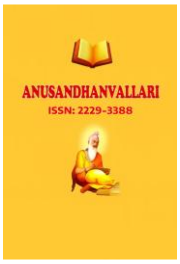


Anusandhanvallari
Vol 2024, No.1
September 2024
ISSN 2229-3388

Enhancing Cardiovascular Disease Detection Through Dataset Integration and TVAE-Driven Synthetic Data Generation with 1D CNN

Swati S Khandalkar[1], Shweta M Barhate[2]

Department of Electronics and Computer Science, RTMNU, Nagpur[1]

Department of Electronics and Computer Science, RTMNU, Nagpur[2]

Abstract

Early diagnosis of cardiovascular disease (CVD) plays a vital role in disease prevention. This paper proposes to predict the risk of developing CVD´s using traditional clinical parameters and some cardiac biomarkers such as C-reactive protein (CRP), Homocysteine, and lipoprotein(a)(Lp(a)). This research works by augmenting two different datasets and make use of Tabular Variational Autoencoder (TVAE) to obtain 10,000 balanced samples. One-Dimensional Convolutional Neural Network (1D-CNN) has been trained to detect intricate relationships within said dataset. Our algorithm was rather efficient, reached accuracy of 91%, precision of 91%, recall of 90% and f1 score 90%. Feature Importance Permutation was employed to determine which features were the most important in making CVD prediction. The study found out that the most critical parameter was Hypertension, while Smoking and Diabetes came in a close second place. Although these conventional factors were responsible for making up the primary CVD prediction model, some specific biomarkers were indispensable for reaching high accuracy.

Keywords: Cardiovascular disease, Synthetic Data, CRP, Homocysteine, Lp(a), 1D-CNN, TVAE, Deep Learning

1. Introduction

Cardiovascular disease (CVD) prediction is important for early diagnosis and treatment. Disease detection is highly significant. If doctors diagnose cardiac diseases at an early stage, then they would be able to treat their patients effectively and save their lives. Conventional risk stratification, such as Framingham Heart Study, depends on simple parameters including age, BMI, and BP [1]. The old approach usually fails to detect complicated associations.

Cardiovascular disorders associated with chronic systemic diseases have major consequences for vascular anatomy and physiology. Hypertension plays an significant role in the onset of heart disease, where the consistent pressure exerted on the arteries causes the arteries to become dysfunctional and more rigid, causing faster plaque deposition [2].Similarly, diabetes mellitus significantly increases the risk of cardiovascular diseases due to systemic hyperglycemia and insulin resistance, which lead to oxidative stress and inflammation of the blood vessels, thus making diabetes not only a metabolic disorder but also a cardiovascular disease equivalent[3]. Other than metabolism-related risks, there are also specific behavioral risks such as smoking. Smoking is a major chemical exposure that exposes one’s body to toxins that will affect endothelium, increase chances of blood clotting, and speed up the formation of atherosclerosis [4].

The latest findings reveal that modern cardiologists discovered novel biomarkers in the bloodstream. This novel group includes CRP, Homocysteine, and Lipoprotein(a). These markers reveal hidden inflammation and genetic risks that standard test failure [5]. The inclusion of the particular biomarkers into the standard health information leads to better insights into the patient’s heart condition.

Systemic inflammation is known to play an important part in cardiovascular pathologies, which can be measured in terms of high-sensitivity C-reactive protein (CRP). Clinical studies conducted recently have found that CRP, measuring residual inflammatory risk, was a more powerful predictor of cardiovascular disease and death compared to residual cholesterol risk in people on statin therapy [6]. Homocysteine is metabolically an important independent marker as a sulfur-containing amino acid. Hyperhomocysteinemia causes endothelial cell injury, oxidative stress, and prothrombosis. Recent studies have established its strong association with atherosclerotic cardiovascular disease, especially in young individuals and those not having any classical risk factors for this condition [7]. Finally, Lipoprotein(a) (Lp[a]) has proven to be an essential risk factor with a genetic basis that needs to be managed carefully through clinical evaluation. It is quite similar to LDL cholesterol except for the presence of the atherogenic apolipoprotein(a). Lp[a] has been identified as a key driver of residual Cardiovascular risk and aortic stenosis regardless of whether other lipid levels are under medical control [8].

The Artificial Intelligence (AI) technology is currently used to predict diseases. Deep learning algorithms are capable of recognizing complicated associations that cannot be detected by a person [9,10]. AI needs massive amount of data to work correctly. Real-world medical datasets are usually very small and unbalanced [11]. So, the lack of data makes it impossible to create highly effective AI models. The proposed paper is aimed at overcoming this challenge. Following steps were implemented while doing experimental work:

i. Two separate data sets will be integrated into one.
ii. An advanced TVAE-based algorithm [12] will be utilized for creating synthetic data.
iii. One-dimensional convolutional neural network will be used to predict the probability of developing a disease.
iv. The integration of routine health information with particular biomarkers enables to develop an effective prediction model.

2. Problem Statement

Detection of cardiovascular disease is difficult due to the absence of adequate data and imbalances in the dataset, as well as the failure to integrate relevant biomarkers. In this paper, we attempt to create a predictive model through the use of synthetic data creation and 1D-CNN with clinical and biomarker data.

3. Objectives

- To generate synthetic data using TVAE to improve dataset size and balance
- To develop a model for enhanced cardiovascular disease detection
- To evaluate model performance using standard metrics

4. Proposed Methodology

This paper presents a systematic framework that uses clinical data and biomarker values for detecting cardiovascular diseases. It involves data gathering, data cleaning, data merging, data synthesis, normalization, model training, and model testing. Data sets are merged, followed by data augmentation using synthetic data, before being normalized and used in training the deep learning model. A one-dimensional convolutional neural network model [13] is applied for making predictions (as shown in figure 1). Methodology proposed is as follows:

4.1 Data Collection

The first process that we had to carry out was the selection of appropriate medical records. The sources of these data were the Kaggle websites. For obtaining the full information about heart conditions, we made sure that there are two different sets of data.

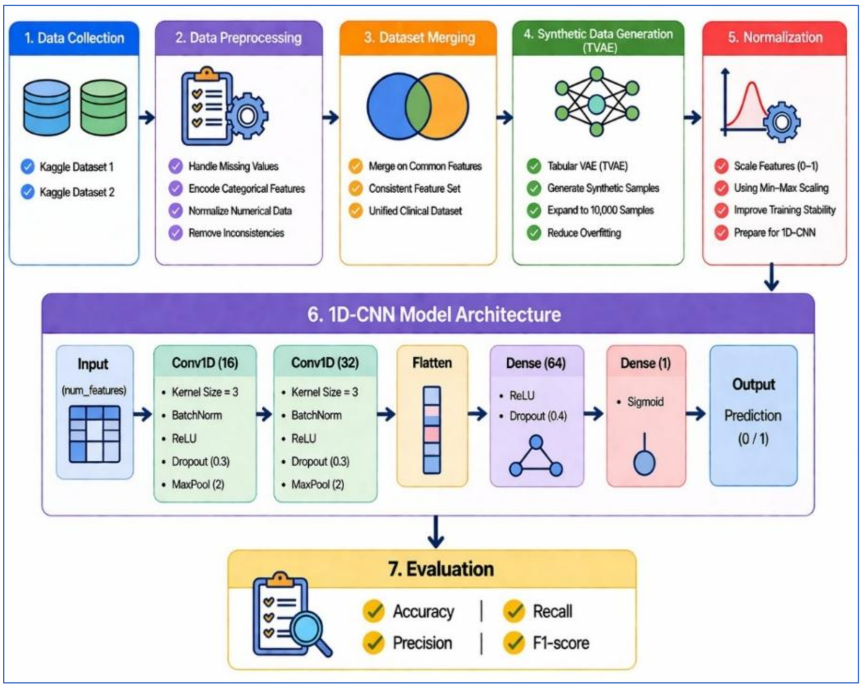


Figure 1: The Proposed Methodology

The First Dataset: The data in this set represent conventional indicators in medical history. The records include such factors as age, body mass index (BMI), Diabetes and Lp(a).

The Second Dataset: The second set contained more sophisticated indicators based on laboratory tests. In particular, these data contain such important indicators as Triglyceride Level, C-reactive protein (CRP) and Homocysteine.

Without using two sets of data, we could not obtain necessary information required to make predictions about heart diseases. Both types of data give us the opportunity to observe both physical and biochemical symptoms of heart condition.

4.2 Data Preprocessing

Raw medical data is rarely perfect. Prior to the combination of the two sets of data, we needed to manipulate the data set for the correct input of the artificial intelligence model. The cleaning process was made up of the following:

Missing Data Imputation: Missing values are a common problem in the actual record of patients. As opposed to discarding the data, we applied some statistical measures to estimate the missing values.

Encoding: Artificial Intelligence models are unable to work with the textual inputs. Therefore, categorical variables were encoded using numerical representations. The text such as "Male/Female" and "Diabetes/Non-Diabetes" was coded into 1 and 0 respectively.

Normalization: There is a huge difference between the scales used in medical tests. For example, blood pressure has a big value, whereas biomarkers have very low values. An initial normalization of the dataset was done, and this helped standardize both the values so that we could integrate them together in the subsequent process.

4.3 Datasets Merging

Both the datasets have been merged into one using common variables. The common variables present in both the datasets are Age, Gender, Cholesterol Level, Diabetes, BMI, and Triglyceride Level. Only these common variables have been considered for maintaining uniformity between both the datasets. This helps avoid any sort of error during the process of merging. Once the common variables have been considered, both the datasets have been merged. Duplicates have been removed from the database. This helps in increasing the number of observations.

4.4 Synthetic Data Generation

The medical datasets are usually not large enough to develop the deep learning algorithm adequately. Hence, we applied the Tabular Variational Autoencoder (TVAE) algorithm [14] to increase the size of our medical dataset. This cutting-edge artificial intelligence identifies the intricate statistical relationships between different characteristics of our data to create an extremely authentic sample. This study used consolidated dataset in the input of TVAE. It analyzed all the variables to create a new set of patients. Some of the specific variables include as shown in Figure 2. Target Variable: Outcome (whether the patient has any form of cardiovascular disease). In addition, TVAE recognized the connection between all these variables and their influence on the "Outcome" variable. In this innovative approach, we expanded our data collection to 10,000 balanced samples. We ensured that our deep learning algorithm would have a sufficient number of data to work on without the risk of overfitting.

4.5 Normalization

After generating 10,000 synthetic samples, we implemented one more scaling step to the final normalized version of our data set. Such an approach is crucial because of the great variability of the sizes of clinical features. For instance, there could be patients' triglycerides level in the hundreds, while the level of CRP will be a small fraction value. Without scaling such data, the machine could misinterpret the information presented, assuming that the higher the number, the greater the feature importance in mathematical terms. In consequence, we managed to normalize the entire dataset and made sure that every feature in the dataset would have the same importance in terms of scale for the 1D-CNN. As a result, our algorithm trained faster and reached its peak performance.

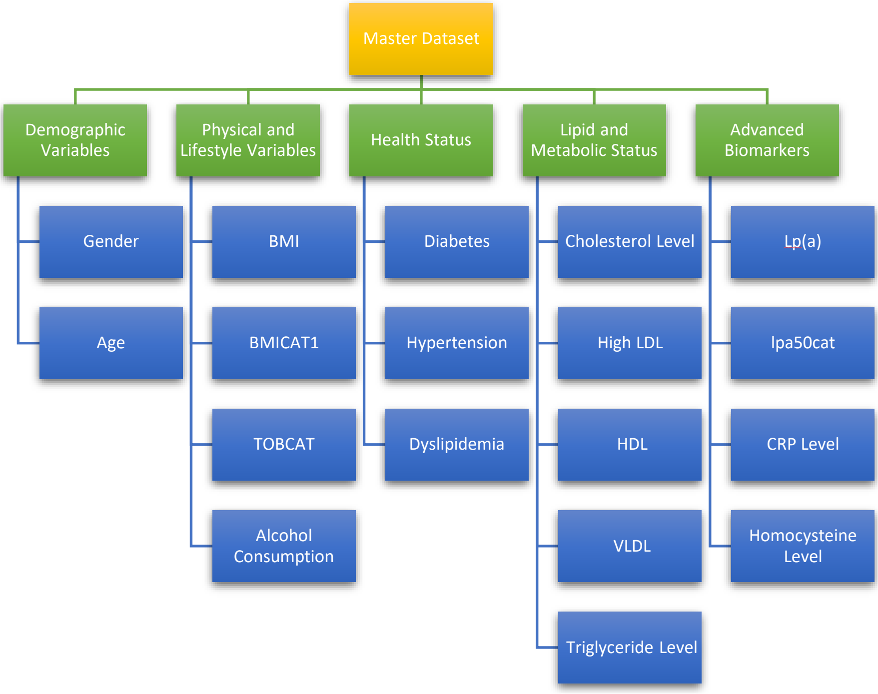


Figure 2: Hierarchical representation of Meta data

4.6 1D-CNN Training

One-Dimensional Convolutional Neural Network refers to a deep neural network architecture used to handle sequential or structured data through application of convolutions in a single dimension. This technique differs from conventional machine learning techniques since one dimensional CNN [15] have the ability to automatically learn the most important features from the raw data.

The 1D-CNN model has been used in this research work to predict cardiac diseases using clinical and biomarker datasets. This model helps in understanding interactions between different features like hypertension, diabetes, and chemical elements, leading to better classification efficiency. Fully connected layers are then used to make predictions based on the feature maps generated. Owing to its capability to generate hierarchical features and manage non-linearities, 1D-CNN models have been found more efficient than other models in healthcare applications.

4.7 Classification-Based Performance Metrics

In order to measure the performance of the 1D CNN model in a quantitative manner, the following classification measures have been obtained using the confusion matrix. The confusion matrix provides the following four basic features: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). With the help of the above-mentioned features, the performance of the model [16] has been evaluated based on four key classification metrics given in Table 1 below.

These performance measures enable an evaluation of the efficiency of the model in prediction, especially for applications in diagnosing diseases, where both false positive and false negative results are critical.

5. Literature Review

Various Artificial Intelligence and Machine learning techniques have been applied in the prediction of CVD because of their excellent predictive capability. Recent research has made use of traditional machine learning models, deep learning methods, and synthetic data generation techniques to enhance the accuracy of the predictions. A few related studies are listed below.

The work of K. Miao and J. Miao [17], proposed a more improved DNN algorithm for the prediction of the occurrence of coronary heart disease based on the clinical features available. The researchers aimed at addressing the problem of overfitting by deploying a multilayer perceptron model that has been strengthened with the application of L2 regularization and a dropout rate of 50%. The proposed DNN was tested on the famous dataset from the Cleveland Clinic Foundation, where its accuracy of classification was found to be 83.67%. Significantly important to a clinical setting, the proposed DNN model obtained a remarkably high sensitivity of 93.51%.

Zawadzki and Parvaneh [18] examined the use of Synthetic Data Generation (SDG) to resolve the common problems of class imbalance and small dataset sizes in cardiovascular predictions, with particular emphasis on enriching training samples for chronic heart failure prediction. The authors considered several SDG approaches, starting from sophisticated Generative Adversarial Networks (GANs) down to more simple architectures such as the Gaussian Copula, to determine their influence on the performance of a Catboost classifier. According to the results, the use of synthetic data resulted in minor gains in general, mainly due to enhanced F1-scores for clinically minority classes (e.g., female patients). Significantly, the simpler generative methods were able to demonstrate better performance compared to the more sophisticated yet optimized GANs, despite their synthetic data exhibiting greater quality. Nonetheless, the work showed considerable variation in the predictive stability of cross validation folds, suggesting that despite being an interesting approach for addressing class imbalances in clinical studies, SDG requires setting specific thresholds for sample sizes.

Table 1: Performance Metrics used

Evaluation Metric	Mathematical Equation	Clinical Interpretation
Accuracy	$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$	Correctness of the model when predicting both sick and healthy states.
Precision	$Precision = \frac{TP}{TP+FP}$	The Probability of a sick person to be detected by the model among all other
Recall	$Recall = \frac{TP}{TP+FN}$	Capacity of the model to detect all the genuinely sick individuals.
F1-Score	$F1\ Score = 2 * \frac{precision * recall}{precision + recall}$	General robustness and stability of the model, which is particularly important for imbalanced data sets.

The problem of the weaknesses of machine learning algorithms has led to the development of an automated Convolutional Neural Network (CNN) by Sajja and Kalluri [19]. While recognizing the shortcomings of existing classification methods in terms of the necessity of manually selecting features and the lack of predictability, the authors employed the Cleveland Heart Disease database to design a CNN with the ability to learn its own features. The deep learning method demonstrated great predictability, registering an accuracy rate of 94.78%. This rate is superior to the best scores of the conventional classifiers including K Nearest Neighbor, Logistic Regression, and Support Vector Machine which achieved only 86.8%.

In order to resolve the limitations related to the application of classical machine learning algorithms within the scope of heart disease diagnosis based on generalization of clinical data, El-Shafiey et al. [20] suggested a novel hybrid deep learning approach. In particular, the research included implementation of 1D CNN to perform automated hierarchical feature extraction using Bi-LSTM for bidirectional dependencies analysis between clinical records. Additionally, to achieve the highest accuracy level and computing effectiveness, Bayesian optimization was utilized to optimize the parameters of the suggested architecture including the sizes of filters and dense layers. After testing the performance of the newly introduced 1D CNN-BiLSTM model on the most widely used Cleveland and Statlog datasets, it became clear that the introduced approach performed impressively in terms of prediction accuracy, reaching 89.01% and 82.72% correspondingly. Consequently, it can be stated that the suggested technique surpassed the existing state-of-the-art models and proved that integration of 1D CNN and Bi-LSTM allows developing a reliable automated system for cardiovascular disease detection.

In the current paper, García-Vicente et al. [21] discussed multiple approaches to the generation of synthetic data through comprehensive analysis and experimentation. The present paper addresses the problem of imbalanced clinical data concerning the prediction of heart disease using data generation methods. Specifically, the authors examined the effectiveness of various data generation algorithms, such as CTGAN, TVAE, and SMOTEN, with regard to specific machine learning models. According to the authors, GAN models proved to be extremely efficient since they outperformed other models in terms of modeling the distribution of categorical data in the real-world environment. Additionally, through implementing different explainability methods, including SHAP, García-Vicente et al. managed to conduct clinical validation of their study results by identifying multiple factors associated with the risk of developing heart disease, including high cholesterol levels, age, BMI, and gender. Therefore, the paper demonstrates the possibility of successfully using deep learning models for the purpose of data generation while maintaining physiological interpretability.

The authors García-Ordas et al. [22] suggested a successful method based on deep learning in combination with feature augmentation to predict heart disease risks. This method takes into account different factors, such as age, gender, cholesterol level, and heartbeat of the individual. Deep learning and feature augmentation are helpful in addressing multi-factor assessment, which yields more positive results and accuracy up to 90%, exceeding other approaches by 4.4%.

Arooj et al. [23] suggested an innovative method for detecting heart disease based on a deep convolutional neural network (DCNN). This algorithm is built based on features extracted from the UCI heart disease data set. This algorithm takes clinical information as inputs and classifies the patient as either healthy or suffering from the disease. It is significant to note that the DCNN proposed by Arooj et al. was able to achieve an impressive validation accuracy of 91.7%.

6. Results and Discussions

In the study, the main purpose was to create an efficient and unbiased deep learning algorithm for CVD prediction. Machine learning algorithms traditionally employed for medical diagnoses are usually affected by poor performance as they use data sets that are usually imbalanced and require manual feature selection. In order to eliminate the limitations of conventional algorithms, the new study introduced a hybrid solution in which a generative algorithm known as Tabular Variational Autoencoder (TVAE) is used along with an autoencoder-based feature extraction technique named 1D CNN.

The clinical information that has been used for this study comes from public sources of heart disease datasets. Similarly to actual medical data, the initial set was imbalanced, as the vast majority of observations belong to the category of people who were not sick at all. Unfixed data imbalance can create an inherent bias in predictive models favoring the majority class. To address this issue, TVAEs have been employed to learn the complicated, non-Gaussian distributions of the original records of the real patients. This way, the network was able to synthesize very authentic, tabular samples, increasing the dataset to a massive size of 10,000 records. Significantly, the generation was perfectly balanced as 5,008 belonged to Outcome 1 class and 4,992 – Outcome 0 class.

After the data augmentation procedure, the customized 1D CNN was set up to learn the multi-dimensional data set. In contrast to the conventional approach, where feature engineering had to be conducted manually, the proposed 1D CNN is able to detect hidden patterns in the multi-dimensional data with its convolutional layers in a systematic manner. In order to ensure consistent learning without overfitting, i.e., merely memorizing the training data, Batch Normalization as well as rigorous Dropout layers (30-40%) were incorporated into the system (as shown in the figure 3).

In order to conduct an analysis of the model's performance and make sure it could be applicable to actual clinical cases, an initial data augmentation of the database with 10,000 patients resulted in splitting before training. The split was conducted on the conventional basis of 80% of data being used for training and only 20% for testing the performance of the 1D CNN. Training for the 1D CNN was carried out on the 8,000 patients using 50 epochs. It is very important to study the learning process because this allows determining whether the model has generalized during the training process.

Figure 4 and 5 shows the gradual convergence of the neural network. Both the training curve and the validation curve converged consistently during the 50 epochs. The model was able to make optimal weight adjustments at epoch 35, denoted by the dashed vertical line labeled "Best Epoch." Since there were no significant discrepancies between the validation and training curves for loss and accuracy, the model did not overfit.

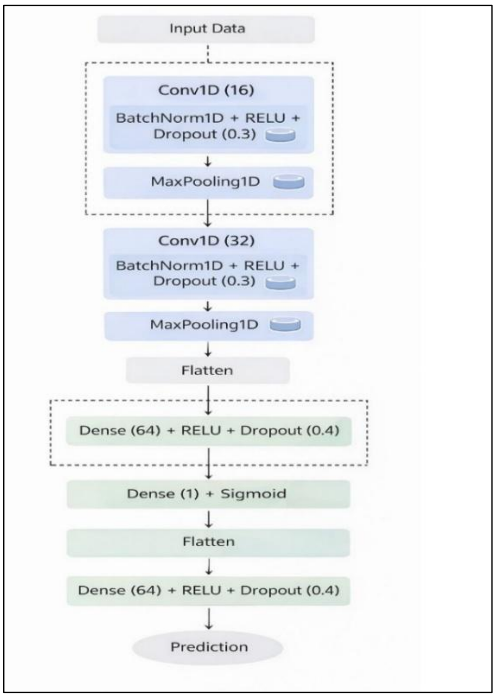


Figure 3: Workflow Diagram of 1D CNN for CVD Prediction

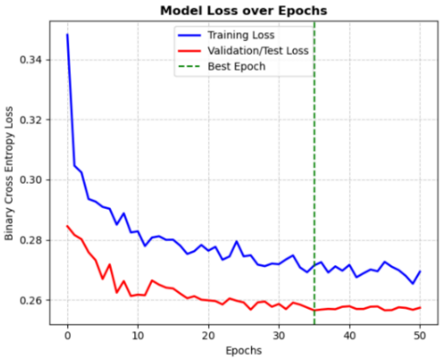


Figure 4: Model Loss over Epochs

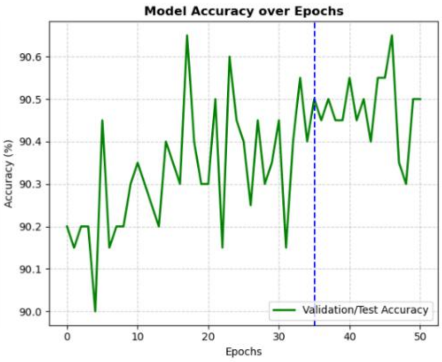
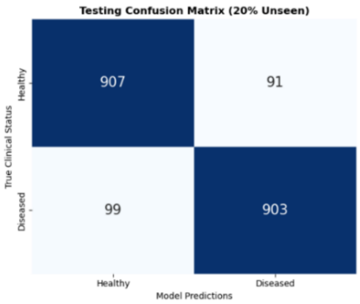


Figure 5: Model Accuracy over Epochs



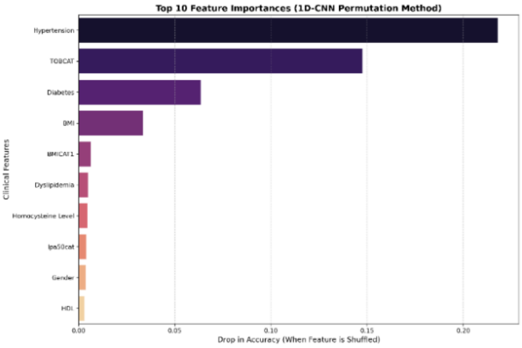


Figure 6: Confusion Matrix

Figure 7: Feature Importance

Table 2. 1D-CNN performance

Metrics	Performance
Accuracy	91
Recall	90
Precision	91
F1-score	90

In order to measure the effectiveness of the framework as a classifier, the trained one-dimensional convolutional neural network was applied to the set of 2,000 unseen test records. The precise probabilities of prediction are indicated by the confusion matrix as shown in figure 6.

In this case, the framework correctly identified almost all unseen subjects, with 903 True Positives and 907 True Negatives, making only 91 False Positives and 99 False Negatives. The final results are displayed in Table 2. This system is characterized by outstanding reliability, with an accuracy level of 90.50%, precision is 91% and its F1-Score is 90.48%. Notably, the recall (sensitivity) of this model was 90.12%. In the case of cardiac disease diagnosis, it is important to minimize errors related to false negative results, where the actual patients who are under the threat of adverse events are not detected. One major drawback associated with the usage of deep learning in the medical field is that of "black box," which means the justification of the reason behind a prediction made by the model remains unclear. To promote clinical clarity, feature importance analysis was performed on the 1D CNN as shown in figure 7.

In terms of the results, it was found out that Hypertension, TOBCAT, Diabetes, and BMI were the best indicators of cardiovascular risk within the network structure. The above-mentioned attributes perfectly correspond to the well-known aspects in real-life medicine and physiology of the cardiovascular system. Since the presented framework proves to be based on clinically reasonable biomarkers and not just mathematical constructs, it provides not only accuracy but also physiological transparency, thus making a good basis for a decision support system.

7. Findings

- 1D-CNN can effectively capture feature interactions in tabular data.
- Data synthesis increases the number and quality of datasets.
- State-of-the-art biomarkers greatly increase prediction precision.
- The model exhibits balanced results in terms of all evaluation criteria.
- This method can be applied in clinical decision support systems.

8. Limitations and Research Gaps

While the proposed CNN architecture for TVAE-1D proves to predict highly accurate results, certain improvements need to be made. Firstly, the study applies retrospective cross-sectional data that only reflects the physiological state of patients at a particular point in time. As heart conditions are dynamic and progressive, static data cannot adequately account for the continuous changes taking place in real-time. Secondly, although the Tabular Variational Autoencoder succeeded in addressing the issue of balancing classes, the process of generating additional samples is constrained by the statistical nature of the original data. While being statistically sound, these new samples will not accurately reflect the rare but possible cases that may occur in real-world settings. Lastly, despite its capabilities, the system is single-modal, only applicable to tabular medical data and biomarkers.

The mentioned constraints highlight several critical deficiencies in the research that should be addressed promptly to enhance its possible applicability in a clinical setting. In the first place, to extend the prediction capability from static cross-sectional analysis to a dynamic approach, it is imperative to adapt the 1D CNN model to accommodate time series information. In addition, to address the issue of restricted input data availability, it would be vital to integrate continuous data obtained via IoMT wearable devices into the predictive algorithm. Finally, it would be essential to perform an empirical experiment based on a real-world prospective sample to address the constraints associated with artificial data usage.

Hence, another significant point to consider while furthering the development of this framework would be its validation using databases from various multi-center hospitals, which could be done on the basis of its validity in diverse patient groups in terms of their demographics, genetics, as well as health regulations around the world. Lastly but most importantly, yet another limitation of the current model lies in the absence of multimodal data integration. One way forward in this case could be developing hybrid ensembles that can incorporate both imaging and physiological data in addition to tabular data.

9. Conclusion

This research was able to develop a diagnostic system with automation and machine learning principles for early detection and classification of CVDs. By mitigating the problems associated with clinical data bias, it was revealed that the implementation of data augmentation before classification can help deep learning models operate at a highly efficient level. The novel architecture of TVAE and 1D CNN showed that using artificial intelligence in combination could drastically increase prediction accuracy.

The key strength of the current study is the high sensitivity of the model to diagnose accurately. The model has a high percentage of recall to ensure that there is no false negative prediction that would lead to fatal errors when dealing with patients who need immediate attention. In addition, the current study made a significant contribution to association of computational models with clinical practice. Unlike other computational models that are treated as black boxes, the model used in the current study was based on evidence from medical studies that Hypertension, Tobacco Usage, Diabetes, BMI, Homocysteine, and Lipoprotein(a) were main contributors of CVD risks.

Finally, this study results in a very scalable, objective, and physiologically clear decision-making support system. As it demonstrates the capability of artificial intelligence technology to be highly accurate and medically understandable, this study creates a highly plausible technique for doctors to improve their methods of intervention at an early stage in cardiac treatment.

References

[1] Connie W. Tsao, Alvaro Alonso, Amelia K. Boehme, et al., “Heart Disease and Stroke Statistics—2022 Update: A Report from the American Heart Association”, 145(1), pp. E153-E639, 2022.
[2] Thomas Unger, Claudio Borghi, et al., “2020 International Society of Hypertension Global Hypertension Practice Guidelines”, 75(6), pp.1334-1357, 2020.
[3] Nikolaus Marx, Massimo Federici, Katharina Schutt, et al., “2023 ESC Guidelines for The Management of Cardiovascular Disease in Patients with Diabetes”, European Heart Journal, 44(1), pp. 4043–4140, 2023.
[4] Giuseppina Gallucci, Alfredo Tartarone, Rosa Lerose, Anna Vittoria Lalinga, and Alba Maria Capobianco, “Cardiovascular Risk of Smoking and Benefits of Smoking Cessation”, Journal of Thoracic Disease, 12(7), pp. 3866-3876, 2020.

[5] Paul Ridker, Deepak L. Bhatt, Aruna Das Pradhan, et al., “Inflammation and Cholesterol as Predictors of Cardiovascular Events Among Patients Receiving Statin Therapy: A Collaborative Analysis of Three Randomised Trials”, The Lancet, 401(10384), pp. 1293-1301, 2023.
[6] Sherin Saheera, “Multifaceted role of cardiovascular biomarkers”, Indian Heart Journal, 75(2), pp. 91-97, 2023.
[7] S.S. Habib, T. Al-Khlaiwi, A. Almushawah, A. Alsomali, and S.A. Habib, “Homocysteine as A Predictor and Prognostic Marker of Atherosclerotic Cardiovascular Disease: A Systematic Review and Meta-Analysis”, European Review for Medical and Pharmacological Sciences, 27(1), pp. 8598-868, 2023.
[8] Florian Kronenberg, Samia Mora, Erik S.G. Stroes, et al., “Lipoprotein(A) in Atherosclerotic Cardiovascular Disease and Aortic Stenosis: A European Atherosclerosis Society Consensus Statement”, European Heart Journal, 43(1), pp. 3925–3946, 2022.
[9] Francisco Lopez-Jimenez, Adelaide M. Arruda-Olson, Panithaya Chareonthaitawee, et al., “Artificial Intelligence in Cardiology: Present and Future”, Mayo Clinic Proceedings, 95(5), pp. 1015-1039, 2020.
[10] Chayakrit Krittanawong, Kipp W. Johnson, Robert S. Rosenson, Zhen Wang, Mehmet Aydar, Usman Baber, James K. Min, W.H. Wilson Tang, Jonathan L. Halperin, and Sanjiv M. Narayan, “Deep Learning for Cardiovascular Medicine: A Practical Primer”, European Heart Journal, 40(1), pp. 2058–2073, 2019.
[11] Marzyeh Ghassemi, Tristan Naumann, Peter Schulam, et al., “A Review of Challenges and Opportunities in Machine Learning for Health”, BMJ Evidence-Based Medicine, 25(2), pp. 67-73, 2020.
[12] Lei Xu, Maria Skoularidou, Alfredo Cuesta-Infante and Kalyan Veeramachaneni, “Modeling Tabular Data Using Conditional GAN”, Advances in Neural Information Processing Systems, 32(1), pp. 1-15, 2019.
[13] Serkan Kiranyaz, Onur Avci, Osama Abdeljaber, Turker Ince, Moncef Gabbouj, and Daniel J. Inman, “1D Convolutional Neural Networks and Applications: A Survey”, Mechanical Systems and Signal Processing, 151(107398), pp. 1-21, 2021.
[14] Asha Abraham, Habeeb Shaik Mohideen, and R. Kayalvizhi, “A Tabular Variational Auto Encoder-Based Hybrid Model for Imbalanced Data Classification with Feature Selection”, IEEE Access, 11(1), pp. 122760-122771, 2023.
[15] Mohammad Mahbubur Rahman Khan Mamun, and Tarek Elfouly, “Detection of Cardiovascular Disease from Clinical Parameters using a One-Dimensional Convolutional Neural Network”, Bioengineering,10(796), pp.1-29, 2023.
[16] Amalia Luque, Alejandro Carrasco, Alejandro Martína, and Ana de las Heras, “The Impact of Class Imbalance in Classification Performance Metrics Based on the Binary Confusion Matrix”, Pattern Recognition, 91(1), pp. 216-231, 2019.
[17] Kathleen H. Miao, and Julia H. Miaoa, “Coronary Heart Disease Diagnosis using Deep Neural Networks”, (IJACSA) International Journal of Advanced Computer Science and Applications,
[18] 9(10), pp.1-9, 2018.
[19] Roy S Zawadzki, and Saman Parvaneh, “Synthetic Data Generation in Small Datasets to Improve Classification Performance for Chronic Heart Failure Prediction”, Computing in Cardiology, 50(1), pp. 1-4, 2023.
[20] Tulasi Krishna Sajja, and Hemantha Kumar Kalluri, “A Deep Learning Method for Prediction of Cardiovascular Disease using Convolutional Neural Network”, Revue d'Intelligence Artificielle, 34(5), pp. 601-606, 2020.
[21] Mohamed G. El-Shafiey, Ahmed Hagag, El-Sayed A. El-Dahshan, and Manal A. Ismail, “A Hybrid Bidirectional LSTM and 1D CNN for Heart Disease Prediction”, IJCSNS International Journal of Computer Science and Network Security, 21(10), pp. 135-144, 2021.
[22] Clara Garcia-Vicente, David Chushig-Muzo, Inmaculada Mora-Jimenez, Himar Fabelo,Inger Torhild Gram, Maja-Lisa Lochen, Conceicao Granja, and Cristina Soguero-Ruiz, “Evaluation of Synthetic Categorical Data Generation Techniques for Predicting Cardiovascular Diseases and Post-Hoc Interpretability of the Risk Factors”, Applied. Sci., 13(4119), pp. 1-23, 2023.

[23] Marıa Teresa Garcıa-Ordas, Martın Bayon-Gutierrez, Carmen Benavides, Jose Aveleira-Mata, and Jose Alberto Benıtez-Andrades, “Heart Disease Risk Prediction using Deep Learning Techniques with Feature Augmentation”, Multimedia Tools and Applications, 82(1), pp. 31759–31773, 2023.

[24] Sadia Arooj, Saif ur Rehman, Azhar Imran, Abdullah Almuhaimeed, A. Khuzaim Alzahrani,

[25] and Abdulkareem Alzahrani, “A Deep Convolutional Neural Network for the Early Detection of Heart Disease”, Biomedicines, 10(2796), pp. 1-15, 2022.