SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Multimodal Data Fusion For Enhanced Predictive Modeling In AI Systems

Gagan Tiwari[1], Mahesh Kurulekar[2], Gayathri B[3], Satish Upadhyay[4], M. N. Nachappa[5], Amaal Samir Abd El Hameed[6], E. Archana[7], Priyadharshini K[8], Rohit Goyal[9]

[1] Professor, Department of Computer Sciences, Noida International University, Greater Noida, Uttar Pradesh, India. Email: gagan.tiwari@niu.edu.in
[2] Assistant Professor, Department of Civil Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra 411037, India. Email: mahesh.kurulekar@vit.edu
[3] Assistant Professor, Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, India. Email: gayathrib@maher.ac.in
[4] Assistant Professor, Department of uGDX School of Technology, Atlas SkillTech University, Mumbai, Maharashtra, India. Email: satish.upadhyay@atlasuniversity.edu.in, ORCID: 0000-0002-2865-014X
[5] Professor, Department of CS & IT, JAIN (Deemed-to-be University), Bengaluru, Karnataka, India. Email: mn.nachappa@jainuniversity.ac.in, ORCID: 0000-0002-4007-5504
[6] May University in Cairo, Egypt. Email: amaal@eng.cu.edu.eg; amaal.Samir@muc.edu.eg, ORCID: 0000-0002-8647-7349
[7] Assistant Professor, Department of Computer Science and Engineering, Panimalar Engineering College, Tamil Nadu, India. Email: archana.athi@gmail.com, ORCID: 0000-0003-3601-9259
[8] Assistant Professor, Department of Management Studies, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, India. Email: kpriyamba@maher.ac.in
[9] Associate Professor, Department of Computer Application, Dev Bhoomi Uttarakhand University, Dehradun, Uttarakhand, India. Email: hod.ca@dbuu.ac.in, ORCID: 0000-0001-7978-4079

Abstract
The growing demand for accurate and early prediction of chronic diseases requires advanced intelligent healthcare systems capable of handling heterogeneous medical data. This research proposes an artificial intelligence-based multimodal framework for chronic disease risk prediction through effective data fusion. The model integrates electrocardiogram (ECG) signals and clinical data, including hemodynamic parameters, laboratory biomarkers, and phenotypic attributes. ECG signals are preprocessed using bilateral filtering to remove noise while preserving structural details and clinical data are normalized using Min–Max normalization. Feature extraction is performed using Local Binary Pattern (LBP) for ECG signals and Principal Component Analysis (PCA) for clinical data. The extracted features are combined using a weighted and attention-based fusion strategy to effectively align multimodal representations. The fused features are processed using a Multimodal Convolutional Neural Network (MM-CNN), which learns hierarchical representations from integrated ECG and clinical data for effective cardiovascular risk prediction. The network captures complex intra- and inter-modality patterns through convolutional feature learning and attention-based fusion. To further enhance performance, a Quokka swarm optimization (QSO) strategy is applied to optimize model hyperparameters, improving convergence, robustness, and generalization. The entire model is implemented using Python. The proposed framework is evaluated on a multimodal cardiovascular dataset and demonstrates superior performance compared to conventional approaches, achieving an accuracy of 94.54%, precision of 92.67%, recall of 93.26%, and an F1-score of 94.64%, with an area under the curve exceeding 0.92. These results highlight the effectiveness of the proposed model in supporting early diagnosis and proactive clinical decision-making.

Keywords: Multimodal Data Fusion, Predictive Modeling, Artificial Intelligence, Chronic Diseases, Risk Prediction, Electrocardiogram (ECG).

1. Introduction

Clinical guidance, lifestyle adjustments, and the speedy aging of the international populace have led to chronic illnesses becoming a major health issue in the domain of geriatric healthcare. The chronic disease represents a heavy economic toll on society as well as on patients [1]. Chronic diseases to proven as one of the most common causes of death and represent about 1.35%. Regarding the age group trends, chronic kidney disease (CKD) prevalence is on the rise and currently impacts about 10% of the world's inhabitants [2]. Heart disease continues to be the main cause of death throughout the world. It is highly linked to many elements like obesity, smoking, improper nutrition, lack of exercise, drinking too much alcohol, and other lifestyle issues [3]. On the other hand, the efficiency of clinical decision support systems is assessed, pointing out some issues and improving their functions for providing efficient assistance to healthcare professionals [4]. The transcatheter aortic valve replacement (TAVR) technique is an important development in the field of cardiothoracic surgery, especially for treating aortic valve disease, due to the development of advanced medical technology and imaging techniques [5, 6]. Nevertheless, some drawbacks of 3D printing based on imaging data of patients involve the incapability of precisely modeling the physical dynamics involved in tissue deformation and blood flow, as well as the mechanical behavior of heart tissues. Moreover, a number of disorders may be diagnosed using oral characteristics by applying fusion deep learning algorithms combining panoramic radiography and clinical oral characteristics [7, 8].

Research Aim: This research aims to design an innovative AI model based on Electrocardiogram (ECG) images and clinical records, incorporating the QSO-MM-CNN model with weighted fusion to improve accuracy, robustness, and interpretability for better clinical decision support in predicting the risk of chronic diseases.
Research Organization: Section 1 provides the background of the research. Section 2 focuses on reviewing existing research. Section 3 discusses the proposed model QSO-MM-CNN. Section 4 includes results obtained from experimental analysis. Section 5 presents the conclusion along with performance, limitations, and future work.

2. Related Work

Multi-modal Chronic Obstructive Pulmonary Disease (COPD) prediction early in its development was introduced in [9]. This Multi-Modal Deep Fusion Network (MMDF-Net) algorithm integrates Computed Tomography (CT) imaging, clinical characteristics, and synthetic environments by means of contrastive learning and dynamic gate fusion. The AUC obtained is 92%, which makes it superior to any of the single-modal algorithms used by approximately 15%. Limitations include dependency on dataset quality and model complexity. Both temporal events and clinical notes in Intensive Care Unit (ICU) mortality prediction are proposed in [10]. The technique used is a combination of deep learning and multimodal data fusion by means of the Time Series Embedding Model. The results show an average increase in performance by about 8-12% regarding accuracy. The limitations associated with the model include data inconsistencies, missing data, and computational complexity. Personalized medicine by modeling the interactions between patients and drugs is considered [11]. This includes the application of a Multi-Modal Data-Driven Chain of Decisions (MDD-CoD). It achieved improvements in AUROC (3.1%), andchallenges include the limited diversity of datasets.Multimodal AI for diagnosis and prognosis using imaging, Electronic Health Records (EHRs), and multi-omics data was investigated [12]. The model proposed employs deep fusion techniques with Attention Mechanisms (AM) and uncertainty quantification.A multimodal framework considering clinical records, laboratory tests, and time-dependent features in EHR to predict chronic diseases was proposed [13]. Themethod involves utilizing Large Language Models (LLMs) and the transformer encoder to account for interaction and sequence relationships. The experiments yielded an accuracy of 94%, andlimitations include the bias of the datasets used.Early prediction of chronic diseases through multimodal methods was examined [14], with the Attention Dual Convolutional-based Random Binary Kepler (ADC-RBK)Algorithm pre-processing a Magnetic Resonance Imaging (MRI) images and using an attention-based dual Convolutional Neural Network(CNN) for feature extraction. The results demonstrate high accuracy, although this research contains limitations in generalization. Gastric cancer that is positive for Human Epidermal Growth Factor Receptor 2 (HER2) response prediction improvement was examined [15]. Scientists suggested developing a deep-learning-based multimodal method combining imaging, pathology, and clinic parameters for 429 patients using the Multimodal Model (MuMo). The accuracy of 82.1% obtained, however, is constrained by the small amount of data and the initial clinical validation stage.Multi-morbidity management in elderly patients was investigated [16]. Incorporate the suggested AI-based analysis of genomics, clinical, and behavioral data to provide personalized treatment plans using CNN. Improved diagnostic precision, lowered risks of polypharmacy, and enhanced patient outcomes with

high accuracy. Challenges are data bias, confidentiality issues, high costs, and a lack of interdisciplinary collaboration.An Explainable Artificial Intelligence (XAI) based ensemble technique framework to predict heart disease employing Classification Techniques was explored [17]. Using classifiers like Logistic Regression(LR) and Naive Bayes(NB) on a 303-instance dataset about cardiovascular diseases. Accuracy achieved is 89%, but some of the limitations include a limited data set, scalability issues, and reliance on traditional feature learning techniques.An effective data-driven machine learning model for CVD Risk Prediction was examined [18]. A machine learning model is proposed that uses different classifiers, like LR, to solve problems of class imbalance in predicting cardiovascular diseases. The findings indicate that the stacking ensemble performed with an accuracy rate of 87.8%. The use of various machine learning methods for the early detection of myocardial infarction was developed in [19]. Using Gradient Boost, eXtreme Gradient Boosting (XGBoost), the best performance is given by the optimized XGBoost method, which gives a precision of 99.14% and an accuracy rating of 98.50%, respectively. A predictive model for cardiac disease based on an group learning technique was proposed [20]. This method employs Random Forest (RF), Support Vector Machine (SVM), and Categorical Boost (CatBoost), along with pre-processing, feature selection, and Multilayer Perceptron (MLP). The output obtained is 86.64%, but one limitation is that this model can be applied to structured data only.

3. Methodology

Figure 1 presents a structured workflow for multimodal chronic disease risk prediction. Data collection includes acquiring both ECG signals and clinical CSV data. The ECG signal is filtered by using bilateral filtering for de-noising purposes, and clinical data are normalized via Min-Max normalization. Feature extraction is then performed using LBP for ECG data and PCA for clinical data. The fused features are input into the QSO-MM-CNN model, where Quokka Swarm Optimization enhances performance, producing accurate and reliable chronic disease risk predictions.

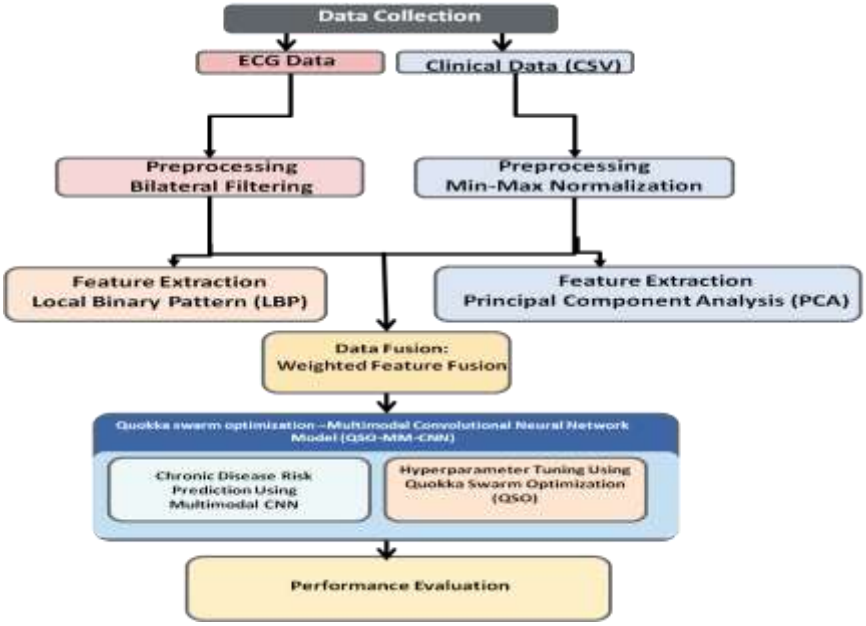

Figure 1: MethodologyWorkflow Multimodal Data Fusion in Disease Risk Prediction

3.1. Data Collection

This research uses a cardiovascular multimodal dataset with 4,327 patient samples, including images of ECG and clinical data in CSV format, which is available at kaggle (https://www.kaggle.com/datasets/zara2099/multimodal-cardiovascular-risk-dataset/data). The dataset contains 22 features on demographic, hemodynamic, laboratory, and ECG characteristics. ECG features are classified as Normal, Abnormal Heartbeat, Myocardial Infarction, and Post-MI History, whereas the CSV contains numerical values of clinical information. Eighty percent of the dataset is used for training, while the remaining twenty percent is used for testing to assess the proposed QSO-MM-CNN model for predicting chronic diseases.

3.2. Data Preprocessing

ECG Images with Bilateral Filter: The procedure entails pre-processing the ECG pictures using a bilateral filter after data gathering in order to help minimize noise without sacrificing any edge details that are crucial for later feature recognition of disease risk. This is because standard linear filters tend to smear the edges of ECG Images signals expressed in Equation (1).

$$k_h(w) = e * H_{\sigma_t}(w) \quad (1)$$

$k_h(w)$denotes the smoothed intensity of the ECG signal at position w, useful for noise reduction without losing the crucial structural information of the cardiac waveform, eis the original ECG image holding raw cardiac electrical signal data, and $*$ represents a convolution operation on the ECG image for applying Gaussian smoothing. $H_{\sigma_t}(w)$ represents the Gaussian kernel function in the spatial domain for denoising and maintaining the ECG waveform. This is achieved by the bilateral filter by adding another intensity-weighted function to compare pixels with the use of Equation (2).

$$x_s(w, z) = H_{\sigma_q}(|\, e(z) - e(w)\,|) \quad (2)$$

$x_s(w, z)$represents the intensity-based weight function that controls the effect of neighboring ECG pixel z on the smoothing of the target pixel w, maintaining the cardiac signal structure, $H_{\sigma_q}()$ represents the Gaussian range kernel that measures the similarity of pixel intensity values for edge-preserving smoothing of ECG images, and $\sigma_q$denotes the range parameter responsible for sensitivity to intensity variation, hence maintaining clinically important features of the ECG during data collection preprocessing. e(z) represents the intensity of neighbor ECG image pixel zlocal cardiac signal information, e(w) represents the intensity of target ECG image pixel w containing current cardiac information for smoothing, and $|\, e(z) - e(w)\,|$ represents the difference in intensities between pixels z and w, ECG waveform variations during preprocessing. The final bilateral filter formula is expressed in Equation (3).

$$J'(w) = \frac{1}{X(w)} \sum_{z \in \Omega} e(z)\, H_{\sigma_t}(|\, w - z\,|) H_{\sigma_q}(|\, e(z) - e(w)\,|) \quad (3)$$

$J'(w)$represents the improved intensity of the ECG pixels at point wis the following noise removal, for maintaining clinically significant patterns of heart waveforms for predicting risk. Ω Nearby ECG pixels at w to collect local ECG signal information, $H_{\sigma_t}(|\, w - z\,|)$ represents geometric distance between pixels to preserve structural ECG features, $H_{\sigma_q}(|\, e(z) - e(w)\,|)$ Gaussian range filter to preserve the ECG Edges, and X(w) to ensure balanced weight of all neighboring pixels in ECG Preprocessing.

Clinical Data Using Min-Max Normalization: ECG images are denoised using bilateral filtering, which effectively removes noise while preserving important edge and structural details of the signal. Concurrently, the clinical data are normalized using Min–Max normalization, a method that rescales feature values to a common range from 0 to 1. This guarantees homogeneity in disparate clinical features, ensuring that the model is trained successfully. This can be illustrated through Equation (4).

$$W_{new} = \frac{W - W_{min}}{W_{max} - W_{min}} \quad (4)$$

Value of clinical feature W_{new}, which is used for normalization to ensure consistency in predicting cardiovascular risk. W refers to the first value determined by clinical measurement in patient data in the dataset, while W_{min} and W_{max} refer to the minimum and maximum values of the particular clinical feature respectively.

3.3 Feature Extraction

ECG Signal Feature Extraction Using Local Binary Patterns (LBP): After the initial image processing step, the next phase involves extracting relevant features from ECG images using local binary patterns. This technique analyzes the pattern of intensities within a particular region, providing information on structural characteristics. The mathematical representation is provided in Equation (5).

$$LBP(w_c) = \sum_{o=0}^{o-1} T(O_o - O_d) 2^o \quad (5)$$

Value of feature $LBP(w_c)$ for LBPto encode local texture information of ECG images for feature extraction, w_c represents the binary pattern computation of ECG image signals, and O_o indicates the local cardiac signals intensity level of the o^{th}position. o_d denotes the intensity of the central pixel against which the neighbor pixel intensity of the ECG image is compared, $\sum_{o=0}^{o-1} T(O_o - O_d) 2^o$ represents the step function whereby 1 results when the neighbor pixel intensity is larger than the central pixel of the ECG image intensity, then otherwise 0, and 2^o represents the encoding of local texture patterns in the ECG image. o denotes the index for the local neighborhoodof the ECG Image.

Dimensionality Reduction Using Principal Component Analysis (PCA):After extracting ECG features using LBP, PCA is applied to clinical CSV data for dimensionality reduction. It preserves essential variance, removes noise,

and improves computational efficiency. The feature reduction allows efficient fusion with the ECG features, which improves the performance and accuracy of heart disease prediction models, as expressed by Equation (6).

$$Y = WX_l \quad (6)$$

Y refers to the PCA-transformed clinical feature vector obtained after reducing the dimensions, W represents the clinical feature vectors derived from the cardiac ECG signal dataset, and X_l denotes the labeled patient data.

3.4Weighted Feature Fusion for Multimodal Cardiovascular Risk Prediction

Following the process of feature extraction, the extracted ECG features via LBP and reduced clinical features via PCA are combined by employing a weighted feature fusion technique. This method considers the proper significance of each modality and enables the effective fusion of different data modalities. Each of the feature sets is weighted with an appropriate weight factor according to their respective prediction capacity and combined to produce a feature vector. The equation for this procedure is formulated in Equation (7).

$$F_{fusion} = \alpha F_{ECG} + \beta F_{Clinical} \quad (7)$$

where F_{fusion} represents the resultant feature vector after fusion, F_{ECG} is the ECG feature vector extracted using LBP, and $F_{Clinical}$ is the feature vector obtained after performing PCA on the clinical features. The weighting factors for both modalities are denoted by α and β, respectively, with $\alpha + \beta = 1$. This approach increases the discrimination capability of the feature vectors, helps in more efficient learning in QSO-MM-CNN, and thus improves disease risk prediction performance.

3.5 Cardiovascular disease risk prediction using Quokka swarm optimized Multi-modal Convolutional Neural Network (QSO-MM-CNN)

QSO-MM-CNN framework is designed based on multimodal characteristics that will be used to predict the probability of having cardiovascular diseases by adopting an attention mechanism-based weighted feature fusion framework normalized by softmax function for balanced contribution from all modes of information. These multimodal features undergo convolution operations to extract more complex interactions among features. In addition, QSO is adopted to tune hyperparameters for achieving better convergence, lower error rate, and generalizability.

Multimodal Convolutional Neural Network (MM-CNN): MM-CNN is a multimodal deep learning model that utilizes different features of various modalities to make risk predictions about heart diseases. In addition, MM-CNN uses an attention-based weighted fusion strategy where softmax normalized weights play an important role in making sure all modalities contribute equally to the process. Further, the multimodal features undergo processing by convolutional layers to extract useful patterns. Moreover, concatenation and mean fusion strategies have also been added to increase the accuracy of risk prediction. In the case of summation, the weighting vector should first be normalized by using the softmax function. The weighted fusion mechanism is mathematically expressed in Equations (8 and 9).

$$\alpha_j = \frac{f^{x^j}}{\sum_{i-1}^{m} f^{x^j}} \quad (8)$$

The relevant weight factor for the j^{th} input modality in cardiovascular disease prediction is weight α_j. Weight f^{x^j} The learned relevance weight of the j^{th} feature in the prediction task. x^j is the learned weight factor for the j^{th} input modality. $\sum_{i-1}^{m} f^{x^j}$. Term used for the normalization of the weights of the modalities within m modalities. m represents the number of input multimodal sources.

$$E_{fusion} = \sum_{i=1}^{m} \alpha_j E_j \quad (9)$$

E_{fusion} Final fused feature representation combining all modalities for cardiovascular disease risk prediction, α_j Attention weight representing the importance of the j^{th} modality in fusion, E_j Feature vector extracted from the j^{th} data source (e.g., clinical, ECG), and $\Sigma_{i=1}^{m}$ Summation over all available modalities used in multimodal fusion. m is the total number of modalities used in the risk prediction method for cardiovascular disease, or Concatenate and Mean strategies, fusion is defined as expressed in Equation (10).

$$E_{concat} = [\alpha_1 E_1, \alpha_2 E_2, \dots, \alpha_m E_m] \quad (10)$$

E_{concat} Fusion vector created by concatenating all weighted modalities for predicting cardiovascular disease risk,$\alpha_1, \alpha_2, \dots, \alpha_m$ represents weights indicating the significance of each modality during fusion, and $E_1, E_2, \dots, E_m$ Feature vectors derived from various data sources, like clinical data, imaging data. [.] Concatenation function employed to integrate weighted feature vectors into one. mis the Number of multimodal inputs utilized for predicting cardiovascular disease risk.

Hyperparameter tuning with Quokka swarm optimization (QSO):Following MM-CNN-based feature learning, QSO is applied to tune key hyperparameters such as learning rate and network weights. This optimization strategy improves convergence stability, minimizes prediction errors, and enhances generalization across patient data. By efficiently exploring the search space, QSO ensures robust and reliable performance of the MM-CNN model. The position update mechanism of QSO is mathematically defined as expressed in Equations (11 and 12).

$$C^{new} = \frac{(S+G)}{(0.8 \times C^{old})} + \Delta x \times rand \times \Delta w \quad (11)$$

$C^{new}$is the Predicted value after fusing and optimizing multiple modalities of information for cardiovascular risk prediction, $C^{old}$is the original value of prediction before optimization for CVD, and $S$is atemporal characteristic value obtained from the time-dependent health-related dataset. G Input health data heterogeneity (that is, input coming from different modalities such as clinical and imaging data), 0.8 Normalization constant to avoid bias in estimating risk, and Δx Change in weight due to the learning process for a certain feature. $rand$ Random value and Δw Variance in feature space parameter.

$$W^{new} = W^{old} + C^{new} \times M \quad (12)$$

$W^{new}$is the fusion of feature representations for prediction of cardiovascular risks. W^{old} Fusion of old feature vectors to predict cardiovascular risks, and $C^{new}$is the cardiovascular risk determinants from new feature adjustments. M represents the normalization constant used to control the risk of feature adjustments.

4. Result

The proposed QSO-MM-CNN framework is quite efficient in predicting the risk factors for chronic diseases through multimodal analysis. Metrics used in evaluating its performance have proved to be highly accurate and reliable, giving balanced predictions. As compared to other models, it ensures accurate predictions, which makes decision-making easier for clinicians. The experiment was conducted using the following machine specifications, the Intel Core i7/AMD Ryzen 7 CPU, NVIDIA RTX 3060 GPU, 16-32 GB RAM, and 512 GB SSD. Implementation is performed on Windows 10/11 via Python 3.8+ with the following libraries: PyTorch/TensorFlow, NumPy, Pandas, Scikit-learn, and Matplotlib. The work is done in Jupyter Notebook or VS Code.

4.1. Exploratory Analysis

In Figure 2(a), the class distribution of the target variable is depicted, which includes Normal (25.7%), Abnormal Heartbeat (25.2%), Post Myocardial Infarction History (25.2%), and Myocardial Infarction (23.8%). Such an evenly distributed dataset facilitates stable convergence of the model and avoids class bias. In Figure 2(b), the Pearson correlation between features like age, heart rate, cholesterol, and glucose provided, suggesting the importance of each feature for information fusion. Figure 2(c) presents the decomposing of the heart rate variability series into observed values, trends, seasonality, and residuals, thereby facilitating a better comprehension of temporal patterns and synchronization. Figure 2(d) depicts the distribution of heart rates in different categories of patients, revealing significant discrepancies that contribute to improved feature distinguishability within the QSO-MM-CNN model.

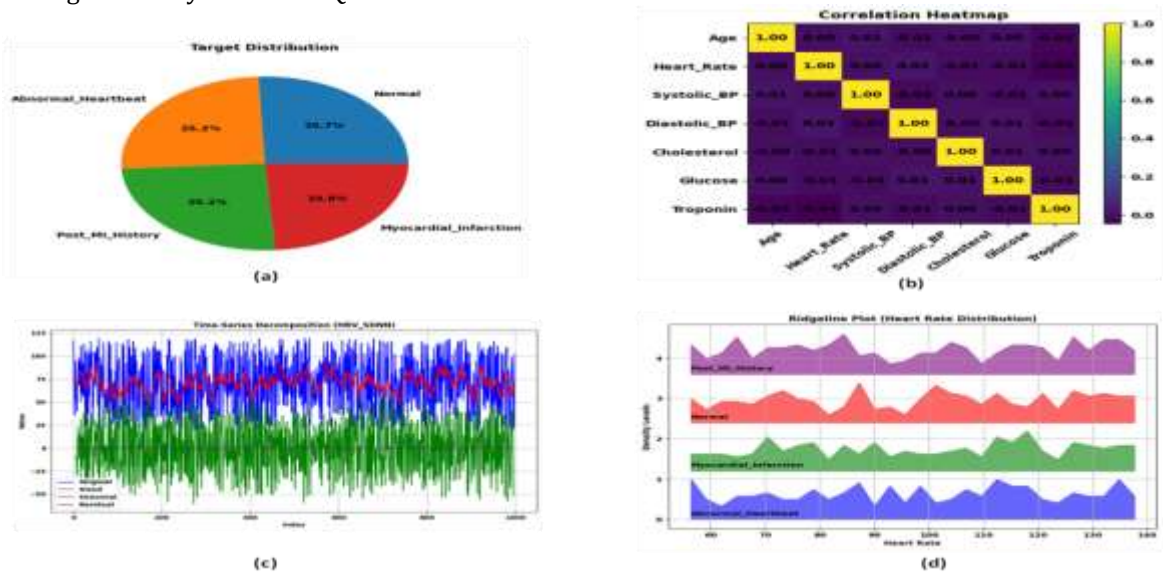

Figure 2:Multimodal Clinical Data Analysis (a) Class distribution of cardiovascular conditions, (b) Correlation heatmap of clinical features, (c) HRV time-series decomposition, and (d) Heart rate distribution across patient groups

4.2. Evaluation Metrics

Classifiers are used in order to analyze how effective the proposed QSO-MM-CNN is in predicting the risks of chronic diseases using multi-modality. First, accuracy is used as a measure of correctness of predictions based on the use of combined ECG and clinical data. Next, precision determines the model's ability to make correct predictions and reduce false positive outcomes so that reliable clinical conclusions can be made. Third, recall reflects how effective the model is in identifying existing diseases for early treatment purposes.

4.3. Performance Evaluation

The following Table 1 shows the comparative research of the proposed QSO-MM-CNN model with various machine learning models, such as CatBoost, SVM, Random Forest, LightGBM, XGBoost, Logistic Regression, and KNN [20] in the prediction of chronic diseases. For instance, CatBoost obtains 86.64% accuracy and an F1 score of 88.18%, whereas SVM gets 85.96% accuracy and an F1 score of 87.67%, suggesting acceptable prediction ability. The proposed QSO-MM-CNN, however, performs better than all benchmark models with much higher accuracy, which is 94.54%, precision, 92.67%, recall, 93.26%, and F1 score of 94.64%, respectively. This reveals the superiority of the proposed technique owing to its better feature extraction and optimization [20].

Table 1: Comparison analyses of Existing and Proposed Method

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CatBoost [20]	86.64	86.75	89.90	88.18
SVM [20]	85.96	85.43	90.14	87.67
Random Forest [20]	85.96	85.69	89.90	87.64
LightGBM [20]	85.41	86.27	87.91	87.00
XGBoost [20]	85.28	85.73	88.43	86.93
Logistic Regression [20]	85.27	86.29	87.42	86.75
KNN [20]	83.92	83.89	88.18	85.86
QSO-MM-CNN [Proposed]	94.54	92.67	93.26	94.64

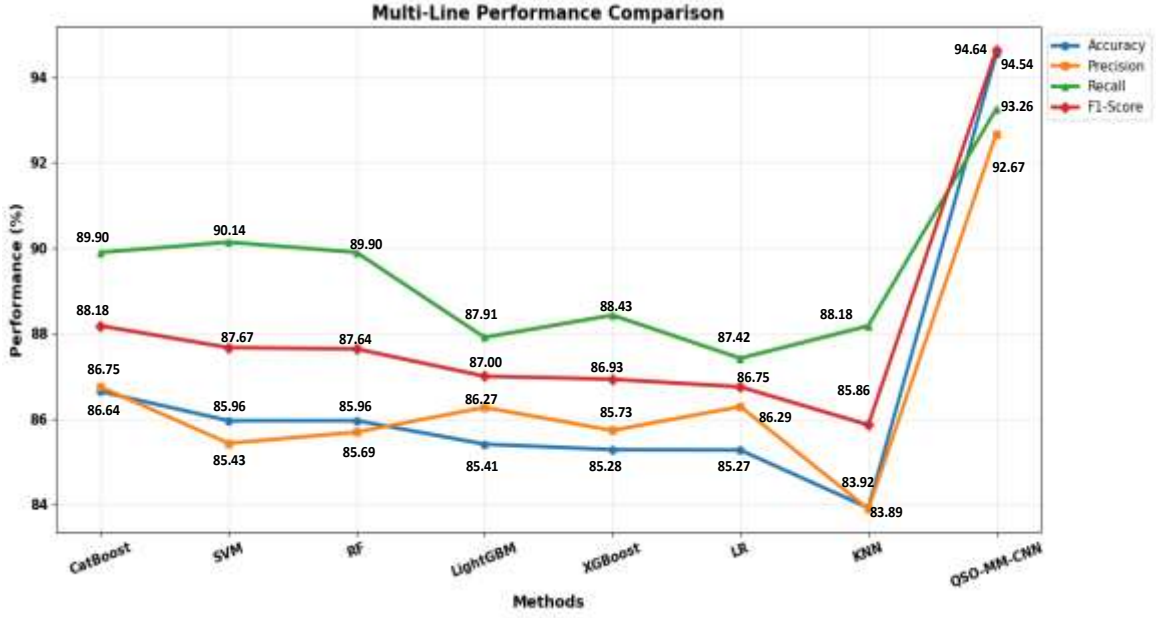

Figure 3: Performance Comparison of Machine Learning Models and Proposed QSO-MM-CNN for Chronic Disease Risk Prediction

4.4. Discussion

The design and implementation of efficient multimodal artificial intelligence architecture is carried out for chronic disease risk prediction by incorporating diverse medical data, including ECG images, hemodynamics, lab results, and clinical phenotypes. Common machine learning approaches, such as CatBoost, SVM, Random Forest, LightGBM, XGBoost, Logistic Regression, and KNN, have been extensively applied in predicting heart diseases [20]. In order to overcome such issues, the proposed QSO-MM-CNN algorithm incorporates multimodal deep learning to extract hierarchical features from ECG and clinical information. Additionally, QSO contributes in tuning network parameters during the training process. The proposed method enhances the predictive accuracy.

5. Conclusion

The proposed research framework uses an AI-based approach to predict CVD risk factors based on multimodal learning and various heterogeneous data types, including ECG signals, hemodynamics, lab results, and clinical phenotype data. Denoising takes place during pre-processing of the ECG signals, while normalization occurs when processing the clinical data. Feature extraction using LBP to extract features from ECG signals and PCA for feature extraction from the clinical data set. Using the QSO-MM-CNN framework, it proved to be efficient in dealing with the modalities involved and achieved significant results, including 94.54% accuracy, 92.67% precision, 93.26% recall, and 94.64% F1-score. Although these were successful findings, the results depended on the quality of the data used and were computationally intensive. The future work will seek to achieve improved generalization through a large and varied dataset.

REFERENCES

1. He, W., Luo, J. and Yang, X., 2025. The Application of Multimodal Data Fusion Algorithm MULTINet in Postoperative Risk Assessment of TAVR. Journal of Clinical Medicine, 14(24), p.8620. https://doi.org/10.3390/jcm14248620
2. Tsai, H., Yang, T.W., Wu, T.Y., Tu, Y.C., Chen, C.L. and Chou, C.F., 2025. Multitask learning multimodal network for chronic disease prediction. Scientific Reports, 15(1), p.15468. https://doi.org/10.1038/s41598-025-99554-z
3. Amal, S., Safarnejad, L., Omiye, J.A., Ghanzouri, I., Cabot, J.H. and Ross, E.G., 2022. Use of multi-modal data and machine learning to improve cardiovascular disease care. Frontiers in cardiovascular medicine, 9, p.840262. https://doi.org/10.3389/fcvm.2022.840262
4. Javed, R., Abbas, T., Sayyed, A., Abbas, S., Shah, A.A. and Adnan, K.M., 2026. Internet of Medical Things Enabled Multimodal Framework: Deep Machine Learning for Chronic Cardiac disease Prediction in Healthcare 5.0. Healthcare Technology Letters, 13(1), p.e70063. https://doi.org/10.1049/htl2.70063
5. Miao, S., Ji, P., Zhu, Y., Meng, H., Jing, M., Sheng, R., Zhang, X., Ding, H., Guo, J., Gao, W. and Yang, G., 2025. The construction and application of a clinical decision support system for cardiovascular diseases: multimodal Data-Driven development and validation study. JMIR medical informatics, 13, p.e63186. https://doi.org/10.2196/63186
6. Lee, Y.C., Cha, J., Shim, I., Park, W.Y., Kang, S.W., Lim, D.H. and Won, H.H., 2023. Multimodal deep learning of fundus abnormalities and traditional risk factors for cardiovascular risk prediction. NPJ digital medicine, 6(1), p.14. https://doi.org/10.1038/s41746-023-00748-4
7. Yang, X.Y., Li, Y.M., Wang, J.Y., Jia, Y.H., Yi, Z. and Chen, M., 2025. Utilizing multimodal artificial intelligence to advance cardiovascular diseases. Precision clinical medicine, 8(3), p.pbaf016. https://doi.org/10.1093/pcmedi/pbaf016
8. Zhao, D., Homayounfar, M., Zhen, Z., Wu, M.Z., Yu, S.Y., Yiu, K.H., Vardhanabhuti, V., Pelekos, G., Jin, L. and Koohi-Moghadam, M., 2022. A multimodal deep learning approach to predicting systemic cardiovascular diseases from oral conditions. Diagnostics, 12(12), p.3192. https://doi.org/10.3390/diagnostics12123192
9. Deng, H., Ding, X., Zhu, S., Song, X. and Guo, Y., 2025. Prediction of chronic obstructive pulmonary cardiovascular disease based on multimodal data and deep learning. APL bioengineering, 9(4). https://doi.org/10.1063/5.0289414
10. Yang, H., Kuang, L. and Xia, F., 2021. Multimodal temporal-clinical note network for mortality prediction. Journal of Biomedical Semantics, 12(1), p.3. https://doi.org/10.1186/s13326-021-00235-3

11. Chu, X., Ye, Y., Tang, S., Han, M., Wang, G., Lin, S., Sun, B., Huang, Q., Zhang, Y., Chu, X. and Bao, K., 2025. Personalized Medication for Chronic diseases Using Multimodal Data-Driven Chain-of-Decisions. Advanced Science, 12(40), p.e04079. https://doi.org/10.1002/advs.202504079
12. Zhang, R., Chen, Y., Yue, W., Zhang, Y., Li, X., Feng, S., Yuan, F. and Luo, M., 2025. Multimodal artificial intelligence in medicine: a task-oriented framework for clinical translation. Frontiers in Medicine, 12, p.1736272. https://doi.org/10.3389/fmed.2025.1736272
13. Dao, C.T., Phan, N.M.T., Ding, J.E., Wu, C., Restrepo, D., Luo, D., Zhao, F., Liao, C.C., Peng, W.C., Wang, C.T. and Chen, P.F., 2025. CURENet: combining unified representations for efficient chronic disease prediction. Health Information Science and Systems, 14(1), p.7. https://doi.org/10.1007/s13755-025-00396-w
14. Shiny Irene, D., Lakshmi, M., Kumar, L.R. and Kishore, T.R., 2024. ADC-RBK: a multimodal approach for early detection of chronic cardiovascular diseases and focusing on Alzheimer's. Progress in Artificial Intelligence, 13(3), pp.187-200. https://doi.org/10.1007/s13748-024-00327-y
15. Chen, Z., Chen, Y., Sun, Y., Tang, L., Zhang, L., Hu, Y., He, M., Li, Z., Cheng, S., Yuan, J. and Wang, Z., 2024. Predicting gastric cancer response to anti-HER2 therapy or anti-HER2 combined immunotherapy based on multi-modal data. Signal transduction and targeted therapy, 9(1), p.222. https://doi.org/10.1038/s41392-024-01932-y
16. Deng, W., Zhang, L.Y., Yue, J.R. and Huang, X.L., 2025. The future of multimorbidity management in the older adults: transforming AI-enabled precision medicine. Frontiers in Public Health, 13, p.1691682. https://doi.org/10.3389/fpubh.2025.1691682
17. Guleria, P., Naga Srinivasu, P., Ahmed, S., Almusallam, N. and Alarfaj, F.K., 2022. XAI framework for cardiovascular disease prediction using classification techniques. Electronics, 11(24), p.4086. https://doi.org/10.3390/electronics11244086
18. Dritsas, E. and Trigka, M., 2023. Efficient data-driven machine learning models for cardiovascular diseases risk prediction. Sensors, 23(3), p.1161. https://doi.org/10.3390/s23031161
19. Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A.M. and Qasem, S.N., 2024. Machine learning-based predictive models for detection of cardiovascular diseases. Diagnostics, 14(2), p.144. https://doi.org/10.3390/diagnostics14020144
20. Bihri, H., Charaf, L.A., Azzouzi, S. and Charaf, M.E.H., 2025. A Robust Stacking-Based Ensemble Model for Predicting Cardiovascular Diseases. AI, 6(7), p.160. https://doi.org/10.3390/ai6070160