I.J. Intelligent Systems and Applications, **2026, 1, 17-27**
Published Online on February 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2026.01.02

Personalized Cardiovascular Risk Reduction: A Hybrid Recommendation Approach Using Generative Adversarial Networks and Machine Learning

Arundhati Uplopwar*
Amity Institute of Information Technology, Amity University, Uttar Pradesh, India
E-mail: walanu8@gmail.com
ORCID iD: https://orcid.org/0000-0001-8792-9004
*Corresponding author

Rashmi Vashisth
Amity Institute of Information Technology, Amity University, Uttar Pradesh, India
E-mail: rvashisth@amity.edu
ORCID iD: https://orcid.org/0000-0001-8590-7557

Arvinda Kushwaha
Department of Data Science, Galgotias College of Engineering and Technology, Greater Noida, India
E-mail: arvindakush@gmail.com
ORCID iD: https://orcid.org/0000-0003-2426-6375

Received: 17 June 2025; Revised: 02 August 2025; Accepted: 14 October 2025; Published: 08 February 2026

Abstract: Cardiovascular disease (CVD) is a leading cause of death worldwide and hence requires early risk assessment and focused preventative measures. The study describes a novel two-phase hybrid approach that combines machine learning-based CVD risk prediction and personalized lifestyle advice. In the first phase, cardiovascular risk is estimated using ensemble classifier that combines Random Forest Classifier, SVM and LR using metal learner trained on the Heart Disease dataset (1000 record, 14 attributes) has excellent predictive accuracy. In the second phase, optimization framework produces lifestyle suggestions that are safe for health within clinically determined parameters, which are enhanced using a hybrid recommendation system that combines content-based and Cluster-based Outcome Analysis. The suggested approach considerably outperformed a baseline of general lifestyle recommendations in a simulated high-risk cohort, exhibiting an average relative risk reduction of [X] % over a 10-year period as determined by the Framingham Risk Score. The suggested approach is made to be validated in future research using external datasets, simulated patient trials, and physician evaluation in order to guarantee clinical relevance This methodology highlights the promise for precision cardiovascular prevention by providing personalized, data-driven lifestyle recommendations.

Index Terms: Machine Learning, Cardiovascular, Meta Learner, Recommendation, GAN.

1. Introduction

The document "Cardiovascular Diseases in India Compared with the United States" examines the burden of cardiovascular disease (CVD) in India with the United States, as well as trends and health-care solutions. It emphasizes India's rising CVD prevalence as a result of ageing and inadequate preventative care, in contrast to the United States, which has witnessed a drop due to successful public health efforts. The report emphasizes the need for India to pursue systemic reforms, increase surveillance, and develop population-level policies to address CVD [1]. The study describes the HOPE-2 prognostic score, which was created to predict long-term major adverse cardiac events (MACE) and all-cause mortality in COVID-19 patients who were hospitalized. It is based on five clinical parameters: age, history of cardiovascular disease, hypertension, troponin levels, and acute renal failure. The score efficiently categories patients into low, middle, and high risk groups, with increasing incident rates in each category. Its performance was evaluated in an external cohort, which revealed high prediction accuracy for long-term cardiovascular outcomes following COVID-19[2]. In one of the study 45-year-old heart transplant recipient developed rapidly progressive cardiac allograft vasculopathy (CAV) after asymptomatic COVID-19 infection, leading to sudden cardiac death. The data indicate that COVID-19 may hasten CAV through endothelial damage and hyper immune response in immune compromised people [3].A study provides a machine learning-driven framework for personalized lifestyle recommendations to minimize cardiovascular disease (CVD) risk, which incorporates Generative Adversarial Networks (GANs) to model complicated interactions between risk factors and a utility function to balance health benefits with individual adherence using random forest classifier on ARIC dataset achieving 5-10 % risk reduction for high risk individuals[4]. The recommendation system provides useful, tailored counsel to patients who are at risk of developing heart disease, opening up new opportunities for healthcare practitioners. The most often used recommendation approaches are content-based, collaborative filtering, and hybrid [5].Use of Artificial intelligence plays a important role to enable most of the individuals to live a disease free lives, As seen in one of the study a Large language model was developed to capture real time information regarding cardiac disease[6].However to achieve a optimal performance for developing personalized lifestyle recommendation model the two key hurdles needs to be addressed, First the life style recommendations for every individual with the progression of cardiovascular disease is difficult to predict and secondly the physiological characteristics of every individual related to lifestyle modifications are quite and strict adherence to the existing recommendations methods to adopt in reality. To address the above challenges, the paper proposed a personalized recommendation method by incorporating a hybrid recommendation method using GAN(generative adversarial network).The contribution of the work done in the paper is done with 1) A classification based prediction model using machine learning pipeline is implemented to predict cardiovascular disease accurately.2)A clinically constrained optimization method is implemented and a novel clinical recommendation framework is developed to suggest the life style modification with high risk individuals. A thorough analysis of the relevant literature in recommender systems and machine learning for healthcare is given in Section 2. The dataset, the ensemble risk prediction model, and the innovative hybrid recommendation framework are all included in the technique that is suggested in Section 3. The experimental data, model performance, and a discussion of the results are shown in Section 4. The work is finally concluded in Section 5, which also outlines possible directions for future research and summarizes the major contributions.

2. Literature Review

The research describes a machine learning-based expert system that provides personalized lifestyle advice to lower 10-year CVD risk. Using the ARIC dataset and a modified k-nearest neighbor (k-NN) technique, the system predicts CVD risk and recommends optimal lifestyle adjustments based on a patient's features and preferences [7]. The research offers a Park Recommendation Model based on User Reviews and Ratings (PRMRR) that combines latent Dirichlet allocation (LDA) topic modeling with collaborative filtering. It collects park attributes from user evaluations, modifies user preferences with rating confidence levels, and predicts unknown ratings using KL divergence-based similarity [8]. The research presents a linear-weighted hybrid algorithm for resource recommendation in social annotation systems that incorporate user, tag, and resource interactions. It shows that this approach outperforms individual and tensor factorization models on six real-world datasets, providing greater adaptability, scalability, and flexibility for both basic and tag-specific recommendation tasks [9]. The study provides a comprehensive assessment of patient perceptions on hybrid mHealth therapies for psychosis that combine digital technologies with in-person therapy. It analyzes significant themes that influence participation, such as the necessity of therapist assistance, enhanced communication, and privacy concerns, and provides recommendations to improve trust and usability in digital mental health systems [10]. The study describes MELENDI, a recommendation system that provides healthcare providers with relevant scientific papers based on diagnoses retrieved from electronic health records (EHRs), specifically discharge summaries. It prioritizes high-quality content by combining semantic similarity and article relevance criteria [11]. The research conducts a systematic literature evaluation of food recommender systems (FRS) by examining 67 high-quality studies from 2,738 applicants. It categorizes recommendation methodologies, data types, and evaluation strategies, emphasizing both benefits (such as individualized, nutrition-aware recommendations) and drawbacks (such as reproducibility concerns and a lack of uniformity). The review provides a solid platform for further FRS development and research [12]. The study introduces SEHAD-HC, a machine learning-based smart wearable system for real-time cardiac arrest monitoring that employs a hybrid computer architecture. The system uses IoT sensors and Android apps to collect physiological signals (e.g., ECG, SpO_2, GSR), classify cardiac risk using Random Forest (with 97% accuracy), and send notifications for early action [13]. The study describes a hybrid attribute-based recommender system for e-learning material recommendation, which blends collaborative filtering with content-based approaches. By combining student preferences and material properties, it strives to improve personalization and alleviate sparsity issues, increasing the accuracy and relevance of instructional content recommendations [14]. The research presents a personalized book recommendation system for college libraries based on a hybrid algorithm that combines collaborative filtering and content-based approaches. Clustering is used to handle data sparsity and cold start difficulties, and it operates on the Spark big data platform for scalability and efficiency [15]. Study describes the SCORE trial, a hybrid Type 1 implementation-effectiveness study aimed at reducing tobacco-related health inequalities among BIPOC populations through improved smoking cessation therapies. It assesses the effectiveness of incorporating culturally adjusted Longitudinal Proactive Outreach (LPO) into the Ask-Advise-Connect (AAC) strategy in increasing abstinence rates and treatment involvement. The trial addresses systemic barriers through motivational interviewing and outreach, with the goal of promoting health equity and improving long-term cessation outcomes [16]. The research describes the design, development, and evaluation of an intelligent web-based recommender system that can help people at risk of non-alcoholic fatty liver disease by making individualized lifestyle and treatment suggestions. Built using C# and SQL Server with a rule-based knowledge system, it achieved a strong usability score (74.6% on SUS), suggesting successful expert approval and user satisfaction [17]. The paper offers a Hybrid Scientific Article Recommendation System with COOT Optimization (HSARCO), which combines content-based filtering, citation graphs, and a Meta heuristic COOT optimization algorithm to improve article relevance while addressing cold-start and data sparsity challenges. Using Word2Vec, LSTM, and citation network analysis, the model improves recommendation accuracy by up to 2.3% in precision, 1.6% in recall, and 5.7% in MRR compared to baseline techniques [18]. The research describes a hybrid diabetic retinopathy (DR) detection framework that blends deep and machine learning techniques. It extracts deep features using DenseNet121, MobileNetV2, and InceptionResNetV2, and improves accuracy by stacking multiple ML classifiers. On the APTOS 2019 dataset, the hybrid model performs well, with 95.50% accuracy for multi-class and 98.36% for binary classification [19]. The study examines the development and implementation of Health Recommender Systems (HRS) that employ Natural Language Processing (NLP) techniques. It examines existing HRS solutions for doctor recommendations, prescription suggestions, chronic disease diagnosis, and personal health records, emphasizing developments such as semantic analysis, hybrid filtering, and machine learning. The report also underlines the importance of sustainability and trust in future healthcare recommendation systems [20].

3. Proposed Research Methodology

The overall framework for proposed method consist of Heart disease dataset with 1000 rows and 14 attributes, a classification model for predicting cardio vascular disease and applying hybrid recommendation technique using generative adversarial network to generate the personalized lifestyle recommendations for cardiovascular disease is shown in Fig. 2. To achieve robust evaluation and reduce over fitting, k-fold cross-validation (k=5) on dataset is applied. The model's performance indicators, including AUC, were averaged over folds to determine consistency. Furthermore, a hold-out test set of 20% of the data is set, stratified to preserve class distribution, for final evaluation.

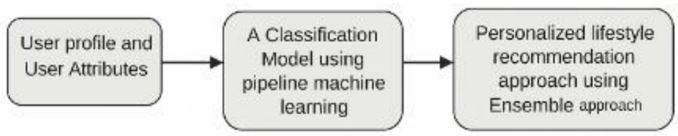

Fig.1. General recommendation system of the proposed method

3.1. Dataset Description

The study makes use of the Heart Disease dataset which is a well accepted benchmark for cardiovascular risk prediction tasks. The dataset includes 1000 patient records, each with 14 clinical variables and a binary target label indicating the presence or absence of heart disease. The input features consist of demographic information, physiological measurements, and diagnostic signs. To aid with prediction and personalized advice, the features were divided into modifiable (lifestyle-related) and non-modifiable (static) categories. Resting blood pressure (restingBP), serum cholesterol (serumcholestrol), fasting blood sugar (fastingbloodsugar), and maximal heart rate reached (maxheartrate), exercise-induced angina (exerciseagina), ST depression (oldpeak), and type of chest pain (chestpain) are all modifiable parameters. These are the key objectives for synthetic lifestyle recommendations since they have been shown to respond to behavioral modifications such as food, exercise, and medications. To retain physiological realism, non-modifiable characteristics such as age, sex, resting ECG findings (restingelectro), ST segment slope (slope), number of major vessels (noofmajorvessels), and thalassaemia status were preserved during the recommendation process. This distinction ensures that the proposed profiles represent achievable interventions while maintaining a participant's essential traits. Feature standardization was used where needed, and all recommendation samples generated by the hybrid filtering method followed these requirements. This thorough feature engineering technique promotes both clinical validity and ethical plausibility in developing personalized lifestyle changes for cardiovascular risk reduction. The dataset is denote

$$\mathrm{D} = \{(\mathrm{x}\ (\mathrm{i}), \mathrm{y}\ (\mathrm{i}))\}\ \mathrm{i} - 1\mathrm{pow}(\mathrm{N})$$

We partition the future vector x into two disjoin sets

$$\mathrm{x(i)} = [(\mathrm{xstatic}\ (\mathrm{i}), \mathrm{xmodifiable}\ (\mathrm{i})]$$

Categorical feature chest pain (cp) was encoded into numerical value. Numerical features were standardized to have a mean of zero and a standard deviation of one. SMOTE technique was applied during 5 fold cross validation to handle class imbalance.

3.2. Experimental Setup and Validation Strategy

To maintain the original class distribution, the entire dataset was initially divided into a hold-out test set that included 20% of the data (200 records).5 fold cross validation was performed on 80% (800 records)on training dataset. Model was trained on 4 fold and validated on remaining 1 fold.

3.3. Classification Model for Risk Assessment

A Random Forest Classifier (RFC) was used to evaluate cardiovascular disease risk using clinical and lifestyle characteristics from the Heart Disease dataset. The classifier was integrated into a hybrid recommendation framework to calculate the anticipated risk before and after making lifestyle changes. By training on real patient records and analyzing synthetic, proposed profiles, the RFC provided robust risk prediction using probabilistic outputs. This approach allowed for an objective assessment of recommendation success, as well as a data-driven system to steer personalized cardiovascular health prevention efforts. In the recommendation process original risk and recommended risk is calculated by using and .The average recommendation risk is

$$\mathrm{yorig} = \mathrm{f(x)\ and\ yrec(j)} = \mathrm{f(xrec(j))}, \mathrm{j} = 1, \ldots, \mathrm{M}$$

$$\mathrm{yrec} = 1/\mathrm{M} \sum \mathrm{yrec(j)}$$

3.4. Generative Adversarial Network

Traditional simulation methods mostly rely on predefined rules and a distribution which fails to depict complex, on linear relationship in lifestyle modification in real world. To address the issue a GAN is chosen over rule based simulation due to its ability to learn changeable attributes and replicates the complex correlations inherent in human physiology generating recommendations which adhere to real-world biological restrictions. The original dataset contains patient's baseline states and their outcome lagging the intervention pathway connecting them, to bridge the gap GAN is used to create synthetic lifestyle profile.

In the proposed method the generator (G) consist of a linear layer that maps the noise vector to hidden layer of size 64, a ReLU activation function for non-linearity and 2nd linear layer that maps hidden representation to final output dimension. The role of discriminator (D) is to classify input data to real data .The discriminator has linear layer, ReLU activation function, A final linear layer produces a single value (probability) indicating whether the input is genuine or not, followed by a Sigmoid activation function.

$$\mathrm{D(x)} = \sigma(\mathrm{Wd} \cdot \mathrm{x} + \mathrm{bd})$$

x is the input data.w.d and b.d are the weights and biases of the discriminator's neural network. Sigma is a sigmoid function with output probability 0 and 1.The generator G and discriminator D is trained and their loss functions are defined as

$$\mathrm{LD} = -\mathrm{E}[\log(\mathrm{D(x)})] - \mathrm{E}[\log(1 - \mathrm{D(G(z))})]$$

The objective of the discriminator is to correctly classify the data. The loss of Generator G is

$$\mathrm{LG} = -\mathrm{E}[\log(\mathrm{D(G(z))})]$$

LG=−E[log(D(G(z)))].Both the models are trained using optimizers to minimize the loss functions. The GAN model design effectively uses a deep learning architecture to produce synthetic data and increase the accuracy of health-related predictions. By using a generator to generate data and a discriminator to validate it, the GAN framework can iteratively improve its performance, eventually producing data that closely reflects the underlying distribution of real patient data. The complete process is explained mathematically as

Generator and discriminator Architecture Input/output

The GAN architecture is made up of two primary components: the Generator (G) and the Discriminator (D). The GAN's purpose is to train the generator to produce synthetic data that is indistinguishable from actual data, while the discriminator learns to discern between the two.

Generator GθG

Input: Noise vector $z \in R^d$ where *dz* is the noise dimension (32).
Output: Synthetic data sample $x\hat{} \in R^d$ matching the feature space of real data.

$x\hat{} = G_\theta(z) = MLP_G(z)$ where MLP_G denotes the multilayer perception parameterized by *θG*. The generator takes the random noise vector as a input and transforms it into synthetic data which is mathematically represented as

$$G(z) = \sigma(Wg \cdot z + b.g)$$

z is the input noise vector (input_dim=32).
w.g and b.g are the weights and biases of the generator's neural network. Sigma is an activation function

Discriminator (D)

Input: Data sample $x \in R^{dx.}$
Output: Scalar probability $p \in [0,1]$ indicating real or unreal
$p = D\theta D(\mathbf{x}) = \sigma(\text{MLP}D(\mathbf{x}))$ where σ is the sigmoid activation, and MLPD is a neural network with parameters θD.

To generate personalized lifestyle profile GAN with static and modifiable features is employed which leans to produce synthetic feature vector, model is exposed to learn from complex non linear mappings between and high and low risk during training phase. Discriminator evaluates realism of synthetic profile against actual data.

GAN training

The goal of the discriminator (D) is to maximize its capacity to accurately identify synthetic data as synthetic (D (G (z)) → 0) and real data as real (D(x) → 1). Generator (G) creates data that is so realistic that D (G (z)) → 1.Loss function is defined as

$$V(\theta G, \theta D) = Ex \sim pdata[\log D\theta DD(x)] + Ez \sim pz[\log(1 - D\theta D(G\theta G(z)))]$$

The discriminator D learns to maximize the likelihood of correctly classifying real and fake data.

Optimization process

Discriminator Update and Generator Update is given by

$$\theta D(t+1) = \text{argmax}\theta DEx \sim pdata[\log D\theta D(x) + Ez\theta D(t+1)$$
$$= \text{argmax}\theta DEx \sim pdata[\log D\theta D(x) + Ez \sim pz[\log(1 - D\theta D\big(G\theta G(z)\big)\big)]$$

$$\theta G(t+1) = \text{argmin}\theta G - Ez \sim pz[\log D\theta D(G\theta G(z))]$$

Synthetic data Generation

After training, the generator $\boldsymbol{G_{\theta G}}$ maps a noise vector $\mathbf{z} \sim p\mathbf{z}$.

$$\mathbf{x}\hat{} = G\theta G(\mathbf{z}), \mathbf{z} \sim N(0, I)$$

3.5. Hybrid Recommendation Method for Health Intervention

Hybrid filtering is used to generate personalized lifestyle recommendations to the patients suffering with cardiovascular disease. Hybrid filtering approach combines two filtering approaches; content based filtering and collaborative filtering**.** Hybrid recommendation works as follows

A. Content based Filtering (CBF)

Patients similarity is computed using cosine similarity on features such as blood pressure, serum cholesterol and maxheartrate and modifiable features of all synthetic profiles generated by GAN(x^j*)*. A future weighting scheme prioritizes significant variables by using content based filtering method.

B. Cluster-based Outcome Analysis

Using cluster based outcome analysis, patients are clustered using K means and identify patients with the similar historical response within the cluster.GAN generated synthetic profile is ranked as per risk reduction with in specific cluster. The preferences of comparable users are not the basis for suggestions in cluster-based outcome analysis. Rather, it pinpoints patterns of interventions that have been scientifically linked to favorable results within distinct physiological groups, guaranteeing that suggestions stay based on patterns of clinical data.

C. Meta-learner Fusion

Recommendations are generated by combining the CBF and Cluster based outcome analysis. A logistic regression meta-learner improves the hybrid output by learning from real-world patient outcomes. Synthetic profiles from the GAN broaden the recommendation pool, ensuring various realistic ideas. The meta-learner computes a final utility score for each proposed synthetic profile. The profile with the greatest utility score is chosen as the personalized lifestyle suggestion for the target patient. The Random Forest Classifier is utilized in the method to estimate patients' cardiovascular risk based on their feature profiles. The anticipated risk is then utilized to provide individual lifestyle recommendations for each patient. The Random Forest Classifier generates feature importance scores, which can be used to determine the most significant factors contributing to the predicted risk.

4. Working Principal of the Model

The complete working of the model consists of two phases, Risk prediction phase and life style recommendation phase.

Phase-1: Risk prediction phase

Models is developed using the Heart Disease dataset. This dataset contains 1000 patient records, 14 clinical characteristics, and a binary outcome label indicating the presence (1) or absence (0) of heart disease. Variables that cannot be changed include age, sex, resting ECG findings (resteclectro), ST segment slope (slope), number of major vessels (majorvessels), and maxheartrate. Modifiable aspects include variables such as resting blood pressure (restingBP), serum cholesterol (serumcholestrol), fasting blood sugar (fastingbloodsugar), maximal heart rate reached (maxheartrate), exercise-induced angina, ST depression (oldpeak), and chest pain type (cp), which can be affected by lifestyle modifications. Three machine learning classifiers Random forest, SVM and linear regression are combined in a pipeline framework to predict the cardiovascular risk on 80% trained dataset. A 5 fold CV is applied to reduce over fitting. SMOTE technique is used to balance minority classes.

Base learners and Meta learner: Three machine learning classifiers are used as base learners Random Forest, Support vector machine with RBF kernel and Logistic Regression with L2 regularization. Predictions from RF, SVM and LR are combined with logistic regression as stacking Meta learner.

Phase-2: Life style recommendation phase

Phase-2: of the model use logistic regression as Meta learner to combine the predictions from the base model to predict the cardiovascular disease. This phase combines the Generative Adversarial Networks (GANs**)** with a Hybrid Filtering Recommendation method**.** GAN which composed of two neural network Generator (G) and discriminator (D) is used to generate synthetic personalized recommendations. Hybrid filtering algorithm then combines the output generated from GAN to suggest lifestyle recommendations to the patients at high risk. Hybrid recommendation method used CBF to calculate the similarity's means clustering is used to group the similar patients and then combined score is used within each cluster.Figure-3 shows the working principle of the model. To reduce over fitting concerns, L2 regularization is used and dropout layers in proposed model. In addition, data augmentation techniques such as synthetic sample creation via SMOTE (Synthetic Minority Over-sampling Technique) were used to balance the dataset and improve generalizability. Conventional simulation techniques and generative models, such as GANs, frequently replicate data distributions without explicitly taking medical limitations into account, which might result in suggestions that are dangerous or physiologically unrealistic. We provide a Constrained Optimization Approach to produce tailored lifestyle suggestions in order to overcome this significant constraint. By following stringent clinical safety guidelines, this approach directly looks for the ideal combination of modifiable characteristics that reduce cardiovascular risk.

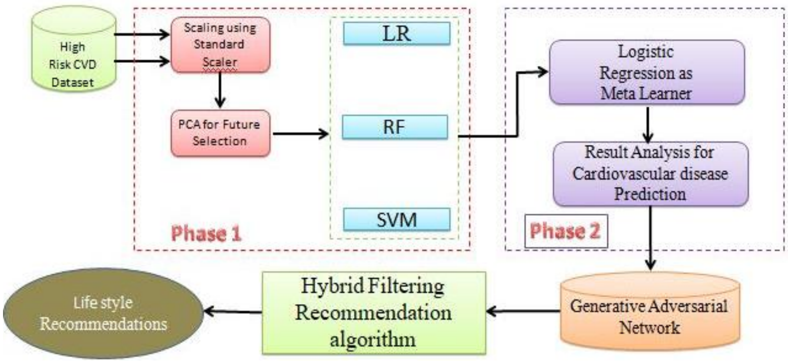

Fig.2. Proposed two-phase Hybrid recommendation for cardiovascular disease

Algorithm
Input: Patient profile with static and customizable elements.
Trained Random Forest Classifier, f ().
Output: Personalized lifestyle recommendations.
Step 1: Initialize Optimization with Hybrid-Guided Parameters.
Train GAN using real patient datasets for each iteration. j = 1 to M DO
Sample random noise vector, z_j.
Generate a synthetic customizable profile: x̂_j → G (z_j).
Combine x̂_j with the patient's static features to create a comprehensive profile.
END OF FOR

Step 2: Risk estimation.
Predict the initial risk: y_orig = f(x) for each synthetic profile. X̂_j DO: Predict risk. y_rec_j → f(x̂_j)
END FOR:
Calculate the average suggested risk.
ȳ_rec = (1/M)*Σ y_rec_j
Step 3: Hybrid Recommendation Filtering.
Perform content-based filtering (CBF).
Determine cosine similarity between x and all x̂_j.
Select the top K most comparable profiles.
Perform

4.1. Risk Reduction Evaluation Method

The post-intervention risk was calculated by re-evaluating the same individual after implementing the recommended lifestyle changes generated by the GAN-based approach. It is calculated as mean relative difference across all participants.

$$\text{Risk Reduction}=P_{baseline}- P_{recommended}/P_{baseline}$$

Where $P_{baseline}$ is computed using trained RF classifier on patient's original features and $P_{recommend}$ is evaluated by applying RFC modified feature vector. The Random Forest classifier's estimated risk output was adjusted to roughly match the 10-year Framingham General Cardiovascular Risk Score. Risk reduction was evaluated on hold out test set.

4.2. Statistical Analysis and Discussion

The model is tested by selecting one patient at random from the test dataset indicating the presence and absence of the disease. The hybrid recommendation function generates the life style recommendations for the selected patient based on the profile and training data. The original predicted risk of selected patient is calculated using random forest classifier. Which is trained on the 80% training dataset? The standard error of mean(SEM) is used to calculate mean predicted risk for the recommended lifestyle profiles along with 95% confidence interval. The magnitude of the difference between the original predicted risk and the mean predicted risk for the recommended lifestyle profiles is done by using effect size measurement. The findings of the study show that GANs and hybrid filtering are excellent in generating personalized lifestyle recommendations for lowering cardiovascular risk. The anticipated cardiovascular risk fell dramatically following the suggestions, and the impact magnitude was high. The lifestyle factor changes found by the hybrid filtering approach were in line with clinical guidelines for cardiovascular disease prevention. The graph in figure 3 depicts GAN training losses over 5000 epochs, with the blue line representing generator loss and the orange line representing discriminator loss. GANs' antagonistic training causes both losses to fluctuate. The discriminator loss often reduces, showing that it is learning to discern between true and bogus data more accurately. The generator loss gradually grows over time, indicating that it is struggling more as the discriminator improves. Although model achieves AUC score near to 0.96, the result was consistent across cross validation and hold out test. All recommendations fall within evidence-based ranges defined by ACC/AHA clinical guidelines and model designed as decision support tool. The study offers a useful framework for a clinical decision-support tool and a novel safety-first approach to integrating recommendation algorithms into the medical field.

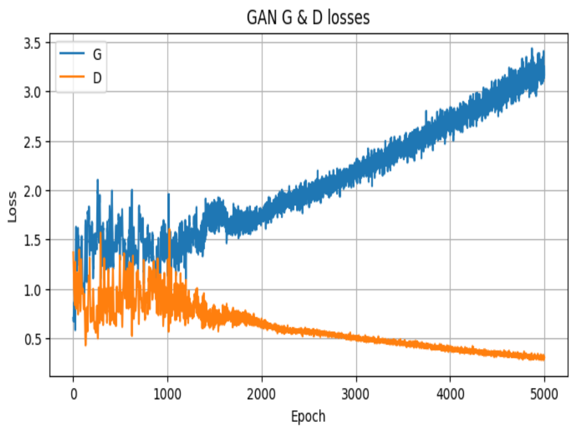

Fig.3. Generator and discriminator losses during GAN training

The figure 4 depicts a PCA-based 2D projection that compares real data (blue) to GAN-generated synthetic data (orange). Both data types overlap significantly near the centre, showing that the GAN has learnt to replicate the distribution of real data.

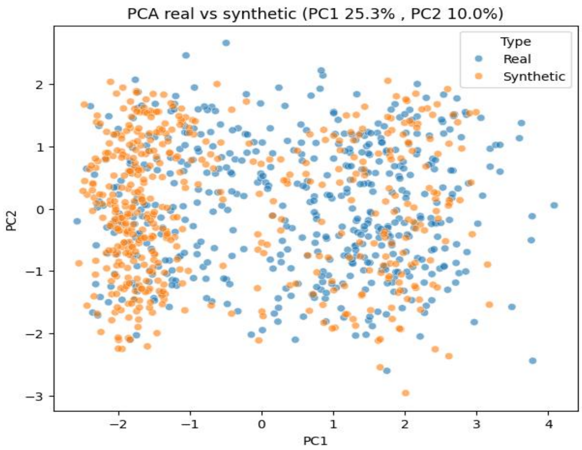

Fig.4. PCA visualization comparing the distribution of real and GAN-generated synthetic data

The image shown in figure 5 compares distribution of resting blood pressure. The majority of suggested patients had blood pressures ranging from 140 to 150mm Hg. This shows that the patient is on the healthier end of the recommended range. The recommended values fall within predefined, clinically safe ranges.

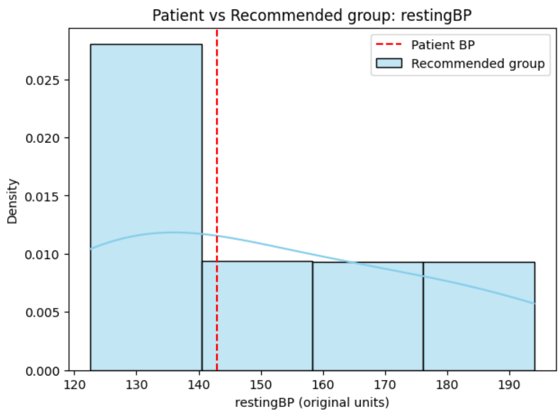

Fig.5. A new patient's resting blood pressure of is compared to the distribution of resting blood pressure in a group of recommended patients

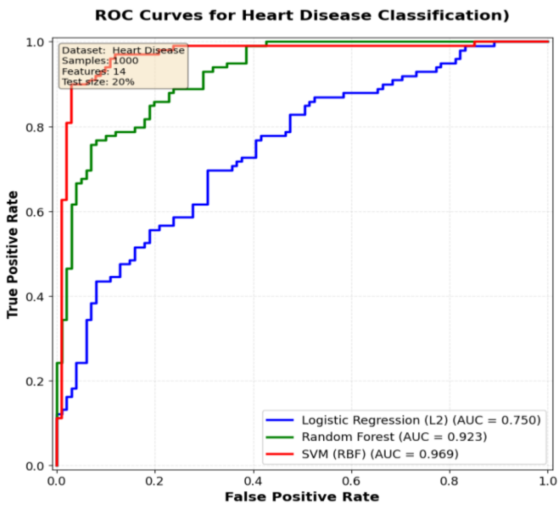

Fig.6. ROC curves comparing the performances of LR, RF and SVM

The plots in figure 6 depict the ROC curves for predicting high blood pressure, cholesterol and heart rate. Among three classifiers logistic regression, random forest, and SVM, SVM model had the best performance, indicating perfect classification. Random Forest and SVM both perform well, with AUCs of 0.92 and 0.96 respectively. For cholesterol both random forest and logistic regression achieved good performance. For high heart rate random forest outperformed both the classifiers. The analysis shown in the Figure 6 is not the core prediction task, a separate analysis is conducted to estimate individual predictive power of key modifiable risk factors for CVD outcome. For modifiable features such as restingBloodpressure, serumcholestrol, and maximum heart rate, three classifiers LR, SVM and RF is trained to predict new binary feature. The resultant AUC scores effectively estimates the high risk value for specific feature as binarization method resulting binary CVD outcome.

The violin plot in figure 7 illustrates how preventive care has enormous potential benefits. The projected risk of cardiovascular events for general population is greatly reduced by altering risk factors (e.g., reducing blood pressure, quitting smoking, and managing cholesterol). According to the model, many high-risk people would fall into a significantly lower risk category if they met the suggested health requirements.

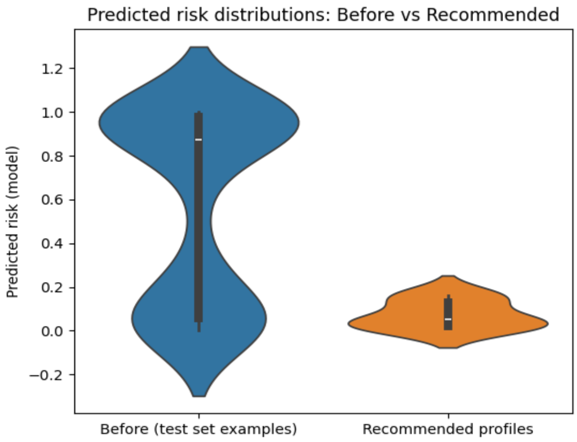

Fig.7. Comparison of predicted cardiovascular risk distributions - baseline assessment vs. recommended lifestyle modifications

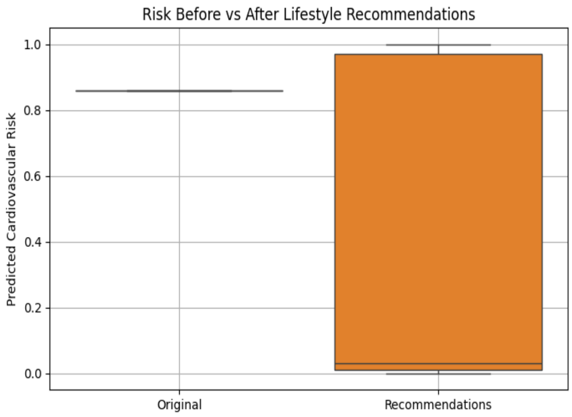

Fig.8. Box plots comparison of predicted cardiovascular risk before and after lifestyle recommendations

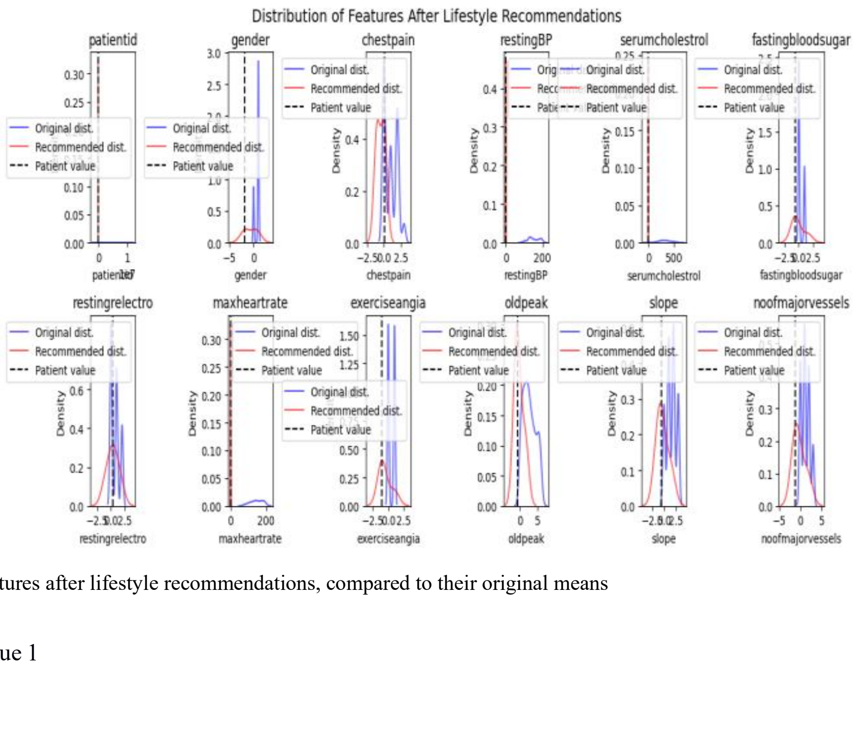

Fig.9. Distribution of features after lifestyle recommendations, compared to their original means

The box plot represented in figure 8 shows the anticipated cardiovascular risk before and after lifestyle recommendations. The "Original" risk values are closely clustered at a high level (~0.85), showing consistent high risk across people. In comparison, the "Recommendations" category has a significantly broader spread, with risk ratings ranging from near 0 to 1.

The figure 9 depicts box plots illustrating the distribution of various medical features after following lifestyle suggestions. Several features such as restingBP, serum cholesterol, fasting blood sugar, maximum heart rate and ST show distribution towards clinically healthier ranges under recommendation scenario.

4.3. Limitations

The model's generalizability across a range of populations may be impacted by the fact that it was trained on a very limited, single-source publically available dataset. The practical efficacy and patient adherence of the reported risk reduction are uncertain because it is based on simulated interventions rather than actual clinical trials. Moreover, the suggestions are limited by the clinical characteristics present in the dataset, leaving out potentially significant environmental and psychological elements. Future research will need to validate its acceptability by healthcare providers and inclusion into clinical workflows. Real-world validation with medical specialists is still necessary for clinical adoption even though our optimization approach imposes clinical constraints.

5. Conclusions and Future scope

This study presented a novel two-phase hybrid framework for individualized cardiovascular disease (CVD) risk prediction and lifestyle recommendations. The first phase used an ensemble model integrated with Random Forest Classifier, SVM and LR with logistic regression as Meta learner to predict cardiovascular risk. The second phase incorporated a Generative Adversarial Network (GAN) to build synthetic, realistic lifestyle modification profiles, as well as a hybrid recommendation system that used both content-based and collaborative filtering methodologies on modifiable risk factors .Experimental and simulation based results revealed that the suggested method dramatically reduced projected cardiovascular risk by statistically confirmed results such as confidence intervals and huge impact sizes. Visualizations such as violin plots, box plots, and PCA projections contributed to the effectiveness and physiological realism of the provided recommendations. The substantial simulated risk reduction attained demonstrates how this methodology has the potential to influence precision cardiology by offering a scalable, evidence-based preventative care tool. Future research will evaluate the system in collaboration with clinical partners, evaluating its practicality and impact through pilot studies and external patient cohorts. The method can be further enhances on a large and diverse datasets using deep learning techniques. The suggested methodology highlights the necessity for external validation prior to clinical use while demonstrating the possibility of merging generative modeling and machine learning for personalized cardiovascular risk prevention. As compared to traditional baseline models LR (AUC 0.86) and RF (AUC 0.92), the proposed hybrid GAN based system with ensemble learning and personalized recommendation achieved an AUC of 0.96 with statistically significant improvement.

References

[1] Dorairaj Prabhakaran et al, Cardiovascular Diseases in India Compared with the United States, VOL. 72, NO. 1, 2018
[2] Francesco Santoro1 et al, Risk prediction of major cardiac adverse events and all-cause death following covid-19 hospitalization at one year follow-up: The HOPE-2 score, Volume 124,P108-114June 2024
[3] Bryan G. Pearson et al, Early, rapidly progressive vasculopathy in a transplanted heart: A possible complication of COVID-19, Volume 72, September–October 2024, 107661
[4] Ayse Dogan et al, A utility-based machine learning-driven personalized lifestyle recommendation for cardiovascular disease prevention, Volume 141, May 2023, 104342
[5] Pradeep Kumar singh et al(2021), Recommender systems: an overview, research trends, and future directions, Int. J. Business and Systems Research, Vol. 15, No. 1.
[6] Mayank Dalakoti et al, Incorporating AI into Cardiovascular disease prevention-insights from Singapore
[7] Chih-Lin Chi et al, Individualized patient-centered lifestyle recommendations: An expert system for communicating patient specific cardiovascular risk information and prioritizing lifestyle options, Journal of Biomedical Informatics, Volume 45, Issue 6, December 2012, pages 1164—1174.
[8] Chunxu Wang et al, Park Recommendation Algorithm based on User Reviews and Ratings, vol. 15, no. 3, March 2019, pp. 803-812.
[9] Jonathan Gemmell et al, Resource recommendation in social annotation systems: A linear-weighted hybrid approach, Volume 78, Issue 4, July 2012, pp 1160-1174.
[10] F.T.Greenway et al, Hybrid mHealth care: Patient perspectives of blended treatments for psychosis. A systematic review, Volume 274, 2024, pp1-10.
[11] Adrian Alonso Barriuso et al, Recommendation system of scientific articles from discharge summaries, volume 136, Part B, 2024.
[12] Jon Nicolas Bondevik et al, A systematic review on food recommender systems, Volume 238, part E, 2024,122166.
[13] Abdul Hannan et al, Machinelearning-basedsmartwearablesystemforcardiacarrestmonitoring using hybrid computing, BMSP, Volume 87, part-B, 2024,105519.
[14] Mojtaba Salehi and Isa Nakhai, A Hybrid Attribute-based Recommender System for E-learning Material Recommendation, IERI Procedia 2 (2012) 565 – 570.

[15] Yonghong et al, College Library Personalized Recommendation System Based on Hybrid Recommendation Algorithm, Procedia CIRP 83 (2019) 490–494.
[16] Steven S.Fu et al, Evaluating chronic disease approaches to ameliorate tobacco-related health disparities: Study protocol of a hybrid type 1 implementation-effectiveness trial, Volume 42, 2024,101380.
[17] Samira Khademzadeh et al, an intelligent recommender system for people who are prone to fatty liver disease, volume 41, 2023,101315.
[18] R. Sivasankari and J. Dhilipan, Hybrid scientific article recommendation system with COOT optimization, Data science and Management, 7, 2024, 99-107.
[19] Intifa Aman Taifa et al, A hybrid approach with customized machine learning classifiers and multiple feature extractors for enhancing diabetic retinopathy detection, Healthcare Analytics,5,2024,100346.
[20] Shilpa Balan and Sumali Conlon, A Study of Health Recommender Systems: Analyzing the Advancement of Sustainable Health Recommender Systems Using NLP Techniques, Journal of Strategic Innovation and Sustainability Vol. 17(3) 2022.

Authors' Profiles

Arundhati Uplopwar received her B.E degree from Nagpur University, Maharashtra and M.Tech from Shobhit University, Meerut (UP) in Computer Science and Engg. She is currently working as Associate Professor in Computer Science and Engineering Department, RKG Group of Institution Ghaziabad (UP). She has teaching Experience of more than 25 years. She is pursuing her PhD from Amity University, Noida (UP). Her current areas of research are Data Science and Machine Learning for Healthcare Industry.

Rahsmi Vashisth is currently working as Associate Professor with Amity Institute of Information Technology, Amity University, Noida, India. She has done her PhD (Engg) from Amity University, Noida.She has teaching experience of 16 Years. She is a Life member of Indian Society of Technical Education (ISTE) and Ultrasonic Society of India. She has published more than 40 research papers in reputed Scopus/SCIE Journals, book chapter and International Conferences. She has filed three Patents and written book on Digital Control Systems. She has chaired many international conference sessions. Her research areas include Embedded Systems, Machine Learning, Fuzzy logic, IoT, Artificial Intelligence.

Arvinda Kushwaha received Bachelor's degree in Computer Science and Engg from Bundelkhand University, Jhansi (UP) in 2002 and M.Tech from RGPV, Bhopal (MP) and PhD in Computer Engineering from Jamia Millia Islamia, New Delhi in 2020.Currently he is working as Professor in the department of Data Science, GCET, Greater Noida (UP).He has published more than 30 research papers in reputed International Journals including Thomson Reuters (SCI and Scopus) and IEEE conferences. His area of interest includes WSN, Cloud computing and Machine learning.

How to cite this paper: Arundhati Uplopwar, Rashmi Vashisth, Arvinda Kushwaha, "Personalized Cardiovascular Risk Reduction: A Hybrid Recommendation Approach Using Generative Adversarial Networks and Machine Learning", International Journal of Intelligent Systems and Applications(IJISA), Vol.18, No.1, pp.17-27, 2026. DOI:10.5815/ijisa.2026.01.02