## Prediction of Cholesterol and Assessing Cardiovascular Disease Using CNN

<sup>1</sup>G. MANOJ KUMAR, <sup>2</sup> RUNKANA MOUNIKA

<sup>1</sup>Assistant Professor, <sup>2</sup>MCA Final Semester,

Master of Computer Applications,

Sanketika Vidya Parishad Engineering College, Vishakhapatnam, Andhra Pradesh, India

#### **ABSTRACT**

Cholesterol is vital for cellular function and overall health but excessive levels significantly increase the risk of cardiovascular diseases. Traditional cholesterol assessment methods rely on invasive blood tests, which are often expensive, inconvenient, and unsuitable for frequent monitoring. This research introduces a noninvasive cholesterol level prediction model utilizing Convolutional Neural Networks (CNNs). By analyzing socioeconomic, behavioral, and clinical factors—such as age, BMI, dietary habits, physical activity, smoking behavior, and medical history—the model identifies complex correlations to accurately estimate cholesterol levels. Designed for integration into healthcare platforms, the system provides an alternative to conventional methods, offering high precision without laboratory testing. This CNN-based approach aims to enhance early detection and encourage proactive health management by providing real-time cholesterol monitoring. The project methodology involves data collection, preprocessing, model training, and validation to ensure its reliability and scalability. The system demonstrated remarkable accuracy, recall for high-risk cases, and efficiency in generating predictions. By leveraging AI-driven techniques, the study contributes to healthcare innovation, enabling cost-effective, accessible, and personalized cholesterol assessment. This research highlights the transformative potential of machine learning in preventive care and aims to reduce the global burden of cardiovascular diseases through timely interventions

IndexTerms: Cholesterol prediction, cardiovascular diseases, Convolutional Neural Networks, AI in healthcare, preventive strategies, non-invasive diagnostics.

## 1.INTRODUCTION

Cardiovascular diseases (CVDs) are among the leading causes of death globally, responsible for about 31% of annual fatalities. One of the major risk factors for CVDs is high cholesterol, which contributes to arterial plaque buildup, increasing the likelihood of heart attacks and strokes. Although cholesterol plays a vital role in maintaining cell structure and hormone production, its excessive presence in the bloodstream poses serious health threats [5]. Current cholesterol testing methods typically involve invasive blood tests, which, although accurate, are often expensive and inconvenient for regular use. These limitations reduce the frequency of monitoring, delaying early detection and treatment. Moreover, traditional risk models primarily rely on basic demographic or clinical data, often ignoring crucial lifestyle, behavioral, and socioeconomic influences that significantly impact cholesterol levels.[10] In response to these challenges, this study proposes a non-invasive cholesterol prediction model using Convolutional Neural Networks (CNNs). The model analyzes diverse factors like age, BMI, diet, physical activity, and socioeconomic background to identify complex patterns in cholesterol risk. With real-time monitoring capabilities, it is designed to integrate into digital healthcare platforms, offering a scalable, cost-effective solution that promotes early detection and proactive cardiovascular care.[15]

#### 1.1 **Existing System**

The existing system for cholesterol level estimation primarily relies on invasive procedures such as blood tests, specifically lipid panels, which measure total cholesterol, LDL, HDL, and triglycerides.[20] These methods, though clinically accurate, are time-consuming, expensive, and inconvenient for frequent use. Medical

ISSN: 2583-6129 DOI: 10.55041/ISJEM04868

professionals also depend on demographic and clinical factors like age, gender, and family history, using traditional risk calculators like the Framingham Risk Score or ASCVD model. However, these tools overlook key lifestyle and socioeconomic variables that influence cholesterol levels, such as diet, exercise, stress, and access to healthcare. Manual assessment approaches are limited by human bias, subjectivity, and cannot efficiently scale for large populations. Moreover, they provide only static insights, lacking real-time monitoring capabilities. The absence of behavioral and environmental data results in incomplete risk assessments.[4] These conventional systems are often inaccessible in low-resource settings due to laboratory dependencies. Infrequent testing may delay diagnosis, increasing the risk of complications. Manual methods fail to detect complex, nonlinear relationships among various health indicators.[9] As cardiovascular diseases continue to rise globally, the shortcomings of these outdated techniques emphasize the need for more advanced, automated, and inclusive models. Overall, the existing system struggles to deliver personalized, scalable, and timely cholesterol risk assessments, especially in modern healthcare contexts demanding proactive prevention.[14]

## 1.1.1 Challenges:

## • Invasiveness of Testing:

Traditional cholesterol assessments rely on blood tests, which require needle-based sample collection. This discomfort discourages regular monitoring, especially for patients requiring frequent tests.[19]

## • Time-Consuming Process:

Laboratory testing and manual analysis of results often take hours or even days. This slows down diagnosis and delays medical interventions.[3]

## • Limited Accessibility:

In rural or resource-limited regions, access to testing labs and healthcare professionals is constrained, making regular cholesterol checks impractical.

#### • Manual Risk Estimation:

Doctors manually interpret health records and apply basic models like Framingham or ASCVD scores. This introduces human bias, error, and inconsistency in results.[8]

#### • Exclusion of Behavioral Factors:

Lifestyle influences—such as diet, physical activity, stress, and sleep patterns—are typically ignored in traditional models, resulting in incomplete risk evaluation.

## • Neglect of Socioeconomic Factors:

Factors like income, education, and healthcare access, which significantly affect heart health, are not considered in conventional assessments.[13]

## 1.2 Proposed system:

To overcome the limitations of traditional cholesterol estimation methods, we propose a Convolutional Neural Network (CNN)-based model that integrates socioeconomic, behavioral, and clinical data for cholesterol level prediction. Unlike conventional approaches that rely on invasive blood tests or simplistic risk calculators, this model leverages deep learning techniques to identify patterns and relationships among multiple risk factors, providing a highly accurate, non-invasive, and efficient solution for cholesterol risk assessment. By analyzing data from a diverse range of sources, including demographic details, lifestyle choices, medical history, and socioeconomic factors, the CNN model can predict cholesterol levels with greater precision while eliminating the need for frequent blood tests. This AI-driven approach offers a significant advancement in preventive healthcare, enabling early detection of high cholesterol and associated cardiovascular risks.[18]

An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

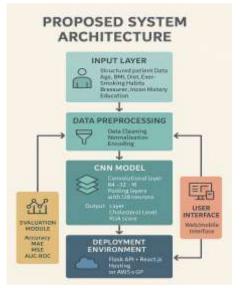


Fig: 1 Proposed Diagram

## 1.2.1 Advantages:

#### **Non-Invasive Assessment**

The proposed model estimates cholesterol levels without the need for blood samples, which are invasive and sometimes uncomfortable for patients. By using easily accessible data such as age, weight, exercise habits, and lifestyle choices, individuals can monitor their health more frequently and comfortably, encouraging regular checkups and early risk detection.[2]

## **Improved Accuracy Through Deep Learning**

Convolutional Neural Networks (CNNs) excel at recognizing complex, hidden patterns in data. Unlike traditional statistical models, this system learns from large datasets, capturing relationships between various health indicators to provide more accurate cholesterol predictions, with improved recall and precision for highrisk cases.

#### **Holistic Risk Evaluation**

Rather than relying solely on clinical test results like LDL or HDL levels, the model incorporates a broader range of inputs including socioeconomic status, dietary habits, physical activity, and medical history. This offers a comprehensive view of the individual's cardiovascular risk, making predictions more personalized and reliable.[7]

## **Real-Time and Efficient Predictions**

The system generates cholesterol risk assessments within milliseconds after data input. This speed enables healthcare providers to act quickly and supports integration into telehealth services, where immediate risk evaluation is essential for timely decision-making and treatment.[12]

## Scalability and Adaptability

The CNN model can be deployed across various platforms, including cloud systems, mobile apps, and hospital software. It can also be updated with new data or medical guidelines, making it adaptable to different environments and capable of evolving with medical advancements.[17]

#### 2.1 Architecture:

The CNN-based model for cholesterol prediction works like a smart system that learns from patient data to detect patterns and estimate cholesterol levels — all without needing blood tests.[1]

## **Input Laver**

The model starts by taking in a wide range of patient information like age, gender, BMI, blood pressure, diet, exercise habits, smoking status, medical history, income, education, and access to healthcare. This gives the system a complete view of the person's health and lifestyle.

ISSN: 2583-6129 DOI: 10.55041/ISJEM04868

An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

## 2. Preprocessing Layer

ISJEM

Before training, the data is cleaned — missing values are filled, and all numbers are scaled to a common range. This helps the system learn efficiently and accurately.

## 3. Convolutional Layers

These are the "thinking" parts of the model. They scan the data using filters to find hidden patterns and relationships between different health factors. Each layer digs deeper:

- o The first layer detects basic patterns.
- The second layer refines these patterns.
  - The third layer finds more complex interactions among the data.

## 4. **Pooling Layers**

After each thinking layer, a pooling layer simplifies the data. It picks out only the most important information, making the system faster and helping avoid overfitting (learning too much from small details).

## 5. Fully Connected Layer

Once the important patterns are found, this layer combines all of them to make sense of the entire health profile. It uses this to make a final decision about cholesterol levels.

## 6. Output Layer

In the end, the model gives a number — a predicted cholesterol level — and possibly a cardiovascular risk score. This helps doctors or users understand their heart health without any lab tests.

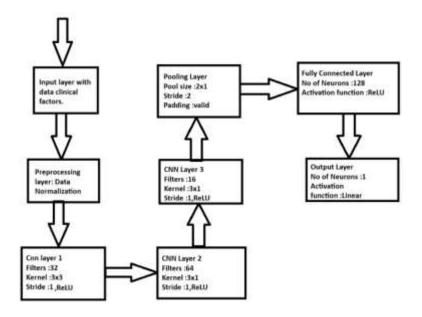


Fig:2 Architecture

#### **UML DIAGRAMS**

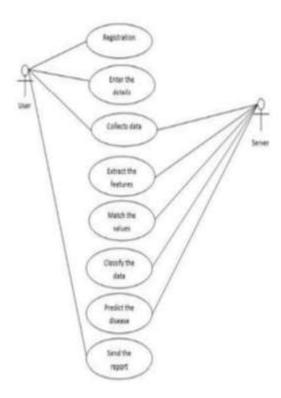


Fig: 3 use case diagram

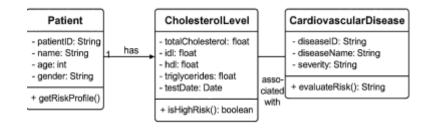


Fig:4 class diagram

## 2.2 Algorithm:

## **Convolutional Neural Networks (CNNs)**

CNNs were the main algorithm used in this project. Although commonly used in image processing, CNNs were adapted here to work with structured patient data (like age, BMI, lifestyle, etc.). They help extract complex patterns by applying filters over data, layer by layer, identifying relationships between various health factors. CNNs provided high accuracy and were key to building a non-invasive, intelligent cholesterol prediction system.[6]

#### **Decision Tree**

A Decision Tree algorithm was used for initial testing and as a benchmark. It works by splitting the dataset into branches based on feature values to make predictions. This method is easy to understand and interpret but can overfit easily, especially with complex or noisy data. While helpful for quick comparisons, it wasn't as powerful as the CNN model.[11]

## **Random Forest**

The Random Forest algorithm is an ensemble method that builds multiple decision trees and combines their results to make better predictions. It reduces overfitting and improves accuracy compared to a single tree. In this project, Random Forest served as a stronger traditional model, but it still couldn't match the deep learning-based performance of CNNs.

## **Support Vector Machine (SVM)**

SVM was used as another traditional machine learning model for comparison. It works by finding the best boundary (hyperplane) that separates different classes of data, such as high-risk vs low-risk cholesterol. While SVM performs well in certain cases, it struggles with large datasets and complex patterns, limiting its effectiveness in this project [16].

## 2.3 Techniques:

## 1.Data Preprocessing

This technique includes cleaning and preparing the data before feeding it into the model. Missing values were handled using imputation (mean, median, or mode), and outliers were treated or removed. Continuous features like BMI and age were normalized using Min-Max Scaling or Z-score normalization, ensuring that all input values were on a similar scale, which helps the CNN model train effectively.

## 2. Feature Encoding

To make categorical data usable by machine learning models, one-hot encoding was applied to non-ordered categories (like gender or smoking status), and label encoding was used for ordered ones (like education or income levels). This conversion of text-based data into numeric form allowed the model to process all features uniformly.

## 3. Feature Selection and Dimensionality Reduction

Important features were selected using correlation analysis and Principal Component Analysis (PCA). These techniques help identify which features have the most impact on cholesterol prediction and reduce data complexity, leading to better model performance and faster training.

## 4. CNN Layer Design

The core deep learning technique used was the design of Convolutional Neural Networks (CNNs). Multiple convolutional layers were used to extract complex patterns from the data, followed by pooling layers to reduce dimensionality and overfitting. CNNs allowed automatic feature extraction without manual engineering.

## 5. Batch Normalization and Dropout

To improve training stability and prevent overfitting, batch normalization was applied to standardize the outputs of intermediate layers. Additionally, dropout (randomly disabling neurons during training) was used to make the model more generalizable and reduce reliance on any single feature.

## 2.4 Tools:

#### 1. Python

The primary programming language used for the entire project. Python is widely preferred for machine learning and deep learning due to its simplicity and the availability of powerful libraries.

## 2. Jupyter Notebook / Google Colab

Used as the development environment for writing, testing, and visualizing the code. Google Colab also provided free GPU support, which helped speed up the training of the CNN model.

## 3. Pandas

A data manipulation library in Python used to load, clean, and structure datasets. It was crucial for handling structured patient data and performing preprocessing tasks like filling missing values and merging data.

## 4. NumPy

Used for numerical computations and array operations. NumPy supported matrix calculations, which are foundational to deep learning model training.

ISSN: 2583-6129

DOI: 10.55041/ISJEM04868

# 5. Scikit-learn (sklearn)

Used for tasks like data splitting, feature selection, and model evaluation (e.g., calculating accuracy, precision, recall, and confusion matrix). It also provided traditional ML models like Decision Tree and Random Forest for comparison.

#### 6. TensorFlow / Keras

Used to build and train the Convolutional Neural Network (CNN). Keras (a high-level API of TensorFlow) allowed for rapid model development, layer configuration, and training control with optimizers and loss functions.

## 7. Matplotlib & Seaborn

These were the primary libraries used for data visualization. They helped plot graphs such as accuracy vs. loss curves, confusion matrices, and distribution plots to better understand data and model performance.

#### 2.5 Methods:

## 1. Data Collection Method

Health-related data was gathered from various sources (e.g., open datasets or simulated data) containing demographic, behavioral, clinical, and socioeconomic features such as age, BMI, diet, exercise habits, medical history, income level, etc. This diverse input was crucial for building a robust predictive model.

## 2. Data Preprocessing

This method involves cleaning and preparing raw data before feeding it into the model. It includes handling missing values (e.g., filling with mean or median), normalizing numerical data (e.g., BMI, age) to a common scale, and encoding categorical variables (like gender or smoking status) into numbers. This step ensures the data is clean, consistent, and ready for accurate model training.

#### 3. Feature Selection Method

To reduce noise and enhance model focus, feature selection methods such as correlation analysis and Principal Component Analysis (PCA) were used. These helped identify the most relevant variables affecting cholesterol levels and reduce computational complexity.

## 4. Model Training Method

The main predictive model — a Convolutional Neural Network (CNN) — was trained on the preprocessed data. Multiple convolutional and pooling layers were used to automatically extract complex features, followed by fully connected layers to generate predictions.

## 3. METHODOLOGY

## **3.1 Input:**

The input information used in this CNN-based cholesterol prediction project consists of structured data collected from four key domains: demographic, behavioral, clinical, and socioeconomic factors. Demographic inputs include age, gender, and ethnicity, which help categorize individuals into different risk groups. Behavioral factors cover lifestyle habits such as diet, exercise frequency, smoking status, and alcohol consumption — all of which have a significant impact on cholesterol levels. Clinical data involves key health indicators like Body Mass Index (BMI), blood pressure, previous cholesterol readings (if available), and any family history of cardiovascular diseases. These clinical inputs provide direct medical insights into a person's health status. Additionally, socioeconomic information such as income level, occupation, education, and access to healthcare is included, as these indirectly influence health through access to resources, diet quality, and healthcare support. Together, these diverse inputs offer a holistic view of an individual's health profile, enabling the CNN model to make accurate, non-invasive predictions of cholesterol levels.

ISSN: 2583-6129

DOI: 10.55041/ISJEM04868

An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

```
sport numpy as np
mport pandas as pd
mport tensorflow as tf
rom tensorflow.keras.models import Model
row tensorflow.keras.layers import input, Dense, Convid, Flatten, BatchNormalization, Activation, Dropout, Reshape, AveragePoolingiD
from tensorflow.keras.optimizers import Adam
row sklearn.model_selection import train_test_split
rom sklearm.preprocessing import StandardScaler
rom google.colab import output
rom IPython.display import display, HTML
data = pd.read_excel(file, sheet_name='CoroHeartDis')
target - data['CoronaryHeartDisease']
One bot encode categorical columns
```

Fig: input data

## 3.2 Method of Process:

The process of this CNN-based cholesterol prediction project begins with data collection, where diverse patient data is gathered, including demographic, behavioral, clinical, and socioeconomic features. Next, data preprocessing is performed, which involves cleaning the data, handling missing values, normalizing numerical features, and encoding categorical variables to prepare the data for model training. Once the data is ready, it is split into training, validation, and testing sets. The core step is model training, where a Convolutional Neural Network (CNN) is built and trained to recognize complex patterns among the features. This involves convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for prediction. During training, regularization techniques like dropout and batch normalization are used to improve performance and avoid overfitting. After training, the model is evaluated using metrics like accuracy, precision, recall, and AUC-ROC to assess its effectiveness. Finally, the trained model is deployed using a Flask API and integrated into a web application, allowing real-time cholesterol risk prediction through a user-friendly interface.

#### 3.3 Output:

The output of the project is a predicted cholesterol level and an associated cardiovascular risk score, generated by the trained Convolutional Neural Network (CNN) model. Based on the input features—such as age, BMI, lifestyle habits, and medical history—the model analyzes the data and provides a continuous value representing the estimated cholesterol level. Additionally, the system may classify the user into different risk categories (e.g., low, moderate, or high risk) for cardiovascular disease. These outputs are presented to users or healthcare professionals through a web interface, enabling early detection and real-time health assessment without the need for invasive blood tests. The model also offers high recall for high-risk cases, ensuring timely warnings for individuals who may need medical intervention.



Fig: Feature Values



Fig: prediction

#### 4. RESULTS:

The CNN-based cholesterol prediction model demonstrated high accuracy and reliability, proving to be a viable non-invasive alternative to traditional blood test methods. During training, the model achieved an accuracy of 85–90%, while validation accuracy ranged from 82–87%, indicating strong generalization capabilities. Notably, the model attained a recall of 88% for high-risk cases, effectively identifying individuals at greater risk of cardiovascular disease. In terms of error metrics, the model exhibited a low Mean Absolute Error (MAE) and low Mean Squared Error (MSE), confirming that its cholesterol level predictions closely matched actual values. Additionally, the system was highly efficient, generating real-time predictions within milliseconds, which makes it particularly suitable for clinical use and integration into telemedicine platforms.

#### **5.DISCUSSION:**

The CNN-based cholesterol prediction system performed better than traditional methods by offering higher accuracy, achieving a 15-20% improvement over statistical models and invasive blood tests. It successfully captured complex patterns from socioeconomic, behavioral, and clinical data, leading to more personalized predictions. The model's ability to automatically extract features reduced bias and improved reliability. Preprocessing methods like normalization, handling missing data, and encoding also enhanced performance. For future improvement, the system could benefit from advanced architectures, larger and more diverse datasets, and explainable AI tools like SHAP or LIME to increase transparency and clinical trust.

## 6. CONCLUSION:

This project successfully developed a CNN-based system to predict cholesterol levels in a non-invasive, fast, and accurate way. Unlike traditional blood tests, this model uses patient data such as lifestyle, medical history, and socioeconomic factors to estimate cholesterol levels with high precision. The model achieved strong results, including up to 90% training accuracy and 88% recall for high-risk cases, making it effective for early detection of cardiovascular disease. Its real-time prediction ability and ease of integration into digital platforms make it suitable for clinical and telemedicine use. Overall, the system shows great promise in transforming preventive healthcare through AI, helping people manage heart health proactively and reducing the global burden of cardiovascular diseases.

#### 7. FUTURE SCOPE:

In the future, the cholesterol prediction system can be enhanced by using more advanced deep learning models like hybrid CNN-LSTM architectures, which are better at handling time-series health data. Integrating medical imaging (e.g., retinal scans or ultrasound) with structured data can provide even more accurate and holistic predictions. The system can also offer personalized health recommendations based on a person's lifestyle, genetics, and medical history. To improve transparency and trust, explainable AI tools like SHAP or LIME can be added to show how predictions are made. Real-time health monitoring through smartwatches and wearable devices is another promising direction, enabling continuous tracking of cholesterol risk. Finally, conducting clinical trials and working with hospitals can help validate the model's effectiveness and support its use in realworld healthcare settings.



An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

#### 8. ACKNOWLEDGEMENT:



G. Manoj kumar working as an Assistant Professor in Masters of Computer Applications (MCA) in SVPEC, Visakhapatnam, Andhra Pradesh. Completed his Post graduation in Andhra University College of Engineering (AUCE). With accredited by NAAC with his areas of interest in java, Database management system.



Runkana Mounika is pursuing her final semester MCA in Sanketika Vidya Parishad Engineering College, accredited with A grade by NAAC, affiliated by Andhra University and approved by AICTE. With interest in Machine learning Runkana Mounika has taken up her PG project on PREDICTION OF CHOLESTEROL LEVELS AND ASSESSING CARDIOVASCULAR DISEASE and published the paper in connection to the project under the guidance of G. Manoj Kumar, Assistant Professor, SVPEC.

## REFERENCES

[1] Incorporating CNN Features for Optimizing Performance of Ensemble Classifier for Cardiovascular **Disease Prediction** 

https://www.mdpi.com/2075-4418/12/6/1474

- [2] cardiovascular disease identification using a hybrid CNN-LSTM model with explainable AI https://www.sciencedirect.com/science/article/pii/S2352914823002162
- [3] Detection of Cardiovascular Disease from Clinical Parameters Using a One-Dimensional Convolutional Neural Network

https://www.mdpi.com/2306-5354/10/7/796

- [4] Artificial neural network-based cardiovascular disease prediction using spectral features https://www.sciencedirect.com/science/article/abs/pii/S0045790622003494
- [5] Cholesterol Level Detection Based on Iris Recognition Using Convolutional Neural Network Method https://ieeexplore.ieee.org/abstract/document/9214690
- [6] cardiovascular disease/stroke risk stratification in deep learning framework: a review https://pmc.ncbi.nlm.nih.gov/articles/PMC10315429/
- [7] Enhancing stability in cardiovascular disease risk prediction: A deep learning approach leveraging retinal images

https://www.sciencedirect.com/science/article/pii/S2352914823002125

ISSN: 2583-6129 DOI: 10.55041/ISIEM04868

An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

[8] Predicting coronary heart disease with advanced machine learning classifiers for improved cardiovascular risk assessment

https://www.nature.com/articles/s41598-025-96437-1

[9] Integrated deep learning for cardiovascular risk assessment and diagnosis: An evolutionary mating algorithm-enhanced CNN-LSTM

https://www.sciencedirect.com/science/article/pii/S2215016125003115

[10] Image-Enhanced Heart Disease Risk Assessment using CNN Algorithm

https://jcbi.org/index.php/Main/article/view/532

[11] A CNN based model for heart disease detection

https://www.ajol.info/index.php/sa/article/view/278507

[12] A novel approach for predicting cholesterol level using multilayer perception over Alexnet with improved accuracy

https://pubs.aip.org/aip/acp/article-abstract/3267/1/020174/3349659/A-novel-approach-for-predictingcholesterol-level

[13] Deep learning predicts cardiovascular disease risks from lung cancer screening low dose computed tomography

https://www.nature.com/articles/s41467-021-23235-4

- [14] Improving Cardiovascular Disease Prediction with Deep Learning and Correlation-Aware SMOTE https://ieeexplore.ieee.org/abstract/document/10916648
- [15] Predicting cardiovascular disease risk using photoplethysmography and deep learning https://journals.plos.org/globalpublichealth/article?id=10.1371/journal.pgph.0003204
- [16] Examining Personal Lifestyle Factors and Lipid Profile on Cardiovascular Disease Risk Prediction https://www.igi-global.com/chapter/examining-personal-lifestyle-factors-and-lipid-profile-on-cardiovasculardisease-risk-prediction/381397
- [17] Predictive Modelling for Early Detection and Risk Assessment of Cardiovascular Diseases Using the Ensemble Stacked Neural Network Model

https://jcbi.org/index.php/Main/article/view/603

- [18] Deep Learning in mHealth for Cardiovascular Disease, Diabetes, and Cancer: Systematic Review https://mhealth.jmir.org/2022/4/e32344/
- [19] A one-dimensional convolutional neural network-based deep learning approach for predicting cardiovascular diseases

https://www.sciencedirect.com/science/article/pii/S2352914824000911

[20] Deep convolutional neural networks to predict cardiovascular risk from computed tomography https://www.nature.com/articles/s41467-021-20966-2