Saba L., Laird J.R., and Suri J.S., *Role of artificial intelligence in cardiovascular risk prediction and outcomes: comparison of machine-learning and conventional statistical approaches for the analysis of carotid ultrasound features and intra-plaque neovascularization.* The International Journal of Cardiovascular Imaging, 2021. **37**(11): p. 3145-3156.

34. Johri A.M., Chitty D.W., Matangi M., Malik P., Mousavi P., Day A., Gravett M., and Simpson C., *Can carotid bulb plaque assessment rule out significant coronary artery disease? A comparison of plaque quantification by two-and three-dimensional ultrasound.* Journal of the American Society of Echocardiography, 2013. **26**(1): p. 86-95.

35. Jamthikar A.D., Gupta D., Mantella L.E., Saba L., Laird J.R., Johri A.M., and Suri J.S., *Multiclass machine learning vs. conventional calculators for stroke/CVD risk assessment using carotid plaque predictors with coronary angiography scores as gold standard: A 500 participants study.* The International Journal of Cardiovascular Imaging, 2021. **37**: p. 1171-1187.

36. Johri A.M., Calnan C.M., Matangi M.F., MacHaalany J., and Hétu M.-F., *Focused vascular ultrasound for the assessment of atherosclerosis: a proof-of-concept study.* Journal of the American Society of Echocardiography, 2016. **29**(9): p. 842-849.

37. Members T.F., Montalescot G., Sechtem U., Achenbach S., Andreotti F., Arden C., Budaj A., Bugiardini R., Crea F., and Cuisset T., *2013 ESC guidelines on the management of stable coronary artery disease: the Task Force on the management of stable coronary artery disease of the European Society of Cardiology.* European heart journal, 2013. **34**(38): p. 2949-3003.

38. Braun T., Spiliopoulos S., Veltman C., Hergesell V., Passow A., Tenderich G., Borggrefe M., and Koerner M.M., *Detection of myocardial ischemia due to clinically asymptomatic coronary artery stenosis at rest using supervised artificial intelligence-enabled vectorcardiography—A five-fold cross validation of accuracy.* Journal of Electrocardiology, 2020. **59**: p. 100-105.

39. Touboul P.-J., Hennerici M., Meairs S., Adams H., Amarenco P., Bornstein N., Csiba L., Desvarieux M., Ebrahim S., and Hernandez Hernandez R., *Mannheim Carotid Intima-Media Thickness and Plaque Consensus (2004–2006–2011) An Update on Behalf of the Advisory Board of the 3rd, 4th and 5th Watching the Risk Symposia, at the 13th, 15th and 20th European Stroke Conferences, Mannheim, Germany, 2004,*

Brussels, Belgium, 2006, and Hamburg, Germany, 2011. Cerebrovascular diseases, 2012. **34**(4): p. 290-296.

40. Deyama J., Nakamura T., Takishima I., Fujioka D., Kawabata K.-i., Obata J.-e., Watanabe K., Watanabe Y., Saito Y., and Mishina H., *Contrast-enhanced ultrasound imaging of carotid plaque neovascularization is useful for identifying high-risk patients with coronary artery disease.* Circulation Journal, 2013. **77**(6): p. 1499-1507.

41. Tandel G.S., Tiwari A., Kakde O.G., Gupta N., Saba L., and Suri J.S., *Role of Ensemble Deep Learning for Brain Tumor Classification in Multiple Magnetic Resonance Imaging Sequence Data.* Diagnostics, 2023. **13**(3): p. 481.

42. Suri J.S., Agarwal S., Saba L., Chabert G.L., Carriero A., Paschè A., Danna P., Mehmedović A., Faa G., and Jujaray T., *Multicenter study on COVID-19 lung computed tomography segmentation with varying glass ground opacities using unseen deep learning artificial intelligence paradigms: COVLIAS 1.0 validation.* Journal of Medical Systems, 2022. **46**(10): p. 62.

43. Jain P.K., Dubey A., Saba L., Khanna N.N., Laird J.R., Nicolaides A., Fouda M.M., Suri J.S., and Sharma N., *Attention-based UNet deep learning model for plaque segmentation in carotid ultrasound for stroke risk stratification: an artificial intelligence paradigm.* Journal of Cardiovascular Development Disease, 2022. **9**(10): p. 326.

44. Jain P.K., Sharma N., Saba L., Paraskevas K.I., Kalra M.K., Johri A., Laird J.R., Nicolaides A.N., and Suri J.S., *Unseen artificial intelligence—Deep learning paradigm for segmentation of low atherosclerotic plaque in carotid ultrasound: A multicenter cardiovascular study.* Diagnostics, 2021. **11**(12): p. 2257.

45. Dubey A.K., Chabert G.L., Carriero A., Pasche A., Danna P.S., Agarwal S., Mohanty L., Nillmani, Sharma N., and Yadav S., *Ensemble Deep Learning Derived from Transfer Learning for Classification of COVID-19 Patients on Hybrid Deep-Learning-Based Lung Segmentation: A Data Augmentation and Balancing Framework.* Diagnostics, 2023. **13**(11): p. 1954.

46. Suri J.S., Agarwal S., Gupta S.K., Puvvula A., Viskovic K., Suri N., Alizad A., El-Baz A., Saba L., and Fatemi M., *Systematic review of artificial intelligence in acute respiratory distress syndrome for COVID-19 lung patients: a biomedical imaging perspective.* IEEE journal of biomedical health informatics, 2021. **25**(11): p. 4128-4139.

47. Wu D.H., Chen Z., North J.C., Biswas M., Vo J., and Suri J.S., *Machine learning pardigm for dynamic contrast-enhanced MRI evaluation of expanding bladder.* Frontiers in Bioscience-Landmark, 2020. **25**(9): p. 1746-1764.

48. Wong T.-T., *Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation.* Pattern recognition, 2015. **48**(9): p. 2839-2846.

49. Duval S., Van't Hof J.R., Steffen L.M., and Luepker R.V., *Estimation of cardiovascular risk from self-reported knowledge of risk factors: insights from the Minnesota Heart Survey.* Clinical Epidemiology, 2020: p. 41-49.

50. Sammut C. and Webb G.I., *Encyclopedia of machine learning.* 2011: Springer Science & Business Media.

51. Revathi T., Sathiyabhama B., and Sankar S., *Diagnosing cardio vascular disease (CVD) using generative adversarial network (GAN) in retinal fundus images.* Annals of the Romanian Society for Cell Biology, 2021: p. 2563-2572.

52. O'Donncha F., Hu Y., Palmes P., Burke M., Filgueira R., and Grant J., *A spatio-temporal LSTM model to forecast across multiple temporal and spatial scales.* Ecological Informatics, 2022. **69**: p. 101687.

53. Suri J.S., Agarwal S., Carriero A., Paschè A., Danna P.S., Columbu M., Saba L., Viskovic K., Mehmedović A., and Agarwal S., *COVLIAS 1.0 vs. MedSeg: artificial intelligence-based comparative study for automated COVID-19 computed tomography lung segmentation in Italian and Croatian Cohorts.* Diagnostics, 2021. **11**(12): p. 2367.

54. Maniruzzaman M., Rahman M.J., Ahammed B., Abedin M.M., Suri H.S., Biswas M., El-Baz A., Bangeas P., Tsoulfas G., and Suri J.S., *Statistical characterization and*

classification of colon microarray gene expression data using multiple machine learning paradigms. Computer methods programs in biomedicine, 2019. **176**: p. 173-193.

55. Saba L., Banchhor S.K., Londhe N.D., Araki T., Laird J.R., Gupta A., Nicolaides A., and Suri J.S., *Web-based accurate measurements of carotid lumen diameter and stenosis severity: an ultrasound-based clinical tool for stroke risk assessment during multicenter clinical trials.* Computers in biology medicine, 2017. **91**: p. 306-317.

56. Saba L., Than J.C., Noor N.M., Rijal O.M., Kassim R.M., Yunus A., Ng C.R., and Suri J.S., *Inter-observer variability analysis of automatic lung delineation in normal and disease patients.* Journal of medical systems, 2016. **40**: p. 1-18.

57. Molinari F., Zeng G., and Suri J.S., *Intima-media thickness: setting a standard for a completely automated method of ultrasound measurement.* IEEE transactions on ultrasonics, ferroelectrics, frequency control, 2010. **57**(5): p. 1112-1124.

58. Molinari F., Pattichis C.S., Zeng G., Saba L., Acharya U.R., Sanfilippo R., Nicolaides A., and Suri J.S., *Completely automated multiresolution edge snapper—a new technique for an accurate carotid ultrasound IMT measurement: clinical validation and benchmarking on a multi-institutional database.* IEEE Transactions on image processing, 2011. **21**(3): p. 1211-1222.

59. Biswas M., Saba L., Chakrabartty S., Khanna N.N., Song H., Suri H.S., Sfikakis P.P., Mavrogeni S., Viskovic K., and Laird J.R., *Two-stage artificial intelligence model for jointly measurement of atherosclerotic wall thickness and plaque burden in carotid ultrasound: A screening tool for cardiovascular/stroke risk assessment.* Computers in biology medicine, 2020. **123**: p. 103847.

60. Cuadrado-Godia E., Maniruzzaman M., Araki T., Puvvula A., Rahman M.J., Saba L., Suri H.S., Gupta A., Banchhor S.K., and Teji J.S., *Morphologic TPA (mTPA) and composite risk score for moderate carotid atherosclerotic plaque is strongly associated with HbA1c in diabetes cohort.* Computers in biology medicine, 2018. **101**: p. 128-145.

61. Saba L., Maindarkar M., Khanna N.N., Johri A.M., Mantella L., Laird J.R., Paraskevas K.I., Ruzsa Z., Kalra M.K., and Fernandes J.F.E., *A Pharmaceutical Paradigm for Cardiovascular Composite Risk Assessment Using Novel Radiogenomics Risk Predictors in Precision Explainable Artificial Intelligence Framework: Clinical Trial Tool.* FRONTIERS IN BIOSCIENCE, 2023. **28**(10).

62. Khanna N.N., Maindarkar M., Puvvula A., Paul S., Bhagawati M., Ahluwalia P., Ruzsa Z., Sharma A., Munjral S., and Kolluri R., *Vascular Implications of COVID-19: Role of Radiological Imaging, Artificial Intelligence, and Tissue Characterization: A Special Report.* Journal of Cardiovascular Development Disease, 2022. **9**(8): p. 268.

63. Saxena S., Jena B., Gupta N., Das S., Sarmah D., Bhattacharya P., Nath T., Paul S., Fouda M.M., and Kalra M., *Role of artificial intelligence in radiogenomics for cancers in the era of precision medicine.* Cancers, 2022. **14**(12): p. 2860.

64. Unnikrishnan P., Kumar D.K., Poosapadi Arjunan S., Kumar H., Mitchell P., and Kawasaki R., *Development of health parameter model for risk prediction of CVD using SVM.* Computational mathematical methods in medicine, 2016. **2016**.

65. Zhou R., Guo F., Azarpazhooh M.R., Hashemi S., Cheng X., Spence J.D., Ding M., and Fenster A., *Deep learning-based measurement of total plaque area in B-mode ultrasound images.* IEEE Journal of Biomedical Health Informatics, 2021. **25**(8): p. 2967-2977.

66. Jain P.K., Sharma N., Giannopoulos A.A., Saba L., Nicolaides A., and Suri J.S., *Hybrid deep learning segmentation models for atherosclerotic plaque in internal carotid artery B-mode ultrasound.* Computers in biology medicine, 2021. **136**: p. 104721.

67. Jain P.K., Sharma N., Saba L., Paraskevas K.I., Kalra M.K., Johri A., Nicolaides A.N., and Suri J.S., *Automated deep learning-based paradigm for high-risk plaque detection in B-mode common carotid ultrasound scans: An asymptomatic Japanese cohort study.* Int. Angiol, 2021. **41**: p. 9-23.

68.   Rumelhart D.E., Hinton G.E., and Williams R.J., *Learning internal representations by error propagation*. 1985, Institute for Cognitive Science, University of California, San Diego La ….

69.   Durstewitz D., *A state space approach for piecewise-linear recurrent neural networks for identifying computational dynamics from neural measurements*. PLoS computational biology, 2017. **13**(6): p. e1005542.

70.   Bengio Y., Simard P., and Frasconi P., *Learning long-term dependencies with gradient descent is difficult*. IEEE transactions on neural networks, 1994. **5**(2): p. 157-166.

71.   Sarra R.R., Dinar A.M., Mohammed M.A., Ghani M.K.A., and Albahar M.A., *A robust framework for data generative and heart disease prediction based on efficient deep learning models*. Diagnostics, 2022. **12**(12): p. 2899.

72.   García-Vicente C., Chushig-Muzo D., Mora-Jiménez I., Fabelo H., Gram I.T., Løchen M.-L., Granja C., and Soguero-Ruiz C. *Clinical synthetic data generation to predict and identify risk factors for cardiovascular diseases*. in *VLDB Workshop on Data Management and Analytics for Medicine and Healthcare*. 2022. Springer.

73.   Dogan A., Li Y., Odo C.P., Sonawane K., Lin Y., and Liu C., *A utility-based machine learning-driven personalized lifestyle recommendation for cardiovascular disease prevention*. Journal of Biomedical Informatics, 2023. **141**: p. 104342.

74.   Ossenberg-Engels J. and Grau V. *Conditional generative adversarial networks for the prediction of cardiac contraction from individual frames*. in *International Workshop on Statistical Atlases and Computational Models of the Heart*. 2019. Springer.

75.   Shokrollahi Y., Dong P., Zhou C., Li X., and Gu L., *Deep Learning-Based Prediction of Stress and Strain Maps in Arterial Walls for Improved Cardiovascular Risk Assessment*. Applied Sciences, 2023. **14**(1): p. 379.

76.   Agarwal M., Agarwal S., Saba L., Chabert G.L., Gupta S., Carriero A., Pasche A., Danna P., Mehmedovic A., and Faa G., *Eight pruning deep learning models for low storage and high-speed COVID-19 computed tomography lung segmentation and heatmap-based lesion localization: A multicenter study using COVLIAS 2.0*. Computers in biology medicine, 2022. **146**: p. 105571.

77.   Suri J.S., Bhagawati M., Paul S., Protogeron A., Sfikakis P.P., Kitas G.D., Khanna N.N., Ruzsa Z., Sharma A.M., and Saxena S., *Understanding the bias in machine learning systems for cardiovascular disease risk assessment: The first of its kind review*. Computers in biology medicine, 2022. **142**: p. 105204.

78.   Narayanan R., Werahera P., Barqawi A., Crawford E., Shinohara K., Simoneau A., and Suri J., *Adaptation of a 3D prostate cancer atlas for transrectal ultrasound guided target-specific biopsy*. Physics in Medicine Biology, 2008. **53**(20): p. N397.

79.   Suri J.S., Agarwal S., Chabert G.L., Carriero A., Paschè A., Danna P.S., Saba L., Mehmedović A., Faa G., and Singh I.M., *COVLIAS 2.0-cXAI: Cloud-based explainable deep learning system for COVID-19 lesion localization in computed tomography scans*. Diagnostics, 2022. **12**(6): p. 1482.

80.   Skandha S.S., Nicolaides A., Gupta S.K., Koppula V.K., Saba L., Johri A.M., Kalra M.S., and Suri J.S., *A hybrid deep learning paradigm for carotid plaque tissue characterization and its validation in multicenter cohorts using a supercomputer framework*. Computers in biology medicine, 2022. **141**: p. 105131.

81.   Saba L., Sanagala S.S., Gupta S.K., Koppula V.K., Johri A.M., Khanna N.N., Mavrogeni S., Laird J.R., Pareek G., and Miner M., *Multimodality carotid plaque tissue characterization and classification in the artificial intelligence paradigm: A narrative review for stroke application*. Annals of Translational Medicine, 2021. **9**(14).

82.   El-Baz A. and Suri J.S., *Big data in multimodal medical imaging*. 2019: CRC Press.

83.   Jena B., Saxena S., Nayak G.K., Saba L., Sharma N., and Suri J.S., *Artificial intelligence-based hybrid deep learning models for image classification: The first narrative review*. Computers in Biology Medicine, 2021. **137**: p. 104803.

84.   Dey R. and Salem F.M. *Gate-variants of gated recurrent unit (GRU) neural networks*. in *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*. 2017. IEEE.

85. Creswell A., White T., Dumoulin V., Arulkumaran K., Sengupta B., and Bharath A.A., *Generative adversarial networks: An overview.* IEEE signal processing magazine, 2018. **35**(1): p. 53-65.

86. Singh J., Singh N., Fouda M.M., Saba L., and Suri J.S., *Attention-Enabled Ensemble Deep Learning Models and Their Validation for Depression Detection: A Domain Adoption Paradigm.* Diagnostics, 2023. **13**(12): p. 2092.

87. Olhosseiny H.H., Mirzaloo M., Bolic M., Dajani H.R., Groza V., and Yoshida M. *Identifying high risk of atherosclerosis using deep learning and ensemble learning.* in *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA).* 2021. IEEE.

88. Suri J.S., Bhagawati M., Agarwal S., Paul S., Pandey A., Gupta S.K., Saba L., Paraskevas K.I., Khanna N.N., and Laird J.R., *UNet Deep Learning Architecture for Segmentation of Vascular and Non-Vascular Images: A Microscopic Look at UNet Components Buffered With Pruning, Explainable Artificial Intelligence, and Bias.* IEEE Access, 2022. **11**: p. 595-645.

89. Donahue J., Krähenbühl P., and Darrell T., *Adversarial feature learning.* arXiv preprint arXiv:.09782, 2016.

90. DeLong E.R., DeLong D.M., and Clarke-Pearson D.L., *Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach.* Biometrics, 1988: p. 837-845.

91. Sun X. and Xu W., *Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves.* IEEE Signal Processing Letters, 2014. **21**(11): p. 1389-1393.

Appendix A

Abbreviation Table

SN	Abbreviation	Description
1	ACC	Accuracy
2	AI	Artificial Intelligence
3	ASCVD	Extreme atherosclerotic cardiovascular disease
4	AUC	Area-under-the-curve
5	BiDL	Bidirectional deep learning
6	BiGAN	Bidirectional generative adversarial network
7	BiGRU	Bidirectional gated recurrent unit
8	BiLSTM	Bidirectional long short term memory
9	BiRNN	Bidirectional recurrent neural network
10	CAD	Coronary artery disease
11	CCVRC	Conventional cardiovascular risk calculator
12	CEUS	Contrast-enhanced carotid ultrasonography
13	cIMT	Carotid intima-media thickness
14	CPHM	Cox proportional hazard model
15	CUSIP	Carotid ultrasound-based phenotype
16	CVD	Cardiovascular disease
17	CVE	Cardiovascular event
18	DICOM	Digital image and communications in medicine
19	DL	Deep learning
20	FRS	Framingham risk score
21	GAN	Generative adversarial network
22	GB	Gradient boost
23	GBM	Gradient boost machine
24	GRU	Gated recurrent unit
25	HDL	Hybrid deep learning
26	IPN	Intraplaque neovascularization
27	LBBM	Laboratory-based biomarkers
28	LSTM	Long short term memory
29	MedUSE	Medication used
30	MPH	Maximum plaque height
31	ML	Machine learning
32	NOF	Number of features
33	OBBM	Office-based biomarkers
34	OR	Odds ratios
35	RF	Random forest
36	RNN	Recurrent neural network
37	ROC	Receiver operating characteristic
38	SCORE	Systematic coronary risk evaluation
39	SDL	Solo deep learning
40	SVM	Support vector machine
41	TPA	Total plaque area
42	UniDL	Unidirectional deep learning

32

Appendix B

The participants with an IPN score of 1.25 or greater (n=159) had 25 CV events during the follow-up period, whereas participants with an IPN score of 1.25 or lower (n=150) had 4 CV events. The table below describes the different CV events namely, cardiac arrest, stroke/transient ischemic attack (TIA), acute coronary syndrome (ACS), heart failure, and cardiac death having 4, 4, 11, 4, and 6 number of patients respectively. The Table A1 was added as follows:

Table A1. Different types of CV events occurred during the 30 days follow-up period.

CV Events	Overall (n=459)
Cardiac Arrest	4
Stroke/TIA	4
ACS	11
Heart Failure	4
Cardiac Death	6
Total	29

Appendix C

Architecture unidirectional DL Models

The unidirectional DL models namely, recurrent neural network (RNN) [83], Gated recurrent units (GRU) [84], Long short-term memory (LSTM) [83], and generative adversarial network (GAN) [85]. Sequence-based applications including speech recognition, natural language processing, and time series prediction are especially well-suited for RNNs. However, because of problems like disappearing or bursting gradients, typical RNNs have trouble capturing long-term relationships (See Figure A1). The LSTM architecture is shown in Figure A2. One kind of RNN architecture called LSTM (See Figure A2) was created to solve the vanishing gradient issue. Because of its memory cells and different gates that regulate information flow, it is more effective at identifying long-range relationships in sequences.

33

Another RNN variation that resembles LSTM but has a slightly different structure is called GRU. It is also equipped with gating mechanisms to regulate the information flow, and it has demonstrated strong performance in a range of sequence-based activities (Figure A3).

GAN is made up of two neural networks: a discriminator and a generator as shown in Figure A4. While the discriminator assesses samples to discern between produced and actual data, the generator synthesizes data by altering random noise in an attempt to replicate real data. The discriminator gains the ability to distinguish between actual and synthetic samples, while the generator becomes more proficient at producing realistic data through adversarial training. As a result of these networks' ongoing interaction, increasingly convincing and high-quality data are produced, which makes GANs effective tools for tasks like data synthesis, picture production, and style transfer (See Figure A4). Generally, GANs are used for data generation and data augmentation [51], however, we have used GANs for classification protocols. The GAN is trained as a usual model, and then it is used for CVD risk stratification using cross-validation.
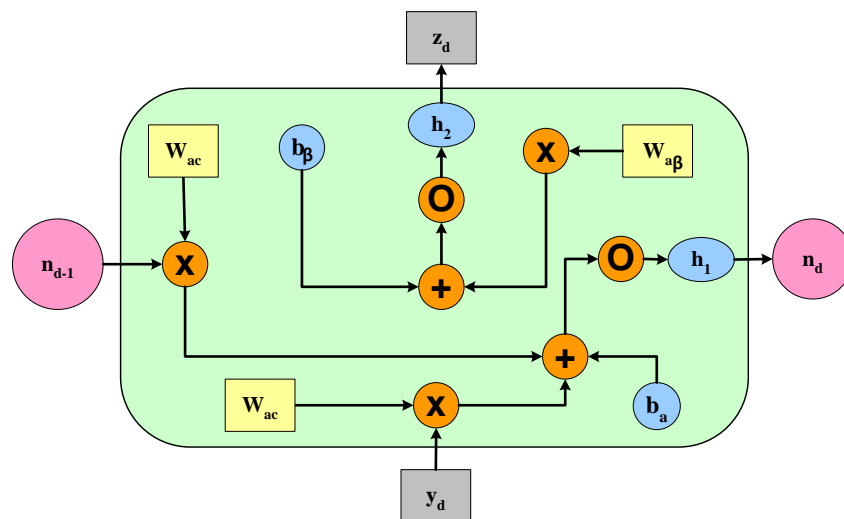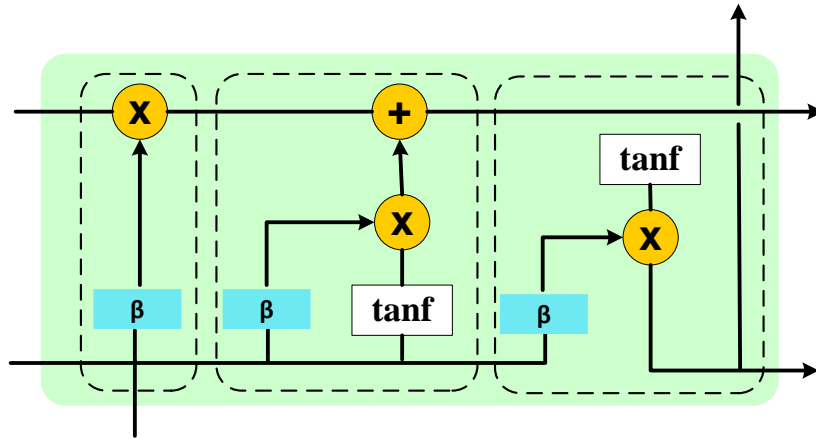

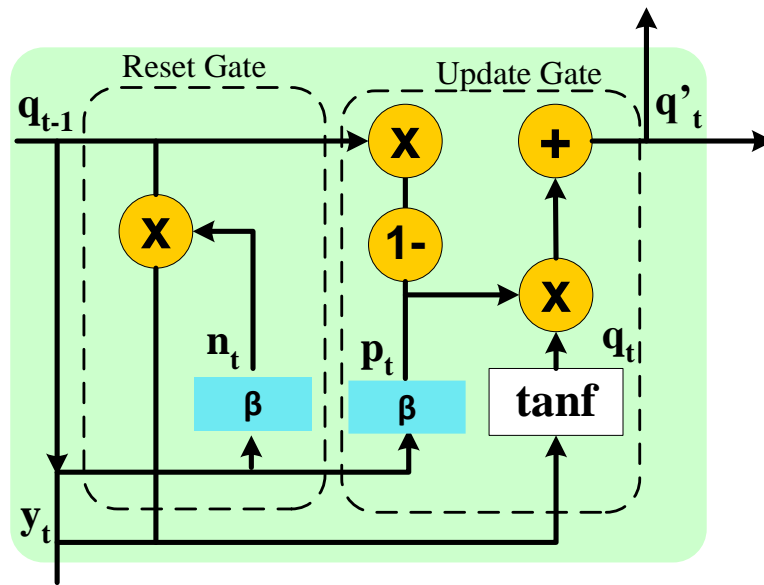Figure A1. RNN Architecture.

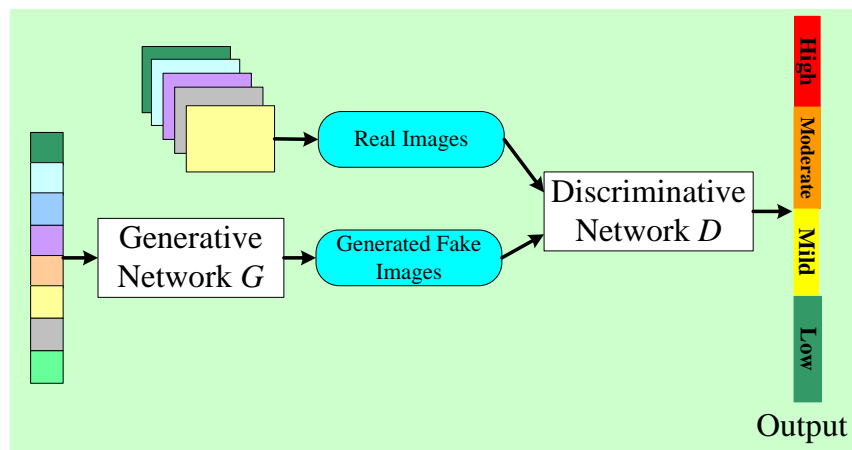Figure A2. LSTM Architecture.

Figure A3. GRU Architecture.

Figure A4. GAN Architecture.

35

Architecture of bidirectional DL Models

The bidirectional DL systems namely, BiRNN [86] (See Figure A5), BiLSTM [86] (See Figure A6), BiGRU [87] (See Figure A7), and BiGAN [88] (See Figure A8). BiRNN analyzes sequential data by traversing the input sequence both forward and backward at the same time. At each time step, the network gathers data from the past during the forward pass and gathers data from the future during the backward pass. The context surrounding each time step is therefore holistically represented by combining the concealed states from both directions. Because of this bidirectional approach, BiRNNs are especially useful for tasks where thorough context understanding is essential, like CVD risk prediction, depression disorder prediction, natural language processing, speech recognition, and time-series analysis. The network is better able to capture dependencies and relationships within the sequential data. To maximize the network's capacity to provide precise predictions or classifications based on the bidirectional contextual data, backpropagation is used to change the network's parameters over training.

An expansion of the conventional LSTM, the BiLSTM uses bidirectional information flow to interpret sequential input. The BiLSTM, which consists of both forward and backward LSTM layers, concurrently records contextual data from previous and future time steps. The LSTM units control input, forget, and output gates to control information flow and maintain memory cells to retain pertinent data at each time step. The network can comprehend dependencies and long-term relationships within the sequence because of its bidirectional nature. Concatenation of the concealed states from both sides usually allows for a more thorough context representation. This design is especially helpful for jobs where good prediction requires knowledge of both previous and succeeding parts, such as time-series analysis and natural language processing. Backpropagation is used to change the network's parameters over time during training, improving its capacity to represent intricate sequential relationships.

BiGRU uses two parallel GRU layers—one for forward processing and the other for backward processing—to analyze sequential input. The forward GRU simultaneously gathers

data from previous elements while the backward GRU concurrently gathers data from subsequent elements at each time step. After that, the concealed states from both sides are usually concatenated to produce a complete context representation. To mitigate the vanishing gradient problems that are frequently found in conventional RNNs, GRU units use gating methods to regulate the flow of input. BiGRU networks can efficiently describe dependencies and connections within sequential data thanks to their bidirectional nature. The network's parameters are updated during training via backpropagation over time, which maximizes the network's capacity to provide precise predictions or classifications based on bidirectional contextual data.

GAN architecture is expanded upon by BiGAN. It includes the generator and discriminator as well as an encoder. The encoder maps produced and actual data back to a common latent space, whereas the generator produces synthetic data. The encoder and generator seek to trick the discriminator, which makes the distinction between produced and actual data. The capacity of the model to capture intricate linkages between the data and latent space is improved by this bidirectional mapping. The network uses adversarial learning to iteratively improve the generator, discriminator, and encoder during training, promoting the creation of more realistic samples and more informative latent space. Applications for BiGANs may be found in many fields, such as representation learning and picture production.
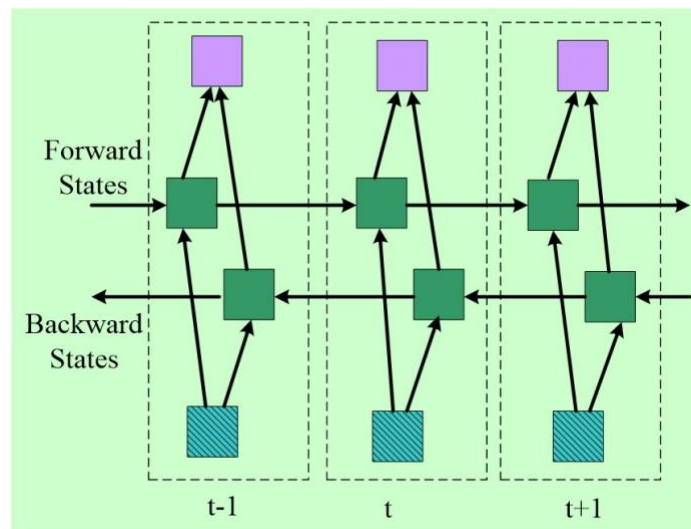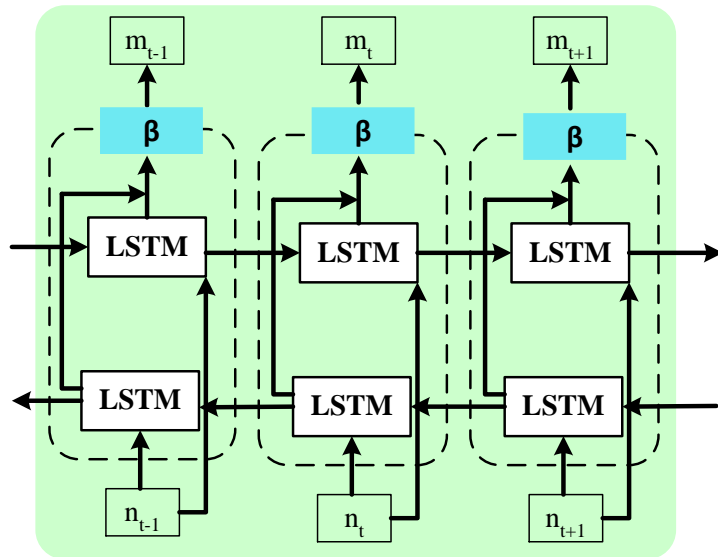

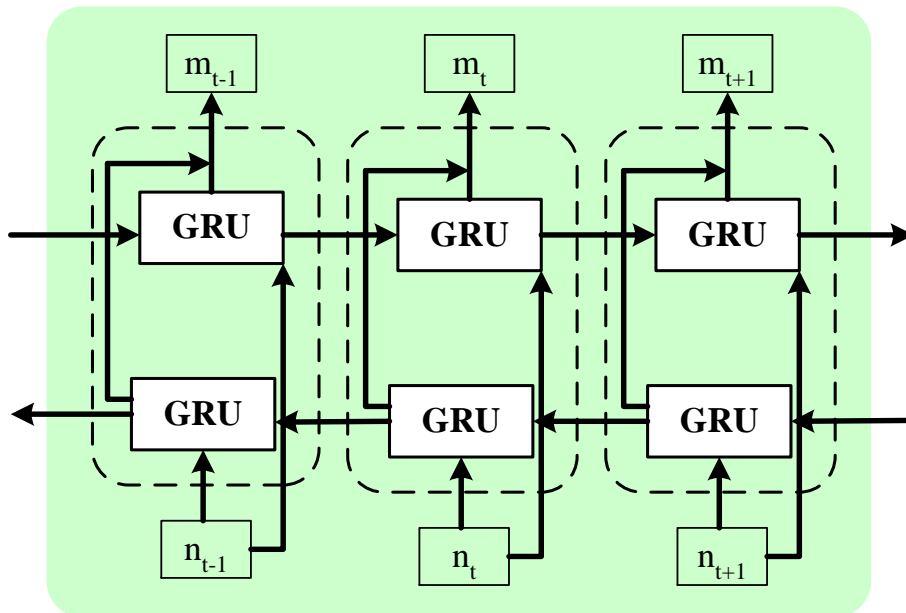Figure A5. BiRNN Architecture.

Figure A6. BiLSTM Architecture.
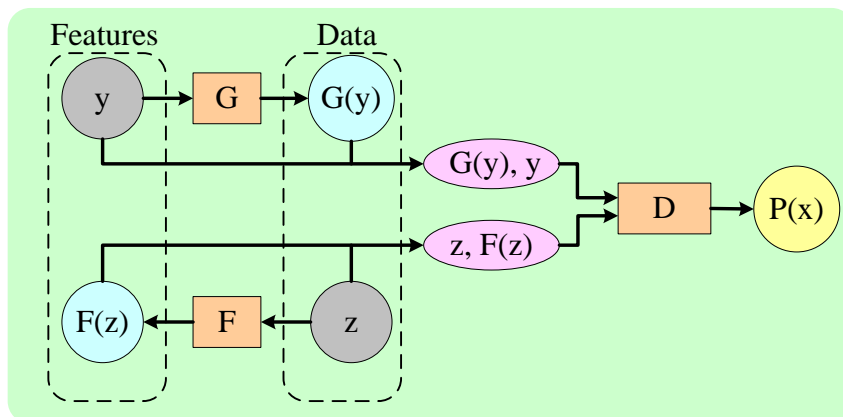

Figure A7. BiGRU Architecture.

Figure A8. BiGAN Architecture [89].

Appendix D

A Batch of 3 samples Data

ID	Age	Sex	BMI	GFR	BP	HLD	DM	SH
1	55	1	25.5	100	1	1	1	0
2	54	1	36.0	100	1	1	0	1
3	55	0	37.2	84	1	0	0	0
Figure A9: Column 1-9; BMI: Body mass index; GFR: Glomerular filtration rate;							
HLD: Hyperlipidemia; DM: Diabetes Mellitus; SH: Smoking Hx;							
BP: Blood pressure;							

FH	Statins	ACEI	ARBA	BB	CCB	AP/AC	Diuretics	MI
1	1	1	0	0	0	1	0	0
1	1	0	0	0	1	0	1	0
1	1	0	0	1	0	1	0	1
Figure A9: Column 10-18: FH: Family history of CVD;							
ACEI: ACE Inhibitors; CCB: Calcium Channel Blockers; MI: Referral: MI;							
ARBA: ARBs Angiotensis; BB: Beta-Blockers;							
AP/AC: Anti-Platelet/Anti-Coagulants							

CP	ST	SB	IMT	MPH	TPA	IPN	Censored	DS
0	1	0	3.6	1.13	0.15	0	0	30
1	0	0	3.98	1.38	0.85	1.33	0	30
0	0	0	3.93	1.14	0.14	8	0	30
Figure A9: Column 19-27: CP: Referral: Chest Pain;							
IMT: IMTperIQR; MPH:MPHperIQR; TPA: TPAperIQR;							
SB: Referral: Shortness of Breath; DS: Days of Survival;							
IPN: IPN Scoreper025; ST: Referral: Stress Test							

Figure A9. A Batch of 3 sample data.

Appendix E

Result of DeLong Test

The DeLong test states whether the AUCs of two models are statistically significantly different or not by comparing [90]. In the DeLong's Test the report states the z-score, p-value and the two AUCs. Under the null hypothesis, z-score can be well approximated by the standard normal distribution. Therefore, if the value of z deviates too much from zero, it is thus reasonable to consider that [theta(A) > theta(B)] with the significance level $p < 0.05$ [91].

We have performed the DeLong test by using the built in library pROC in R and obtained the results as AUC for DL is 0.920 and for ML 0.759, z-score=1.485, and the p-

value=0.080. For the obtained z-score it is well proved the AUCs of two models (ML and DL) are statistically <u>significantly different</u>. Further, we noticed that DL model performs better the ML model. The Table A2 displays the DeLong's test result.

Table A2. DeLong's test Result.

Models	AUC		p-value		z-score	
	ML	DL	ML	DL	ML	DL
DeLong Test	0.759	0.920	0.080		1.459	

Difference between DeLong Test and Standardized ROC curves:

The same hypothesis has been proven from the AUCs of the two models (DL and ML) as presented in the manuscript. The ROC plots demonstrate the AUCs and the difference between them very clearly. Table A3 shows the comparison of both the methods.

Table A3. Comparison of DeLong's test and Standardized ROC curves.

Models	AUC		p-value		z-score	
	ML	DL	ML	DL	ML	DL
Standardized ROC curves	0.762	0.929	<0.0001	<0.0002	-	-
DeLong Test	0.759	0.920	0.080		1.459	