**Improving Cardiovascular Risk Prediction through Data Augmentation using VAE-GAN and Deep Neural Networks.**

**Abstract.** This study explores the application of Convolutional Neural Networks (CNN) for predicting cardiovascular disease (CVD) risk, utilizing both real and synthetic data. The VAE-GAN technique was employed to generate a robust dataset of 2 million synthetic data points, enhancing the training process. The original dataset, sourced from IEEE Dataport, comprises 70,000 records with key physiological features. Extensive feature engineering, including the addition of Body Mass Index (BMI), was performed to improve model robustness. The CNN model with 1D convolutional layers demonstrated significant accuracy improvements, achieving a training accuracy of 97.45% and a validation accuracy of 98.03%. The evaluation on the original dataset yielded an accuracy of 92%, highlighting the model's efficacy. This approach underscores the potential of synthetic data in enhancing predictive models and offers promising directions for future research in CVD risk assessment.

**Keywords:** Convolutional Neural Network, Cardiovascular Disease, Data augmentation, Synthetic Data Generation, Feature Engineering, Physiological Data Analysis

1 Introduction

Cardiovascular diseases (CVDs) pose a serious threat to global public health, being the leading cause of mortality and hospitalization worldwide. In 2021, researchers in Mexico attributed 220,000 deaths to CVDs, while the World Health Organization (WHO) estimates that these diseases are responsible for nearly 18 million annual deaths, approximately one-third of all deaths globally.

Events such as heart attacks and strokes are generally acute and mostly occur due to blockages that prevent blood flow to the heart or brain. These blockages are often caused by fat deposits in the walls of the blood vessels that supply these organs. Additionally, strokes can result from bleeding or clots in the brain's blood vessels. Various behavioral risk factors significantly contribute to the development of CVDs, including smoking, unhealthy diets, obesity, physical inactivity, and harmful alcohol consumption. These risk factors can lead to conditions such as high blood pressure, hyperglycemia, hyperlipidemia, and overweight, thereby increasing the risk of severe cardiovascular events.

It is proven that adopting a healthy lifestyle, such as quitting smoking, reducing salt intake, consuming more fruits and vegetables, maintaining regular physical activity, and avoiding excessive alcohol consumption, can significantly reduce the risk of CVDs. In certain cases, pharmacological treatment is also necessary to control diabetes, hypertension, and hyperlipidemia, to prevent heart attacks and strokes.

Furthermore, there are underlying determinants of chronic diseases, known as "the causes of the causes," which reflect broader social, economic, and cultural forces such as globalization, urbanization, and population aging. Other determinants include poverty, stress, and hereditary factors.

Efforts to mitigate the impact of CVDs have traditionally focused on lifestyle interventions and pharmacological treatments. However, the increasing complexity and multifactorial nature of these diseases require innovative approaches to improve outcomes in diagnosis and treatment. Machine learning and artificial intelligence (AI) have emerged as powerful tools, offering the potential to enhance predictive accuracy, personalize treatment plans, and ultimately reduce the burden of CVDs [1].

The field of machine learning plays a crucial role in data analysis and comprehension, offering advantages like trend prediction, error minimization, process automation, and big data management. Recent studies emphasize the importance of machine learning in managing various pandemic outbreaks, enhancing collaboration between healthcare professionals and computerized systems to ensure optimal medical care for diseases such as Ebola and HIV [2,3]. Technological advancements have enabled the incorporation of neural networks in medical practices, particularly in combating cardiovascular diseases (CVD), highlighting the significance of early detection due to their high mortality rate [4].

Research such as [5] showcases the effectiveness of machine learning techniques like SVM, Decision Trees, and Logistic Regression, along with feature selectors like "Boruta," achieving an impressive accuracy of 88.52% in detecting CVD. Similarly, [6] highlights the remarkable accuracy of 99.04% achieved by the random forest algorithm in CVD detection. These results underscore the necessity of choosing the appropriate machine learning algorithm for accurate and meaningful outcomes. Moreover, [7] proposes a disease prediction system employing seven machine learning algorithms to detect and predict diseases based on symptoms. In addition, [8] presents a patient monitoring scheme utilizing an IoT-modified neural network to aid in CVD monitoring.

The study [9] introduces an efficient neural network with convolutional layers designed to classify significantly unbalanced clinical data, focusing on predicting coronary heart disease (CHD). Their approach, featuring a two-layer CNN architecture, achieves a high classification accuracy of 77% for detecting the presence of CHD and 81.8% for correctly identifying the absence of CHD in test data. Additionally, [10] employs machine learning algorithms, including decision trees, random forests, logistic regression, Naïve Bayes, and support vector machines, for accurate prediction and decision-making in CVD patients.

Furthermore, [11] explores the detection and classification of heart disease using IoT and machine learning methodologies in real-time environments. Their proposed RNN classification algorithms demonstrate the highest classification accuracy, reaching approximately 98.50% for both synthetic and real-time datasets. Lastly, [12] contributes to the field by developing an AI-based heart disease detection system using machine learning algorithms. Their random forest classification algorithm achieves an accuracy rate of about 83% on training data.

Today, continuous technological advancement has enabled the implementation of innovations such as neural networks in the medical field, aiding health personnel in combating various diseases, notably cardiovascular diseases (CVD). Due to the high mortality rate of cardiovascular diseases, early detection is paramount [13].

To diagnose such diseases, it is typically necessary to visit a medical center where a specialist conducts various diagnostic tests to reach a conclusion. Given the time-consuming nature of this process, technology involving machine learning algorithms has been developed. Understanding the speed, quantity, and variety of collected data is essential to save as much time as possible in diagnosing and monitoring these physiological parameters. Accordingly, [14] proposes a disease prediction system using seven machine learning algorithms to detect and predict health issues based on symptoms. Similarly, [15] proposes a patient monitoring scheme using a modified IoT-focused deep learning neural network to assist in CVD monitoring.

In [16] an IoT-focused device focused on collecting information before and after an LCD is made, the neural network was trained using deep learning. This technique has an effectiveness of 99.03%, while [17] machine learning techniques with IoT and cloud-related infrastructure are presented. A public healthcare dataset was considered to build a system that enables real-time remote monitoring.

A cloud-focused prediction system aimed at accurately identifying CVD is proposed in [18], utilizing machine learning. This system achieved an accuracy level of 97.53% and a specificity of 97.5%. Additionally, an Arduino-based system was incorporated for patient monitoring, capturing physiological parameters such as body temperature, blood pressure, and heart rate. [19] suggests a method that identifies significant features through machine learning, enhancing the accuracy of CVD prognosis. The main risk factors related to CVD, such as smoking, hypertension, cholesterol, and diabetes, are examined in [20]. It also shows that body mass index (BMI) and systolic pressure are the critical factors most influencing hypertension. [21] establishes a direct relationship between age, gender, BMI, and heart rate with hypertension. Additionally, studies in a general population indicate that high blood creatinine levels can increase the risk of CVD.

Considering a broad range of risk factors complicates CVD prediction, increasing computational costs. Consequently, there is a constant effort to adapt various neural network architectures and machine learning techniques to predict these conditions. [22] compares several machines learning models, including the Multilayer Perceptron (MLP) and the radial basis function model, for predicting CVD in a group of 1245 patients, finding the MLP to be the most efficient, producing an accuracy of 78%. [23,24] propose a hybrid direct selection technique that aims to select smaller subsets to enhance model accuracy by reducing the number of attributes. Techniques such as fuzzy logic, deep learning, and artificial neural networks are reported in [25] to increase model accuracy.

In our study, we explore the integration Variational Autoencoder (VAE) and Generative Adversarial Network (GAN), for synthetic data generation. This approach aims to improve the accuracy of neural networks in predicting cardiovascular risk by addressing data limitations and enhancing the robustness of predictive models. By leveraging these advanced data augmentation techniques, we seek to contribute to developing more reliable and effective tools for assessing CVD risk, thereby helping to reduce morbidity and mortality associated with CVDs.

2. Materials and Methods

The project is based on a well-structured methodology for the detection and assessment of the risk of cardiovascular diseases (CVD) observed in Figure 1, using advanced machine learning techniques. The process begins with the collection of the dataset, specifically the "cardiovascular disease dataset" obtained from IEEE Dataport. This dataset contains critical physiological metrics, categorized into demographics (age, height, weight, gender), medical examinations (systolic pressure, diastolic pressure, cholesterol and glucose levels), and social history (smoking, alcohol consumption, and physical activity). The characteristics are divided into categorical and numerical, with the "cardio" column being the objective variable that indicates the presence or absence of CVD.

In the feature engineering stage, preprocessing of the data is performed, starting with a statistical analysis to better understand the distribution of the dataset. A new column is generated that calculates the Body Mass Index (BMI), adding an additional relevant feature. Next, we get the learning curve of the original dataset using a convolutional neural network (CNN). Based on the results obtained, synthetic data are generated to improve the robustness and generalizability of the model. The validation of these synthetic data, carried out using the VAE-GAN technique, confirms that the data generated maintain the same distribution and behavior as the original data, which is essential to guarantee the reliability of the model.

In the training phase of the model, a CNN architecture is employed that includes multiple convolutional and pooling layers, followed by dense layers. Training is done entirely with synthetic data, distributing 85% for training and 15% for validation. Data preprocessing includes scaling and resizing to ensure compatibility with the CNN architecture.

Finally, the trained and validated model is used to calculate and classify the level of risk of CVD in new patients. This predictive capacity makes it possible to identify those individuals with the highest probability of developing cardiovascular diseases, thus facilitating early intervention and the planning of preventive treatments. The methodology followed in this project, from the selection and preprocessing of the dataset to the generation of synthetic data and the training and evaluation of the model, ensures a rigorous and effective approach to address the problem of cardiovascular diseases using advanced machine learning techniques.

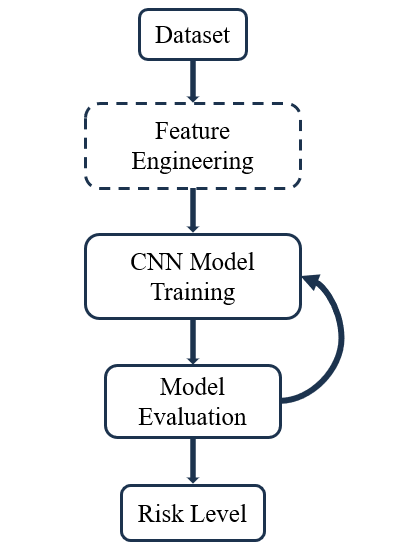


Figure 1. Methodology Flowchart

The detailed methodological workflow ensures a complete and effective approach to addressing the prediction of cardiovascular diseases using advanced machine learning techniques. Every stage, from data collection to risk assessment and calculation, has been meticulously executed to maximize the accuracy and reliability of the model. With the generation and validation of synthetic data complementing the original data, this comprehensive methodology provides a solid foundation to move towards a more detailed analysis of each component, starting with the exploration of the dataset used.

2.1 Dataset

The dataset used in this research, obtained from IEEE Dataport, consists of 70,000 records with 11 independent characteristics. The data were collected during medical examinations and provided by patients. The features incluyen datos demográficos, results of medical examinations and social habits of patients. The target variable, "cardio", indicates the presence (1) or absence (0) of cardiovascular disease.

Dataset features:

* Demographics:
* Age: Modified to show age in years instead of days
* Height: Measured in centimeters
* Weight: Measured in Kilograms
* Medical examinations:
* Systolic Pressure (ap\_hi): Measured in mmHg
* Diastolic pressure (ap\_lo): Measured in mmHg
* Cholesterol: Rated 1 (normal), 2 (above normal), 3 (well above normal)
* Glucose (Gluc): Rated as 1 (normal), 2 (above normal), 3 (well above normal)
* Historial social:
* Alcohol consumption (alcohol): 0 (No), 1 (Yes).
* Smoking: 0 (No), 1 (Yes).
* Physical Activity (active): 0 (No), 1 (Yes).
* Gender: 1 (Female), 2 (Male).

Descriptive Statistics:

Descriptive statistics from the original dataset, detailed in Table 1, reveal key information about the distribution and variability of the features. For example, the average age of patients is approximately 53 years old, with a range of 29 to 65 years old. The average height is 165.86 cm, and the average weight is 75.96 kg. The systolic and diastolic pressures have means of 126.18 mmHg and 81.34 mmHg, respectively. The distribution of cholesterol and glucose levels shows that most patients have normal levels, but there is a significant presence of elevated levels. Drinking and smoking habits are less common in the dataset, with 9.68% and 5.75% of patients reporting these habits, respectively. Most patients (80.50%) report being physically active. The dataset is slightly unbalanced, with 54.74% of the records indicating absence of cardiovascular disease and 45.26% indicating its presence.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Statistics** | **Age** | **Height** | **Weight** | **Ap\_hi** | **Ap\_lo** |
| **Mean** | 52.93 | 165.86 | 75.96 | 126.18 | 81.34 |
| **Std** | 6.77 | 6.53 | 13.19 | 14.62 | 8.82 |
| **min** | 29 | 155 | 56 | 70 | 55 |
| **max** | 65 | 185 | 143 | 160 | 130 |

Table 1. Dataset Descriptive Statistics

2.2 Feature Engineering

Feature engineering is a critical step in machine learning that involves transforming raw data into meaningful features that enhance model performance. This process includes creating new features, modifying existing ones, and selecting the most relevant attributes to improve the robustness and predictive power of the dataset.

Initially, we created an additional column for the dataset: the Body Mass Index (BMI). This feature was derived from the weight and height of the individuals and serves to provide a more comprehensive understanding of the patient’s health status. The BMI is calculated using the formula:



Adding this feature is crucial as BMI is a well-known indicator of overall health and a significant factor in cardiovascular risk assessment. By integrating BMI, we aim to improve the model's ability to capture the relationship between a patient's physical attributes and their cardiovascular health, thereby enhancing the dataset's robustness.

2.2.1 Learning Curve Analysis

Learning curves are essential for understanding how the performance of a machine learning model improves as it is exposed to more training data. They provide insights into whether a model is overfitting or underfitting and help in deciding whether more data could enhance model performance. In this study, we performed a learning curve analysis using a Convolutional Neural Network (CNN) to evaluate its performance on varying sizes of the dataset.

Steps Involved:

* **Data Loading and Preprocessing:**
* The original and synthetic datasets were loaded.
* Features and labels were extracted and scaled using StandardScaler.
* The data was reshaped to fit the input requirements of CNN.
* **Model Definition:**
* A CNN model was defined with several convolutional and dense layers, using the Adam optimizer and binary cross-entropy loss function.
* **Incremental Training:**
* The dataset was divided into increments, and the model was trained and validated at each increment.
* Metrics such as training loss, validation loss, training accuracy, and validation accuracy were recorded.
* **Visualization:**
* Learning curves for loss and accuracy were plotted against the sample sizes to evaluate the model's performance with increasing data.

The results showed that as the sample size increased, both training and validation losses decreased, while accuracies improved, indicating that the model benefits from more data and generalizes well as we can see in figure 2.

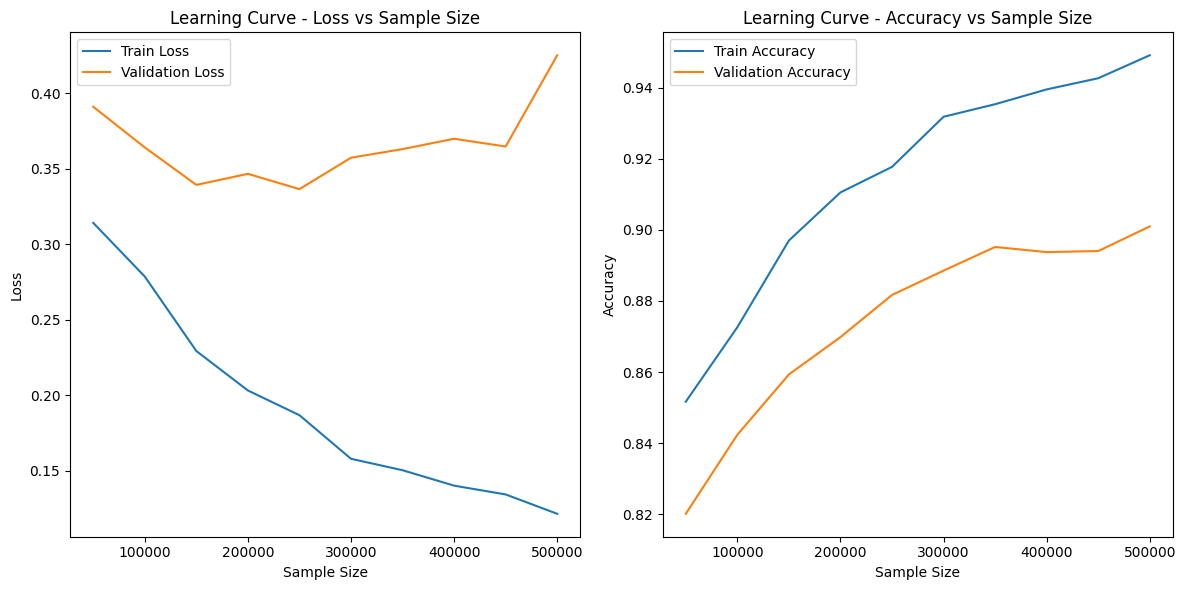


Figure 2. CNN Learning Curve

2.2.2 Synthetic Data Generation & Validation

To address data scarcity and enhance model robustness, synthetic data was generated using a Variational Autoencoder-Generative Adversarial Network (VAE-GAN). This technique combines the probabilistic representation learning of VAEs with the adversarial training of GANs to generate realistic synthetic data.

**VAE-GAN Technique**

Variational Autoencoder (VAE):

A Variational Autoencoder (VAE) is a type of neural network that learns a compressed representation of the input data (latent space) and then reconstructs the input from this latent representation. The VAE consists of two main components:

* Encoder: Maps the input data to a latent space, producing a mean and a standard deviation for the latent variables.
* Decoder: Reconstructs the input data from the latent variables sampled from the distribution defined by the mean and standard deviation.

The VAE introduces a regularization term in the loss function that forces the learned latent space to follow a Gaussian distribution. This regularization enables the model to generate new samples by sampling from the latent space.

**Generative Adversarial Network (GAN)**

A Generative Adversarial Network (GAN) comprises two competing neural networks:

* Generator: Creates synthetic data samples from random noise.
* Discriminator: Evaluates whether the samples are real (from the original dataset) or fake (generated by the generator).

The generator and discriminator are trained simultaneously in a minimax game: the generator aims to produce samples indistinguishable from real data, while the discriminator strives to correctly classify real and fake samples.

**Combining VAE and GAN**

The VAE-GAN architecture leverages the advantages of both VAEs and GANs:

* The VAE provides a structured latent space and generates initial synthetic data samples.
* The GAN refines these samples, ensuring they closely resemble real data.

The combined approach enables the generation of realistic synthetic data with the following steps:

* The VAE encoder compresses the input data into a latent representation.
* The latent representation is fed into the VAE decoder to reconstruct the input.
* The GAN generator refines the VAE-generated samples.
* The GAN discriminator evaluates the refined samples, guiding the generator to improve.

**Data Augmentation**

Using VAE-GAN for data augmentation is analogous to image data augmentation, where new variations of the data are generated to enhance model training. This approach generates new scenarios and combinations rather than entirely new data points, thus improving the model's robustness and generalization by providing diverse training examples. Based on the observations from the learning curve analysis, it was deduced that increasing the dataset size would enhance the model's performance.

In this study, the VAE-GAN technique was used to generate a dataset of 2 million synthetic data points. The synthetic data was validated by comparing its statistical properties with the original dataset. The comparison revealed that the synthetic data maintained the same distribution characteristics as the original data, ensuring its reliability for training machine learning models. This approach is detailed in our previous publication presented at CoDIT24, titled "Analysis of Physiological Parameters for Assessing the Risk Level of Cardiovascular Diseases Using Machine Learning Algorithms". Table 2 shows the comparison of metrics between the original and synthetic datasets [26].

**Table 2. Comparison of metrics between original and synthetic datasets**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metric** | **Original** | | **Synthetic** | |
| Healthy | Disease | Healthy | Disease |
| **Precision** | 69% | 76% | 70% | 76% |
| **Recall** | 81% | 63% | 80% | 63% |
| **F1-Score** | 75% | 69% | 75% | 69% |
| **Accuracy** | 72% | | 72% | |

**Table 3. Comparison of mean and std between original and synthetic**

**Datasets for the most relevant features**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Feature** | **Original** | | **Synthetic** | |
| Mean | Std | Mean | std |
| **Systolic Pressure** | 125.66 | 14.3123 | 125.68 | 14.2888 |
| **Age** | 53.09 | 6.7470 | 53.08 | 6.7473 |
| **Weight** | 75.36 | 13.1386 | 75.43 | 13.1647 |

Upon examining the results in Table 2, it is evident that the metrics for precision, recall, F1-score, and accuracy are very similar between the original and synthetic data, for both "healthy" and "disease" classes. This indicates that the synthetic data generator produces reliable data that accurately reflects the characteristics of the original dataset [26].

Table 3 provides a comparison of the mean and standard deviation between the original and synthetic datasets for the most relevant features. The results show that the synthetic data closely align with the original data in terms of systolic pressure, age, and weight, further supporting the reliability and validity of the generated synthetic data. [26].

In conclusion, the generation and validation of synthetic data through VAE-GAN has proven to be a valuable tool for enhancing the robustness and generalization of machine learning models used in the analysis of cardiovascular diseases. This strategy allows us to overcome the limitations of the original data, providing an expanded and diversified dataset for training more accurate and reliable models [26].

3 Results and Discussion

**Convolutional Neural Networks for Tabular Data**

**Convolutional Neural Networks (CNNs):**

Convolutional Neural Networks (CNNs) are a class of deep neural networks commonly used for processing structured grid data, such as images. They consist of convolutional layers that apply filters (kernels) to the input data, capturing spatial hierarchies and patterns by sliding the filters over the input. This technique allows CNNs to automatically and adaptively learn spatial hierarchies of features from low to high levels.

**Application of CNNs to Tabular Data:**

Although CNNs are typically associated with image data, they can also be effectively applied to tabular data using 1D convolutional layers. In this context, each feature of the tabular data is treated as a channel, and the 1D convolutions help in capturing the relationships between adjacent features. This method can enhance feature extraction and improve the model's ability to detect complex patterns within the data.

**Characteristics of the Model Used:**

In our study, we utilized a CNN model with the following characteristics:

* Input Layer: Receives the input features, including age, height, weight, IMC, ap\_hi, ap\_lo, cholesterol, gluc, alco, smoke, active, and gender.
* Convolutional Layers: The model includes several convolutional layers with 128 filters each and a kernel size of 3. Each convolutional layer is followed by a ReLU activation function to introduce non-linearity and max-pooling layers to downsample the data.
* Dense Layers: Fully connected layers with 128 neurons each, followed by ReLU activation functions, are used to integrate the features extracted by the convolutional layers.
* Output Layer: The final layer is a single neuron with a sigmoid activation function for binary classification (predicting the presence or absence of cardiovascular disease).

**Data Division and Training:**

The synthetic dataset of 2 million data points was divided into training and validation sets, with 85% of the data used for training and 15% for validation. The model was trained for 10 epochs with a batch size of 32, using the Adam optimizer and binary cross-entropy loss function.

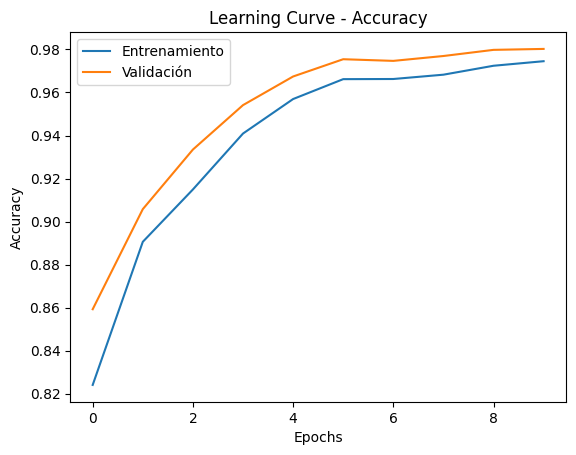


Figure 3 CNN Synthetic Data Learning Curve

Figure 3 presents the learning curves for the CNN trained with the synthetic data, showing the accuracy progression for both training and validation datasets over 10 epochs. Initially, both the training and validation accuracy increase rapidly, indicating effective learning and improvement in performance. Around the fifth epoch, the accuracy begins to plateau, suggesting the model is approaching its optimal performance.

The training accuracy continues to improve slightly, reaching a maximum precision of 0.9745 (97.45%), while the validation accuracy stabilizes around a maximum precision of 0.9803 (98.03%). This minimal gap between training and validation accuracy indicates that the model generalizes well to unseen data, suggesting it effectively captures the underlying patterns without overfitting.

**Evaluation on Original Data:**

After training the CNN with the synthetic data, the model was evaluated using the original dataset. The classification report showed high precision, recall, and F1-scores for both classes, indicating that the model performed well in distinguishing between healthy and disease cases. Specifically:

**Table 4. CNN Original Data Classification Report**

|  |  |  |
| --- | --- | --- |
| **Metric** | **Original Dataset** | |
| Healthy | Disease |
| **Precision** | 92% | 93% |
| **Recall** | 94% | 90% |
| **F1-Score** | 93% | 91% |
| **Accuracy** | 92% | |

Table 4 summarizes the performance metrics of the CNN when evaluated on the original dataset after being trained on the synthetic data. The metrics include precision, recall, F1-score, and overall accuracy for both healthy and disease classes.

4 Conclusions

In this study, we explored the application of a Convolutional Neural Network (CNN) to predict cardiovascular disease (CVD) using a combination of real and synthetic data. We employed the VAE-GAN technique to generate a robust dataset of 2 million synthetic data points, enhancing the training process and ultimately improving the model's performance. The synthetic data was validated to ensure it maintained the statistical properties of the original dataset, confirming its reliability for model training.

The learning curve analysis revealed that the CNN model benefits significantly from the large synthetic dataset, showing substantial improvements in both training and validation accuracy. The model demonstrated excellent generalization capabilities, achieving high precision, recall, and F1-scores on the original dataset. The overall accuracy of 92% indicates that the model is effective in distinguishing between healthy and diseased patients.

The results underscore the potential of synthetic data generation techniques, like VAE-GAN, in addressing data scarcity and improving model robustness in medical applications. The successful application of a CNN for tabular data further highlights the versatility of deep learning architectures beyond traditional image processing tasks.

Future research can explore other data augmentation techniques to further enhance the diversity and robustness of synthetic datasets. Techniques such as GAN variations, SMOTE, or combinations thereof can be investigated. Further optimization of the CNN architecture could be explored, including hyperparameter tuning, alternative activation functions, and different optimizer algorithms to enhance model performance. More advanced feature engineering techniques, such as feature selection and extraction, could be applied to identify the most relevant features and reduce the dimensionality of the dataset, potentially improving model accuracy and efficiency. Applying the trained model to other CVD datasets from different populations could validate its generalization capabilities and robustness across diverse patient demographics. Combining the CNN with other machine learning models, such as ensemble methods or hybrid models, could be explored to leverage the strengths of different algorithms and further improve prediction accuracy. Developing a real-time prediction system for clinical settings, integrating the trained model with healthcare information systems, could provide immediate risk assessments and support clinical decision-making. Enhancing the explainability and interpretability of the CNN model is crucial for clinical adoption. Techniques such as SHAP values or LIME could be applied to provide insights into the model's decision-making process, ensuring transparency and trustworthiness in clinical practice.

In conclusion, this study demonstrates the viability of using synthetic data and deep learning models for CVD prediction, paving the way for future advancements in this critical area of healthcare.

5. References

1. World Health Organization. "Cardiovascular Diseases (CVDs)" WHO, 2021. <https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)>.

2. Kumar, P. M., \& Gandhi, U. D. (2018). A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases. Computers \& Electrical Engineering, 65, 222-235.

3. Karayılan, T., & Kılıç, Ö. (2017, October). Prediction of heart disease using neural network. In 2017 international conference on computer science and engineering (UBMK) (pp. 719-723). IEEE.

4. Agrebi, S. y Larbi, A. (2020). Uso de inteligencia artificial en enfermedades infecciosas. Inteligencia artificial en salud de precisión, 415 438. doi:10.1016 / b978-0-12-817133-2.00018-5

5. Manikandan, G., Pragadeesh, B., Manojkumar, V., Karthikeyan, A. L., Manikandan, R., \& Gandomi, A. H. (2024). Classification models combined with Boruta feature selection for heart disease prediction. Informatics in Medicine Unlocked, 44, 101442.

6. Ansari, G. A., Bhat, S. S., Ansari, M. D., Ahmad, S., Nazeer, J., \& Eljialy, A. E. M. (2023). Performance Evaluation of Machine Learning Techniques (MLT) for Heart Disease Prediction. Computational and Mathematical Methods in Medicine, 2023.

7. Sanjay, J. P., Tummalapalli Naga Deepak, and M. Manimozhi. "Prediction of Health Problems and Recommendation System Using Machine Learning and IoT." 2021 Innovations in Power and Advanced Computing Technologies (i-PACT). IEEE, 2021.

8. Sarmah, S. S. (2020). An efficient IoT-based patient monitoring and heart disease prediction system using deep learning modified neural network. Ieee access, 8, 135784-135797.

9. Dutta, A., Batabyal, T., Basu, M., \& Acton, S. T. (2020). An efficient convolutional neural network for coronary heart disease prediction. Expert Systems with Applications, 159, 113408.

10. Khan, A., Qureshi, M., Daniyal, M., \& Tawiah, K. (2023). A Novel Study on Machine Learning Algorithm-Based Cardiovascular Disease Prediction. Health \& Social Care in the Community, 2023.

11. Patil, R. S., Arjariya, T., \& Gangwar, M. (2023). Detection of Cardiac Abnormalities and Heart Disease Using Machine Learning Techniques. International Journal of Intelligent Systems and Applications in Engineering, 11(5s), 598-605.

12. Chang, V., Bhavani, V. R., Xu, A. Q., \& Hossain, M. A. (2022). An artificial intelligence model for heart disease detection using machine learning algorithms. Healthcare Analytics, 2, 100016.

13. Venkatesan, M., Lakshmipathy, P., Vijayan, V., & Sundar, R. (2022). Cardiac disease diagnosis using feature extraction and machine learning based classification with Internet of Things (IoT). Concurrency and Computation: Practice and Experience, 34(4), e6622.

14. Sanjay, J. P., Tummalapalli Naga Deepak, and M. Manimozhi. "Prediction of Health Problems and Recommendation System Using Machine Learning and IoT." 2021 Innovations in Power and advanced Computing Technologies (i-PACT). IEEE, 2021.

15. Sarmah, S. S. (2020). An efficient IoT-based patient monitoring and heart disease prediction system using deep learning modified neural network. Ieee access, 8, 135784-135797.

16. Al-Makhadmeh, Z., & Tolba, A. (2019). Utilizing IoT wearable medical device for heart disease prediction using higher order Boltzmann model: A classification approach. Measurement, 147, 106815.

17. Kaur, P., Kumar, R., & Kumar, M. (2019). A healthcare monitoring system using random forest and internet of things (IoT). Multimedia Tools and Applications, 78, 19905-19916.

18. Nashif, S., Raihan, M. R., Islam, M. R., & Imam, M. H. (2018). Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system. World Journal of Engineering and Technology, 6(4), 854-873.

19. Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. IEEE access, 7, 81542-81554.

20. Muniasamy, A., Begum, A., Sabahath, A., & Karunakaran, G. (2024). Coronary heart disease classification using deep learning approach with feature selection for improved accuracy. Technology and Health Care, (Preprint), 1-17.

21. Fava, A., Plastino, M., Cristiano, D., Spanò, A., Cristofaro, S., Opipari, C., … Bosco, D. (2013). Insulin resistance possible risk factor for cognitive impairment in fibromialgic patients. Metabolic Brain Disease, 28(4), 619–627. doi:10.1007/s11011-013-9421-3

22. Kurt, I., Ture, M., & Kurum, A. T. (2008). Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. Expert Systems with Applications, 34(1), 366–374. doi:10.1016/j.eswa.2006.09.004

23. Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., & Sun, R. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mobile Information Systems 21 pages. doi: 10.1155/2018/3860146

24. Kahramanli, H., & Allahverdi, N. (2008). Design of a hybrid system for the diabetes and heart diseases. Expert Systems with Applications, 35 (1-2), 82–89. doi: 10.1016/ j.eswa.20 07.06.0 04 .

25. Venkatesh, B. A., et al. (2017). Cardiovascular event prediction by machine learn- ing. The multi-ethnic study of atherosclerosis. Circulation Research, 121 (9), 1092–1101. doi: 10.1161/CIRCRESAHA.117.311312 .

26. López-Saynes, J., Escobar-Gómez, E. N., Velázquez Trujillo, S., Venturino de Coss Pérez, C., Marroquín-Cano, S. F., Chandomi-Castellanos, E., & Hernández-Gutiérrez, C. A. (2024). Analysis of Physiological Parameters for Assessing the Risk Level of Cardiovascular Diseases Using Machine Learning Algorithms. Presented at the 2024 IEEE International Conference on Control, Decision and Information Technologies (CoDIT24).