RESEARCH ARTICLE

Check for updates

Fused-CNN-LSTM: A Software-Oriented Multimodal Deep Learning Framework for Intelligent Hypertension Risk Prediction

Wajiha Farooq[1], **Aamir Ali**[1,*], **Hafza Mehreen Fatima**[2], **Wasif Rafiq**[1], **Noor E Zainab**[1] and **Misbah Ali**[1,*]

[1] Department of Computer Science, COMSATS University Islamabad, Sahiwal 57000, Pakistan
[2] Department of Software Engineering, Riphah International University, Islamabad 44000, Pakistan

Abstract

Hypertension, a life-threatening global health challenge, requires early detection to prevent severe cardiovascular complications. Fundus imaging reveals microvascular alterations, yet conventional diagnosis often misses subtle early changes. This study introduces a multimodal deep learning framework that integrates clinical data, fundus images, and demographic features to improve hypertension prediction. Unlike single-modality approaches, our method captures complementary risk factors from both structured and unstructured data. We evaluate machine learning and deep learning models on clinical data, confirming DL's superior accuracy. For fundus images alone, a CNN achieves 74.44% accuracy, highlighting the limitations of unimodal image analysis. To overcome this, we propose a fused CNN-LSTM architecture that models both spatial biomarkers and temporal clinical trends. The framework achieves robust performance, with an overall accuracy of 98% and minimal variation across datasets. Implemented in TensorFlow/Keras, the system adopts a modular, software-oriented design, ensuring flexibility, ease of maintenance, and seamless integration into clinical workflows. This holistic approach enables timely intervention, improves patient outcomes, and reduces healthcare burdens.

Keywords: convolutional neural network, long short-term memory, hypertension, machine learning, deep learning.

1 Introduction

Hypertension, or high blood pressure, is one of the most common and serious chronic conditions, affecting over 1.3 billion people worldwide. In the United States, hypertension contributes to more than 400,000 deaths annually, while globally it is estimated to account for 7.5 million deaths each year, primarily through its role as a leading risk factor for cardiovascular disease [1, 2]. Despite its high

Submitted: 04 August 2025
Accepted: 17 November 2025
Published: 30 January 2026

Vol. 2, **No.** 1, 2026.
10.62762/JSE.2025.995217

***Corresponding authors:**
Aamir Ali
amirali4436823@gmail.com
Misbah Ali
talktomisbah.ali@gmail.com

Citation
Farooq, W., Ali, A., Fatima, H. M., Rafiq, W., Zainab, N. E., & Ali, M. (2026). Fused-CNN-LSTM: A Software-Oriented Multimodal Deep Learning Framework for Intelligent Hypertension Risk Prediction. *ICCK Journal of Software Engineering*, 2(1), 11–29.

prevalence, many patients remain undiagnosed or untreated due to its largely asymptomatic progression. This underscores the need for early diagnosis and intervention to minimize hypertension morbidity and mortality [3].

Hypertension is often termed a “silent disease” because most individuals remain asymptomatic until advanced stages, underscoring the importance of routine measurement and early detection. While some might experience symptoms like headaches, chest pain, or dizziness. Identifying and monitoring blood pressure remains a key research challenge especially in low-income countries with limited healthcare infrastructure [4, 5].

Traditional diagnostic methods relying on only pressure measurements are prone to variability and might omit early disease indicators [6, 7]. This underscores the need for automated and integrated screening tools, capable of identifying early-stage hypertension before complications arise, especially in low-income settings where proactive healthcare is limited. The classification of blood pressure Stages based on systolic and diastolic readings for hypertension diagnosis is shown in Table 1.

Table 1. Stages of hypertension [8].

Stages	Systolic	Diastolic
Normal	Below 120	Below 80
Elevated	120-129	Below 80
Hypertension stage 1	130-139	80-89
Hypertension stage 2	140-180	90 higher
Hypertension stage 3	Above 180	Above 120

The line chart illustrating the progression of systolic and diastolic blood pressure levels across different hypertension stages, highlighting a steady increase from normal to stage 3, is shown in Figure 1.

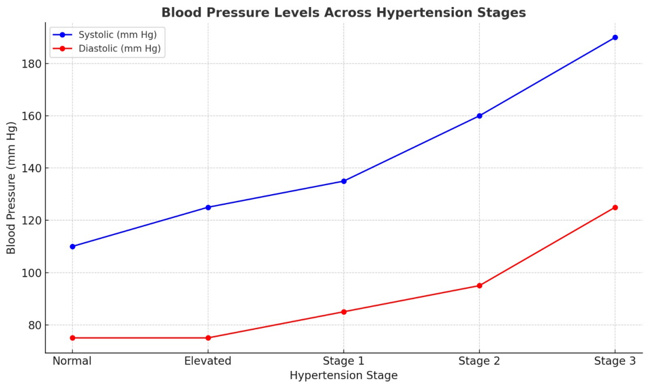


Figure 1. Blood Pressure levels across hypertension stages.

Hypertension is a common but manageable disease after early detection. Early detection and correct treatment can decrease preventable health complications and improve overall wellness. A potential method for evaluating changes in the retinal blood vessels is through fundus photography which supplies an excellent window for viewing hypertensive-induced changes in the body’s circulatory system. Retinal images often reveal very slight changes in the tiny blood vessels; changes so subtle that they can be easily overlooked by the human eye, especially in the early stages of disease. Detecting these minor abnormalities requires advanced tools, as they are not always visible through simple observation. These drawbacks in manual detection capabilities have sparked research that has since explored automated methods utilizing ML and DL technologies to enhance early hypertension discovery [9, 10].

There are some machine learning (ML) algorithms that demonstrate possible applications in the detection of hypertension by analyzing clinical and physiological datasets. One study explained that ML algorithms based on PhotoPlethysmographic (PPG) signals and some demographic information suggested the absence of truly integrated multimodal systems [8]. Clinical datasets encompass structured and semi-structured health information, such as the patient’s age, sex, body mass index (BMI), smoking status, alcohol intake, familial history of hypertension, blood glucose level, and other relevant vitals. These parameters are statistically and clinically validated to have a relationship with hypertension and are commonly employed in the screening and risk assessment processes [11].

Combining different information is possible with ML and Deep Learning (DL). Convolutional Neural Networks (CNNs), originally developed for image recognition and later applied to various domains have shown exceptional capacity in extracting spatial patterns [12]. Modern ML and DL systems can analyze and identify the signals associated with the increase of hypertension and diagnose issues very early, sometimes even in advance. Predictive modeling and diagnosis have greatly benefited from the application of traditional ML techniques: Support Vector Machine (SVM), k-Nearest Neighbors, and Decision Trees. These algorithms are based on mathematical models that classify datasets by recognizing and leveraging patterns and relationships existing in the data [8, 13]. Retinal fundus photography is a non-invasive and relatively inexpensive technique to evaluate the microvascular structure of the circulatory system. Certain changes in retinal vessels, for instance, the narrowing of arterioles, Arteriovenous (AV) nicking, and retinal hemorrhages, are recognized as indicators of hypertension and its complications [10]. Manual examination of fundus images tends to have a high degree of variability, and often fails in the detection of hypertensive damage because of its subtle early changes. In recent years, DL methods, and in particular CNNs, have been shown to effectively capture and extract complex spatial and advanced features of retinal images. However, these approaches are usually constrained when applied to single-source datasets; specifically, fundus images or demographic information in isolation. Such limitations can impede the ability to capture the multidimensional complexity of hypertension.

In addressing the issue of challenge limitations, the first work sought to evaluate the performance of ML algorithms (KNN and SVM) and DL algorithms; CNN and Long Short-Term memory (LSTM) on clinical datasets. The second one applies the DL model (CNN) on fundus images, and as it is known, CNNs spatially extract features and patterns indicative of hypertension. Recent advancements in multimodal approaches have fostered more comprehensive methodologies to hypertension detection. These multimodal methods integrate data from several sources, including fundus images, clinical measurements (including blood pressure and cholesterol levels), and demographic information (age and gender). Through these diverse data modalities, the multimodal approach is capable of revealing intricate patterns and relationships that would otherwise remain unnoticed with conventional ML techniques.

This is the main purpose of this research, to evaluate the effectiveness of conventional ML techniques against different approaches involving DL and multimodal in hypertension detection using fundus and clinical datasets. Furthermore, the study seeks to evaluate the effectiveness of these approaches in identifying individuals at high-risk of developing hypertension using AI. The research findings would aid in providing the most efficient and trustworthy methods for the primary and secondary preventive strategies of hypertension; thereby improving healthcare outcomes and financial burdens on healthcare systems. The importance of early and accurate detection of hypertension is paramount in addressing the burden of cardiovascular disease.

Most existing approaches for hypertension prediction are uni-modal, relying either on clinical records or on fundus images. Clinical models capture structured demographic and physiological factors but miss subtle retinal changes, while image-based models detect micro-vascular biomarkers yet suffer from class imbalance, variable image quality, and lack of contextual patient data. To address these limitations, we propose a multimodal deep learning framework, FusedCNN-LSTM, that integrates clinical and imaging features. By combining temporal clinical signals with spatial retinal biomarkers, our approach achieves significant performance gains over uni-modal models and introduces a perspective not yet systematically explored for hypertension prediction.

The remainder of this paper is organized as follows. Section 2 reviews related work on hypertension prediction using clinical data, fundus images, and multimodal approaches. Section 3 describes the datasets, preprocessing steps, and the proposed FusedCNN-LSTM methodology. Section 4 presents the experimental results and discussion, including clinical deployment and limitations of the study. Section 5 concludes the study by summarizing the key findings, highlighting the clinical relevance of the proposed framework, and outlining directions for future research.

2 Literature Work

Conventional clinical assessment depends on prompt blood pressure readings alongside segmented patient information, which potentially neglects critical underlying risk factors. Recently, applications of artificial intelligence (AI), particularly in ML and DL, have been increasingly successful in predicting hypertension from diverse data sources including clinical data, fundus images, and signals from sensors. This section synthesizes related literature in three domains: models based on clinical data, models based on fundus images, and multimodal fusion models.

2.1 Machine Learning (ML) Unimodal Approaches

The predictive capabilities of ML algorithms have been examined in relation to hypertension with clinical and demographic features as input variables. One research examined, k-nearest neighbors (KNN), and decision tree classifiers and emphasized on their interpretability; while noting a predictive accuracy of 81% across the several datasets used [1].

In another research, more advanced models including logistic regression, artificial neural networks, random forest, and Extreme Gradient Boosting (XGBoost), were trained on a dataset with 612 participants. XGBoost was the best performer in terms of accuracy citing age, weight, fat, income, and BMI as important predictors [4]. Another study applied Long Short-Term Memory (LSTM) networks to longitudinal Electronic Health Record (EHR) data from over 233000 patients, demonstrating a notable improvement in prediction performance with sequential modeling [13]. A further study proposed a hybrid approach that combined LSTM networks with tree-based feature engineering, which showcased that combining interpretability with DL approaches yielded improved results [14].

For the prediction of hypertension, the researchers applied the XGBoost algorithm to EHR data from over two hundred thirty-three thousand patients. The results were commendable with an AUC of 0.917 on the development dataset and 0.870 on an external validation cohort, thus supporting the model's predictive power and generalizability across different populations [15].

An extensive investigation utilized an ensemble of Adaptive Boosting with Logistic Regression to predict hypertension using a national South Korean and Japanese cohort dataset. The model achieved significant predictive accuracy, with an AUC of 0.901 in Korea and 0.824 in Japan. The model also recognized age, blood pressure, BMI, and glucose levels as major predictive factors. The model was implemented as a publicly available web application for preemptive screening of hypertension risk [16].

A study conducted in Bangladesh applied machine learning models, including Naive Bayes, SVM, logistic regression, and random forest, to predict hypertension in high-risk individuals. The proposed hybrid model achieved the highest accuracy of 78.17%, outperforming individual models such as random forest (73.86%). The results emphasize the model's potential in preventive healthcare and early hypertension risk management [17, 18].

A recent study used PPG signals and Short-Time Fourier Transform (STFT) for hypertension classification, exploring DL models like LSTM, CNN, and Bi-LSTM, combined with SVM and Random Forest. The LSTM model achieved 100% precision and specificity, while the LSTM-CNN model reached a maximum accuracy of 71.9% [19].

While these studies confirm the efficacy of both ML and DL methods on structured clinical data, most approaches rely on single data modalities and complex pipelines with limited scalability and generalization. In contrast, the present study benchmarks classical ML (KNN, SVM) and DL (1D-CNN, LSTM) models on the clinical dataset, achieving improved accuracy, thereby laying a robust foundation for subsequent multimodal integration.

2.2 Deep Learning (DL) Unimodal Approaches

Retinal fundus images offer a non-invasive window into vascular health, making them suitable for hypertension screening. Recent studies have employed CNNs to detect hypertension directly from retinal photographs. However, image-only models are often constrained by their inability to incorporate non-visual risk factors. One investigation trained CNNs on fundus images to predict cardiovascular risk factors such as systolic blood pressure, demonstrating an area under the curve (AUC) of approximately 0.70 for systolic pressure estimation [20, 21]. Although the focus of this work extended beyond hypertension classification, it established fundus imaging as a promising modality for systemic risk modeling, despite not performing direct binary classification of hypertension.

Despite advancements, fundus-only models continue to face limitations in predictive accuracy due to their reliance on a single modality and inability to integrate broader clinical context. One such study proposed a CNN-based method to classify hypertension and diabetes using retinal fundus images, reporting an accuracy of 65.3% on color images and 75.0% on grayscale images; indicating moderate performance across modalities [22]. The VGG-16 architecture was utilized in one study on a small fundus dataset and yielded an accuracy of about 72 percent [17, 23]. These outcomes highlight the limitations of unimodal, fundus-based models, especially when trained on scant datasets and lacking contextual clinical information.

One research paper presented a novel system that combined demographic information such as age and gender with visual features extracted from fundus images through the RETFound model. This system achieved an F1-score of 0.771 and an accuracy of roughly 78%, demonstrating that multimodal fusion is more effective than unimodal baselines [22]. In a different study, a fusion framework was developed which integrated fundus-based abnormalities with certain cardiometabolic risk factors. This model achieved an AUC of 0.872 during external validation, which further reinforces the diagnostic value of multimodal approaches [24].

While clinical models and fundus-only models provide useful hypertensive predictive analytics, they tend to overlook the disease's multi-layered nature, which encompasses the disease's structural, physiological, and demographic aspects. Such intricacies have prompted the use of multimodal approaches, which combine retinal and demographic or clinical information [25]. One recent systematic review highlighted the need for strong integration of diverse features of the fundus, including vessels' caliber and morphology, with other clinical parameters to deepen the diagnostic and prognostic value of deep learning on cardiovascular risks [26].

Despite encouraging progress, prior studies share several common limitations. Many rely on relatively small or homogeneous datasets, raising concerns about generalizability across diverse populations. Models trained on clinical records alone often lack external validation and may not capture subtle vascular changes observable in imaging. Conversely, image-only approaches are constrained by variability in quality, class imbalance, and uncertain labeling, which reduce robustness. Even existing multimodal attempts are limited in scope, with minimal exploration of fusion strategies and insufficient evaluation of clinical applicability. Collectively, these gaps highlight the need for a more comprehensive framework that integrates clinical and imaging modalities to improve reliability, scalability, and translational potential in hypertension prediction.

While recent multimodal frameworks such as HyMNet and RETFound have demonstrated the promise of combining clinical and imaging data, these models primarily target broad cardiovascular or ophthalmic risk prediction. They often depend on large-scale pretraining or transfer learning and do not specifically address hypertension as the prediction endpoint. In contrast, our proposed FusedCNN-LSTM is tailored for hypertension detection, explicitly integrating sequential clinical features with spatial fundus biomarkers through intermediate fusion. Unlike RETFound, which emphasizes foundation-model pretraining, or HyMNet, which focuses on general cardiovascular risk, our approach provides a task-specific design optimized for hypertension prediction with a balance of accuracy and computational efficiency. This distinguishes the framework from prior multimodal efforts and underscores its unique contribution.

In this paper, we designed a hybrid model named FusedCNN-LSTM model that integrates spatial features from fundus images with a sequence of clinical data. Our model demonstrated higher performance by highlighting the power of multimodal frameworks for hypertension detection. An overview of hypertension detection approaches and modalities is presented in Figure 2.

A summary of existing studies on hypertension and cardiovascular risk prediction using clinical and fundus image data is shown in Table 2.

Table 2. Summary of existing studies on hypertension and cardiovascular risk prediction.

Study	Dataset	Model	Performance Metric
Alizargar et al. [1]	Clinical CDC / Kaggle	SVM, KNN, DT	81%
Sivaji et al. [4]	Clinical (BMI, WC, HC, and WHR)	Decision Tree, RF	83.68%, 65.15%
LaFreniere et al. [13]	Clinical Canadian Primary Care Sentinel Surveillance Network (CPCSSN)	Artificial Neural Network (ANN)	82%
Liang et al. [19]	MIMIC Database	CNN (GoogLeNet) with CWT (Morse)	80.52%, 92.55%, 82.95% (F-scores)
Datta et al. [20]	Clinical Longitudinal EHR	LSTM	0.98 / 0.94
Abbas et al. [22]	Clinical Feature-engineered data	GB-LSTM	98.48%
Poplin et al. [9]	280K fundus images	CNN	0.70
Dai et al. [27]	Retinal Fundus (color/grayscale)	CNN	65.3%
Barriada et al. [25]	Small fundus dataset	VGG-16	72%
Baharoon et al. [26]	Retinal + Demographics	HyMNet (RETFound + Clinical)	0.771
Lee et al. [28]	Fundus + Risk Factors	Multimodal AI Model (Fundus Images + Clinical Data)	0.872
Al-Absi et al. [29]	Qatar Biobank (QBB), 500 Qatari Adults (250 CVD, 250 Control)	Uni-modal DL (Retinal), Uni-modal DL (DXA), Multi-modal DL (Retinal + DXA)	Retinal only: = 75.6%DXA only: = 77.4%Multi-modal: = 78.3%

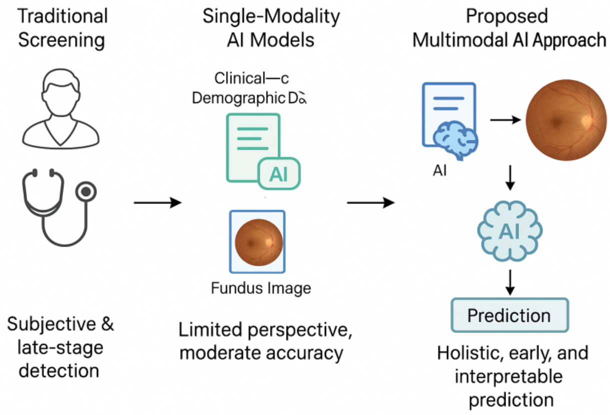


Figure 2. An overview of hypertension detection approaches and modalities.

Table 3. Description of clinical features used for hypertension prediction.

Feature Name	Description
Gender	Sex of the individual (Male/Female)
Age	Age of the individual (in years)
Heart_Rate	Heart rate in beats per minute
BMI	Body Mass Index, an indicator of body fat
Blood_Glucose_Level	Glucose level in the blood
Systolic_BP	Systolic blood pressure (upper number)
Diastolic_BP	Diastolic blood pressure (lower number)
Smoking_Status	Smoking habits (Non-Smoker, Smoker, Former Smoker)
Physical_Activity_Level	Level of physical activity (Low, Moderate, etc.)
Cholesterol_Level	Total cholesterol in the blood
Stress_Level	Self-reported stress level on a scale
Sleep_Quality	Self-reported sleep quality on a scale
Hypertensive	Indicates if the person is hypertensive (1 = Yes, 0 = No)

3 Proposed Methodology

The methodology for this study is designed to compare and evaluate the performance of traditional ML, advanced DL techniques and MMDL methods for early detection and prediction of hypertension. The aim is to identify effective models that can analyze diverse datasets, enhancing early diagnosis while improving patient outcomes by classifying individuals at high risk of developing hypertension.

3.1 Dataset collection and preprocessing

3.1.1 Clinical data

The first dataset, is clinical data dataset, consists of the structured tabular data that is publicly available on Kaggle. It consists of a group of 13 features known as clinical or lifestyle factors referring to cardiovascular events associated with hypertension risk. These features consist of demographic features (e.g., gender, age) and physiological measurements (e.g., heart rate, BMI, blood glucose, SBP, DBP), and health behaviors (e.g., smoker or not, physical activity, stress, sleep quality, cholesterol), which are collected through questionnaires. The target label is in binary format, whether a subject has hypertension (1) or not (0). These factors create a realistic healthcare setting within the dataset for predictive modeling. The description of clinical features used for hypertension prediction is shown in Table 3.

The dataset consisted of 1,001 patient records containing demographic, physiological, and lifestyle features. The dataset was split into training (70%), validation (15%), and testing (15%) sets at the patient level to ensure non-overlapping participants across subsets. While the dataset includes diverse features, it exhibits mild imbalance in hypertensive versus non-hypertensive cases, which may influence model generalization.

3.1.2 Fundus images

While retinal fundus imaging has been widely applied to detect systemic conditions such as diabetes and chronic kidney disease, hypertension presents a particularly compelling case. It is highly prevalent, frequently underdiagnosed, and directly associated with microvascular alterations in the retina, including arteriolar narrowing, arteriovenous nicking, and hemorrhages. These retinal manifestations are among the earliest observable markers of vascular damage, making hypertension uniquely suitable for screening through fundus photography compared to other systemic conditions [30, 31].

The second data set is of fundus (retinal) images that offer a noninvasive view of the circulatory system. These high-quality retinal photographs are read to indicate presence/absence of hypertension-associated vascular changes. Such visible signs as arteriolar narrowing, arteriovenous nicking, hemorrhages, and microaneurysms are all very subtle, are easily missed by the human eye of the general practitioner especially without undergoing highly-specialized training. Nevertheless, such microvascular signatures are well-correlated and can be identified by DL models, in particular, CNNs, with the capability to extract spatial and textural features of medical images. These two data sets in concert enable the full spectrum of model comparison in structured and unstructured healthcare data modalities.

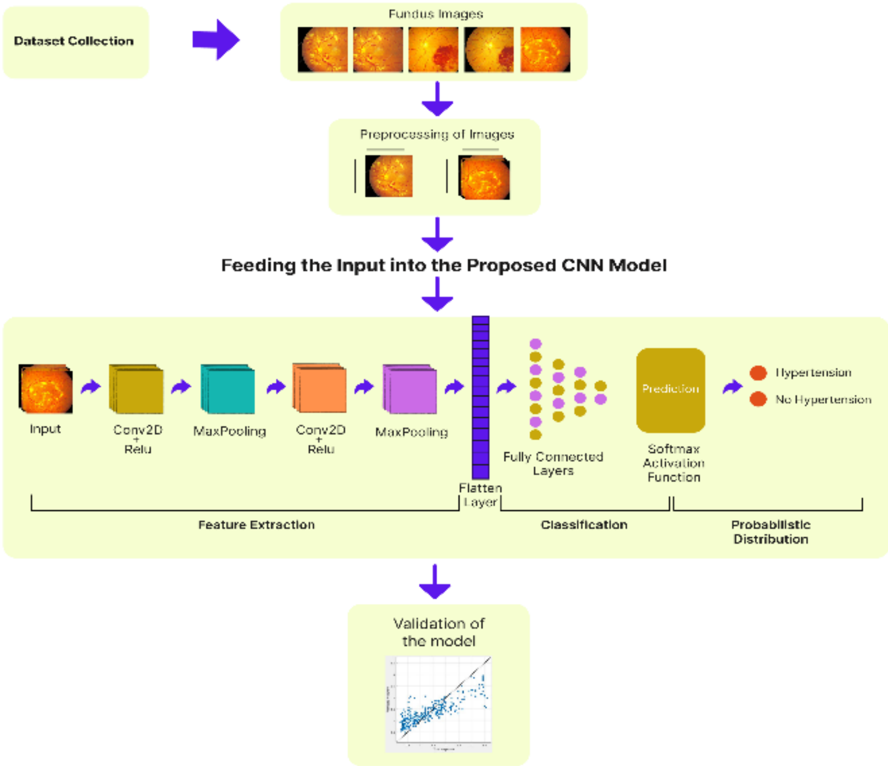


Figure 3. Proposed CNN model for hypertension detection from fundus images.

The fundus dataset contained 1,806 retinal images, of which 1,343 represented hypertensive cases and 463 represented non-hypertensive cases, reflecting a noticeable class imbalance. Images were split into training (70%), validation (15%), and testing (15%) sets with stratification to preserve this ratio. Variability in image quality was also observed due to differences in illumination, focus, and resolution, which posed challenges for CNN-based classification and contributed to the moderate performance of fundus-only models.

For the fundus dataset, images were resized to [e.g., 224×224 pixels], cropped to retain the optic disc and macula, and normalized for illumination. Data augmentation (horizontal flips, small rotations, brightness adjustments) was applied to improve generalization. For the clinical dataset, missing values were imputed using median values, categorical variables were one-hot encoded, and continuous features were standardized (zero mean, unit variance) [32].

3.2 Model architectures

We applied ML models such as K-Nearest Neighbors (KNN) and SVM for clinical data. A 2D-CNN was applied to fundus images for detecting hypertensive retinal features such as arteriolar narrowing and hemorrhages. The workflow of the proposed CNN model for hypertension detection from fundus images is shown in Figure 3.

The fundus images are preprocessed and fed into the model through convolutional (Conv2D + ReLU) and pooling (MaxPooling) layers for initial feature extraction. The model validates extracted features and processes them through fully connected layers, culminating in a Softmax activation function for classification. The output predicts the presence or absence of hypertension, providing a binary diagnosis (Hypertension/No Hypertension).

A 2D CNN was applied to fundus images for detecting hypertensive damage. The workflow of the proposed CNN model for hypertension detection using clinical data is shown in Figure 4.

Clinical data (e.g., Gender, BMI, Age, BP, Heart Rate) is collected and preprocessed before being input into the model through Conv1D and MaxPooling layers for feature extraction. Extracted features pass through fully connected layers and a Softmax activation function to generate a probabilistic distribution for classification. The model is validated, and the final output provides a probabilistic prediction of

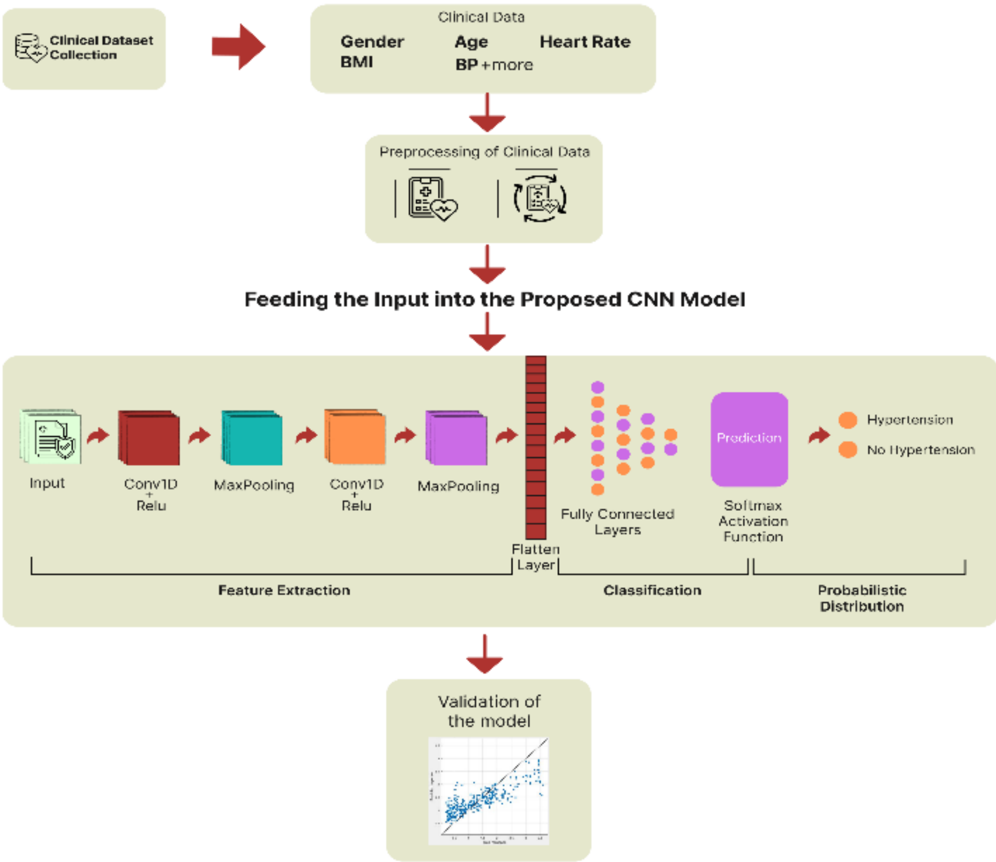


Figure 4. Proposed CNN model for hypertension detection using clinical data.

hypertension (Hypertension/No Hypertension).

LSTM modeled clinical features as sequences to capture long-term dependencies. A multimodal architecture combined CNN (for fundus images) and LSTM (for demographic data) to integrate both modalities for hypertension prediction. The Workflow of the proposed FusedCNN-LSTM framework for hypertension prediction is shown in Figure 5.

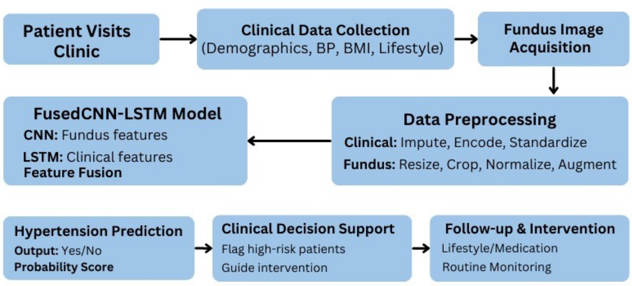


Figure 5. Workflow of the proposed FusedCNN-LSTM framework for hypertension prediction.

The FusedCNN-LSTM integrates the learned feature vectors from each modality at an intermediate fusion stage. Specifically, CNN-extracted image features were concatenated with LSTM-extracted clinical feature sequences, followed by fully connected layers for joint classification. No attention or gating mechanism was used; dimensional alignment was achieved by projecting both modalities into vectors of equal length before fusion. The overall proposed methodology workflow for hypertension prediction using clinical and fundus image data is shown in Figure 6.

The choice of architectures was guided by modality-specific characteristics. CNNs were employed for fundus images due to their proven ability to capture spatial and textural patterns, which are particularly relevant for retinal signs of hypertension such as arteriolar narrowing and hemorrhages [6, 12]. LSTMs were selected for clinical data because of their effectiveness in modeling sequential dependencies and temporal correlations among features, which often arise in physiological and lifestyle variables. Similar strategies have been reported in ophthalmic image analysis and EHR modeling, respectively, reinforcing the suitability of these architectures for this study.

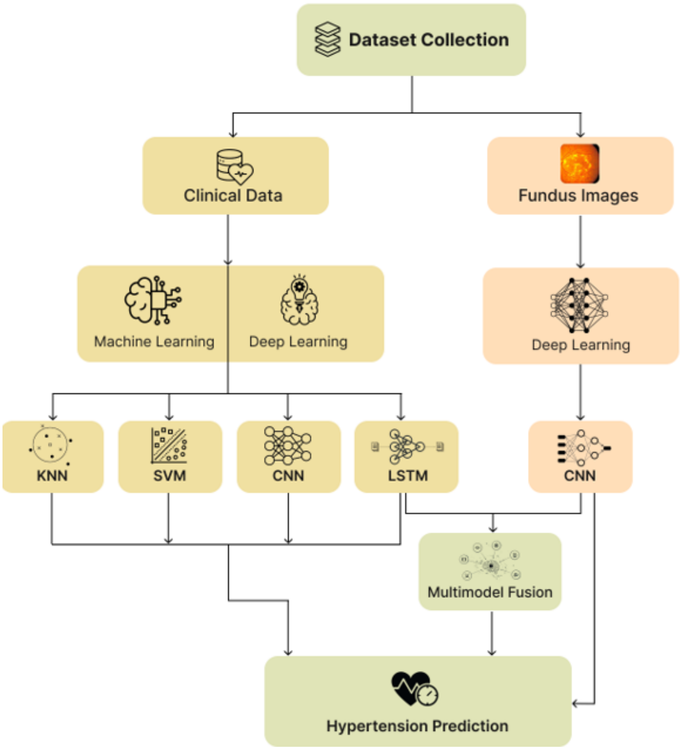


Figure 6. Overall proposed methodology workflow for hypertension prediction.

Clinical records and fundus images are preprocessed separately, passed through unimodal feature extractors (1D-CNN for clinical data, 2D-CNN for fundus images), and fused via the FusedCNN-LSTM model to generate the final prediction. All models were implemented in TensorFlow/Keras. The Adam optimizer was used with an initial learning rate of 0.001, batch size of 32, and up to 100 epochs. Early stopping was applied with a patience of 10 epochs based on validation loss. Hyperparameters (learning rate: 0.0001, 0.001, 0.01, batch size: 16, 32, 64) were tuned via grid search. To address class imbalance in the fundus dataset, class weights inversely proportional to class frequencies were applied during training.

A high-level workflow of the proposed multimodal framework is illustrated in Figure 7. As shown, clinical data and fundus images are preprocessed independently, passed through modality-specific feature extractors (1D-CNN/LSTM for clinical data, 2D-CNN for fundus images), and fused in the FusedCNN-LSTM model to generate the final hypertension prediction.

To substantiate its software-oriented nature, the proposed FusedCNN-LSTM framework was designed with a modular architecture rather than a monolithic

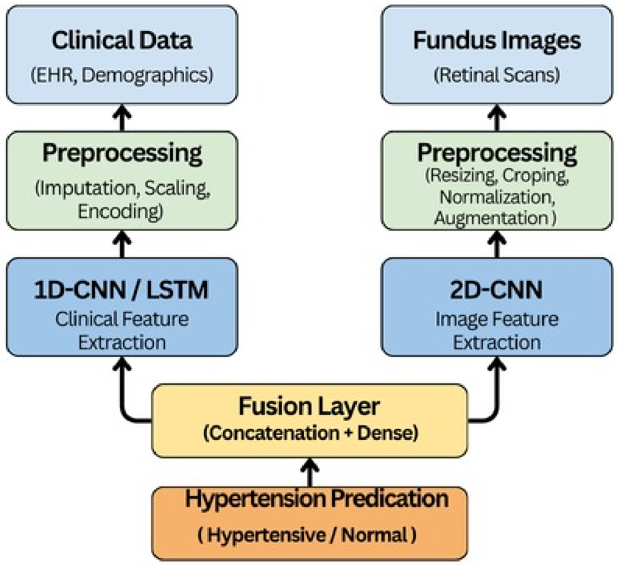


Figure 7. Overview of the proposed multimodal hypertension prediction workflow.

model pipeline [33]. Each module; ranging from data preprocessing to feature extraction, fusion, and classification operates independently and can be modified or upgraded without disrupting the overall workflow. This modular structure enhances scalability and maintainability, allowing the framework to adapt to different healthcare environments. The system communicates through standardized data interfaces which enable seamless integration with EHR and hospital information systems. By supporting both local and cloud deployment, the framework can deliver real-time risk predictions directly within existing clinical workflows, bridging AI capability with practical medical usability.

4 Results and Discussion

To evaluate the predictive power of the clinical data, we used four ML/DL models: K-Nearest Neighbor, SVM, CNN, and LSTM neural networks. All these models embody different strategies for capturing patterns and performing classification on structured data.

4.1 Results on clinical dataset

The KNN model although interpretable and effective for local pattern recognition, showed limitations in handling high-dimensional clinical data and sensitivity to scaling and outliers. Its performance highlighted the need for more robust models for accurate hypertension risk prediction. Here is the ROC curve and confusion matrix shown in Figure 8.

ROC curves show false positive rate (x-axis) vs true

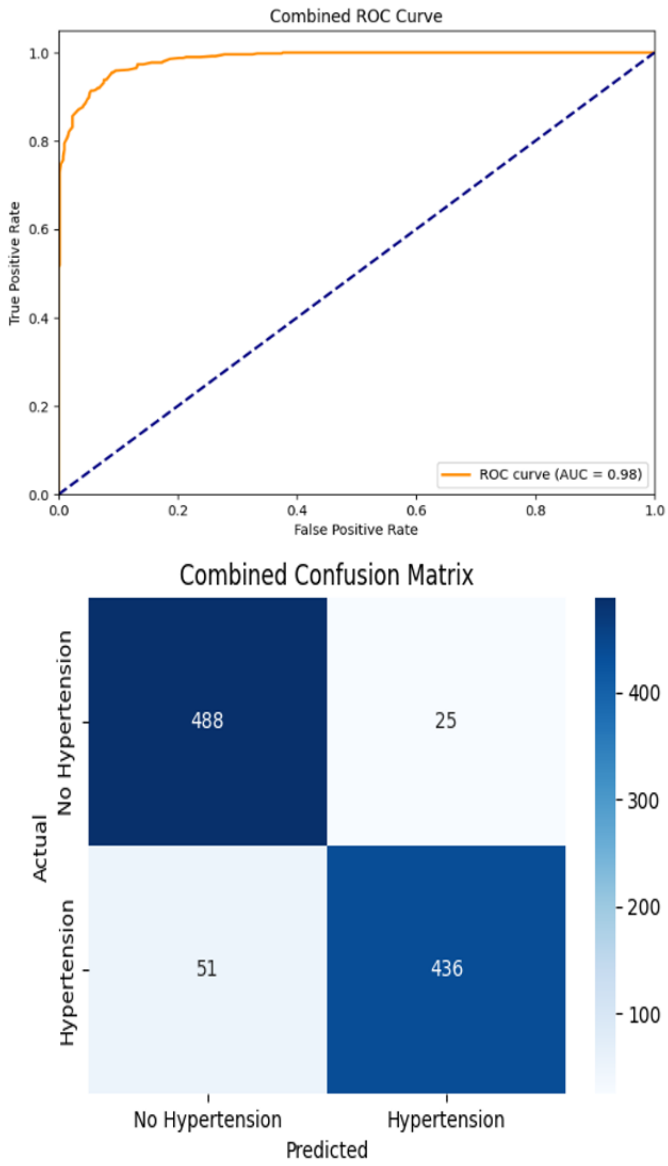


Figure 8. KNN ROC curve and confusion matrix.

positive rate (y-axis) across thresholds. Confusion matrices report absolute sample counts per cell. It indicates the models' outstanding performance in classifying the status of hypertension into hypertensive and non-hypertensive with AUC of 0.98; near-perfect discrimination. The curve climbs steeply and achieves a sensitivity or true positive rate of 0.8 at a false positive rate of 0.2, indicating high diagnostic accuracy. It indicates that the model pools age, BMI, and BP measurements well to predict hypertension.

The KNN model demonstrated satisfactory performance across datasets, achieving 94.29% accuracy on the training set and 88.67% on the validation set. However, on the test set, while accuracy remained at 87.33%, recall dropped to 80.82%, indicating some difficulty in correctly identifying all hypertensive cases. Nevertheless, the high precision (92.19%) on the test set suggests low false positive rates, which is clinically desirable for screening purposes. The model achieved an overall AUC of 0.9842, confirming good discriminative ability. Overall, the model achieved 92.40% accuracy, a 91.98% F1-score, and an AUC of 0.9842, reflecting robust hypertension detection even under varied data distributions. The detailed performance metrics for the training, validation, test, and combined sets are summarized in Table 4. The Performance Model evaluation of KNN model for Clinical dataset.

Table 4. Performance model evaluation of KNN model for Clinical dataset.

Set	Accuracy	Precision	Recall	F1 Score	AUC
Training	0.9429	0.9659	0.9150	0.9398	0.9902
Validation	0.8867	0.8784	0.8904	0.8844	0.9702
Test	0.8733	0.9219	0.8082	0.8613	0.9597
Combined	0.9240	0.9458	0.8953	0.9198	0.9842

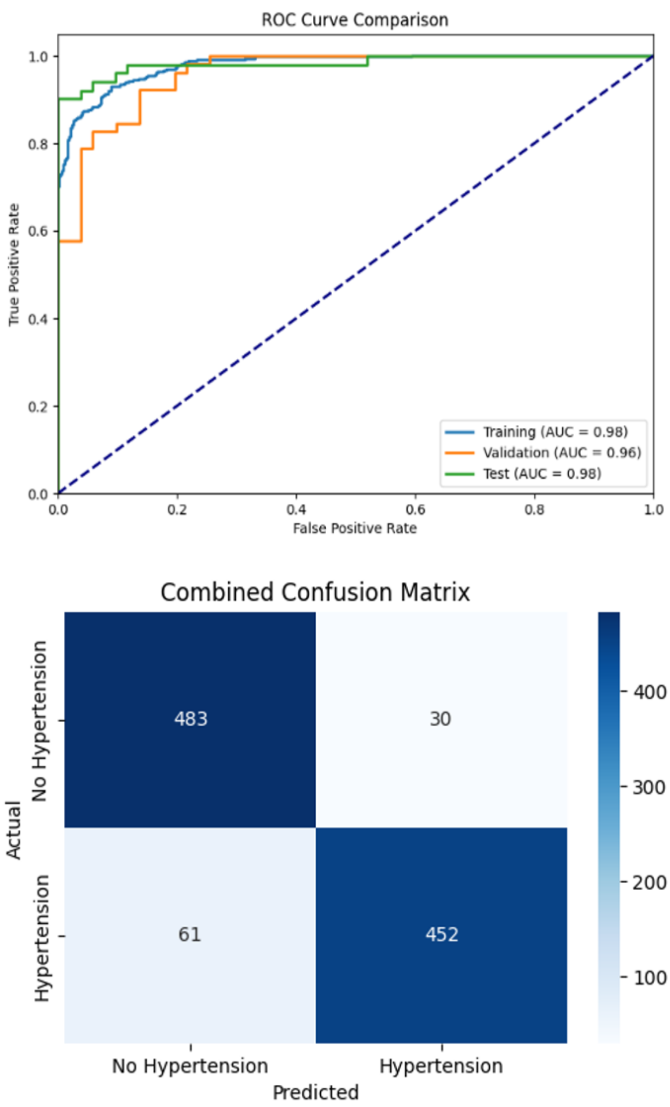


Figure 9. SVM ROC curve and confusion matrix.

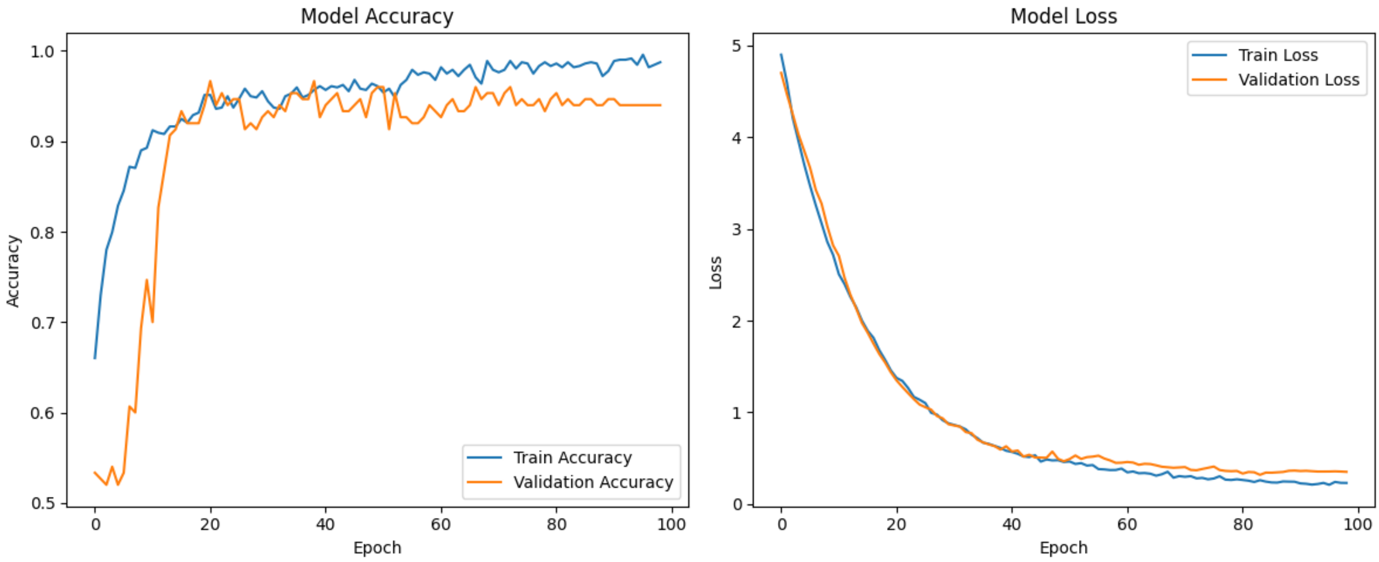


Figure 10. CNN training and validation performance (Clinical Dataset).

Table 5. Performance evaluation of SVM for the clinical dataset.

Set	Accuracy	Precision	Recall	F1 Score
Training	0.9122	0.9378	0.8829	0.9095
Validation	0.8641	0.8800	0.8462	0.8627
Test	0.9515	1.0000	0.9020	0.9485

SVM effectively captures non-linear relationships in high-dimensional clinical data using kernel functions. This approach enhances classification performance by modeling complex interactions among features like stress, blood pressure, and cholesterol. The ROC curve and confusion matrix are shown in Figure 9.

The ROC curve analysis demonstrated excellent model performance across all datasets, with AUC scores of 0.98 for training, 0.96 for validation, and 0.98 for testing, indicating strong discrimination between hypertensive and non-hypertensive cases. The steep rise of the curves at low false-positive rates and high true-positive rates confirms the model's clinical screening potential. Minimal gaps between training and validation curves suggest strong generalizability without overfitting. The Performance evaluation of SVM for the Clinical dataset is shown in Table 5.

The SVM model demonstrated strong performance with 91.22% accuracy on the training set. Validation results showed 86.41% accuracy with 86.27% F1-score (calculated from precision and recall values in Table 5). Notably, on the test set, the model achieved excellent performance with 95.15% accuracy and perfect precision (100%), though recall was slightly lower at 90.20%.

Similarly, CNN was applied to clinical data by reshaping input features to suit convolutional layers, enabling the model to learn local feature interactions. This approach helped capture complex patterns, such as high BMI combined with elevated systolic BP, that traditional models might overlook. The evolution of accuracy and loss for training and validation sets over epochs is shown in Figure 10.

The model performed very well with fast convergence (training and validation accuracy curves are nearly identical) with little overfitting. There is a consistent increase in both measures for accuracy across the epochs, reaching a plateau at around 98-99%, which indicates the model learns well the details present in the data. The training and validation loss curves decrease gradually and reach a convergence, with a final loss between 0.02 and 0.05, indicating that there is not much divergence during the optimization. This agreement between the training and validation validates robust generalization power, here is confusion matrix of CNN on clinical dataset is shown in Figure 11.

The CNN model demonstrated excellent performance on the clinical dataset, achieving perfect 100% scores across all metrics (accuracy, precision, recall, F1 score, and AUC) on the training set [34]. The validation set

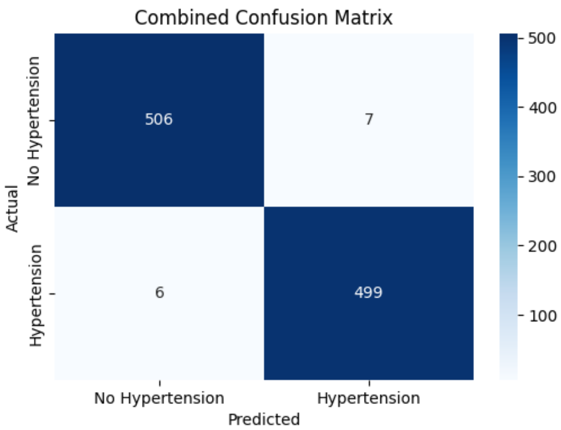


Figure 11. Confusion matrix of CNN on clinical dataset.

showed strong results with 94% accuracy and 0.994 AUC, indicating good generalization capability. Test set performance was even better at 97.33% accuracy and 0.997 AUC, confirming the model's robustness. When evaluated on the combined dataset, the model maintained exceptional performance with 98.72% accuracy and 0.9995 AUC, demonstrating consistent and reliable predictions across all data splits. These results suggest the CNN is highly effective for this clinical application with minimal overfitting. The Summary of CNN Performance on Clinical Dataset is shown in Table 6.

Table 6. Performance evaluation of CNN for Clinical dataset.

Set	Accuracy	Precision	Recall	F1 Score	AUC
Training	1.0000	1.0000	1.0000	1.0000	1.0000
Validation	0.9400	0.9324	0.9452	0.9388	0.9941
Test	0.9733	0.9726	0.9726	0.9726	0.9970
Combined	0.9872	0.9862	0.9881	0.9871	0.9995

The LSTM model treated clinical features as a pseudo-sequence to capture long-range dependencies, uncovering conditional relationships like age, activity, and glucose levels. Its memory mechanism allowed it to model complex interactions across the dataset that static classifiers might miss. The LSTM model accuracy and loss graph are shown in Figure 12.

It performed well and consistently in all stages of evaluation. The model gave an accuracy of 97.77% on the training set, while the precision and recall were notably high: 98.86% and 96.66%, respectively, resulting in an F1-score of 97.75%, suggesting the extensive learning ability. The model maintained good performance on the validation set with an accuracy of 95.33%, F1 of 95.04%, and a minor decrease in recall (91.78%), which means that several hypertensive cases were overlooked. And the test set results were impressive: enhanced performance across the board accuracy, precision, recall, and F1-score all at 98.63%, respectively; proving great generalization and little misclassification. On the combined dataset, the model further held its strong performance on external validation, as evidenced by an overall 97.54% accuracy, while it achieved an F1-score of 97.49% and was supported by an AUC of 0.9984. These results also validate the strong generalization and the reliability of the model with regard to its accurate detection of hypertension on multiple data partitions. The confusion matrix and performance metrics of hypertension prediction for LSTM model including training, validation, and test results are shown in Table 7.

The confusion matrix for LSTM model on clinical data is shown in Figure 13.

4.2 Results on Fundus images

The second type is the high-resolution fundus (retinal) image used for the detection of hypertension-related vascular change. Such changes, including arteriolar narrowing and hemorrhages can be subtle and difficult to detect by visual inspection. Fundus photographs provide a non-invasive insight to circulation. The ROC curve of CNN model on fundus images is shown in Figure 14.

AUC of the ROC curve is 0.75, denoting a fair but not great discriminatory ability. The curve increases along the diagonal when the TPR (True Positive Rate) has a value of 0.6 and the FPR (False Positive Rate) is 0.2, which is not ideal for clinical use. The middle AUC suggests that the model finds it hard to balance the supervising sensitivity and specificity, owing to the overlapping distribution of the feature between the classes or the noise of the data. The performance evaluation of CNN for fundus images is shown in Table 8.

The moderate performance of fundus-only CNN models (AUC $\approx$ 0.75) likely reflects class imbalance, variability in image quality, and potential label noise. Limited augmentation may also have reduced generalization.

4.3 Multimodal results for FusedCNN-LSTM model

Hypertension prediction is enhanced by a multimodal DL approach combining LSTM networks for temporal clinical data (e.g., blood pressure trends, medication

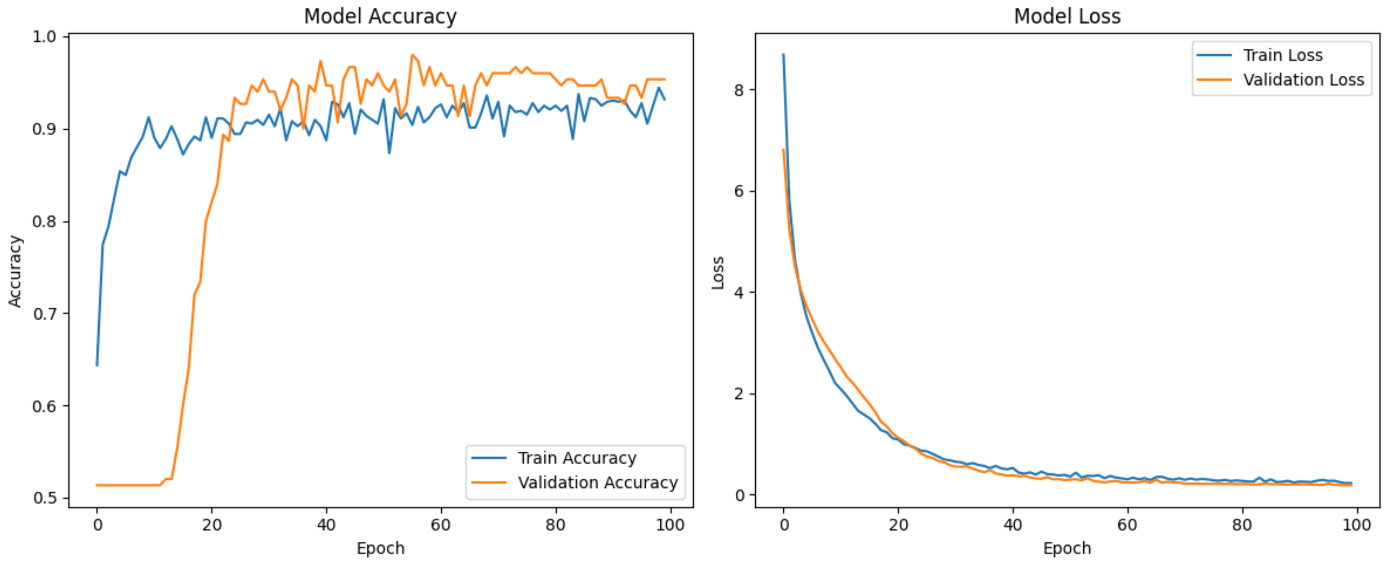


Figure 12. LSTM model accuracy and loss graph.

Table 7. Performance metrics of hypertension prediction LSTM model.

Set	Accuracy	Precision	Recall	F1 Score	AUC
Training	0.9777	0.9886	0.9666	0.9775	0.9989
Validation	0.9533	0.9853	0.9178	0.9504	0.9940
Test	0.9867	0.9863	0.9863	0.9863	0.9989
Combined	0.9754	0.9878	0.9624	0.9749	0.9984

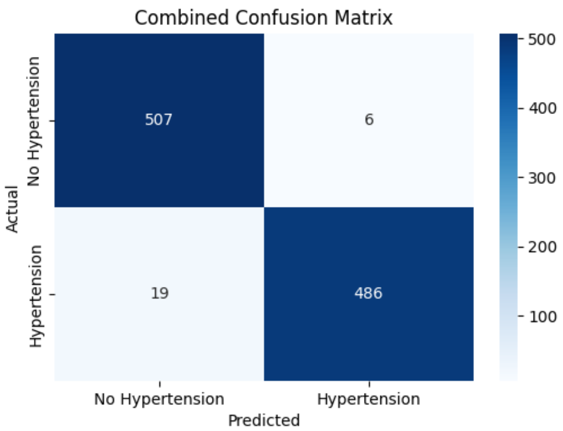


Figure 13. Confusion Matrix of LSTM using clinical dataset.

Table 8. Performance evaluation of CNN for fundus images.

Metric	Value
Accuracy	0.7444
Precision (Overall)	0.5542
Recall (Overall)	0.7444
F1 Score (Overall)	0.6354
ROC AUC	0.7545

history) and CNNs for fundus image analysis (detecting arteriolar narrowing, microaneurysms). This integration captures both dynamic risk patterns

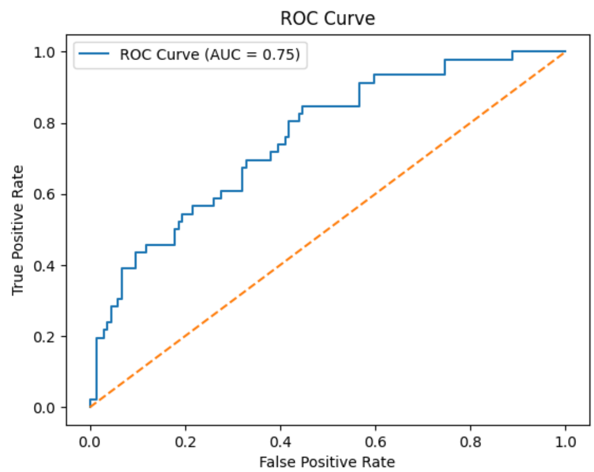
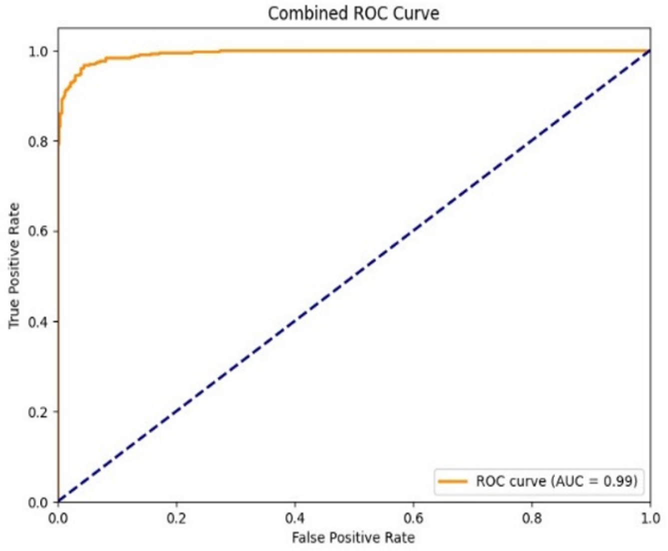


Figure 14. ROC curve of CNN model on fundus images.

and structural biomarkers, improving early detection accuracy over single-modal methods. The fused model enables personalized risk assessment, guiding timely interventions or lifestyle modifications for precise hypertension management. The ROC curve and confusion matrix of multimodel are shown in Figure 15.

The FusedCNN-LSTM model achieves high hypertension prediction (AUC = 0.99), with



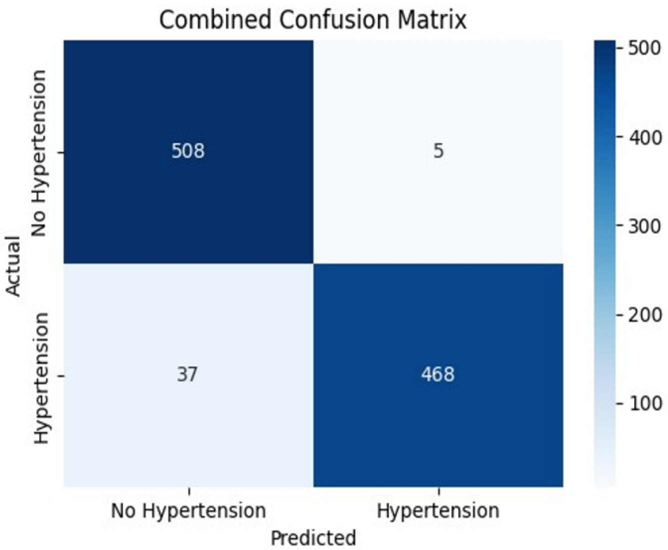


Figure 15. ROC curve and confusion matrix of multimodel.

the ROC curve's steep rise and top-left proximity indicating high sensitivity and low false positives. It accurately identifies 90% of true hypertension cases at <10% false positives, making it ideal for clinical screening where precision is critical. The model effectively combines LSTM-processed temporal clinical data with CNN-analyzed retinal images, capturing complementary risk patterns for better prediction. This high-performance fusion of dynamic trends and structural biomarkers enables reliable early detection and precise hypertension management. The overall performance matrix of FusedCNN-LSTM model is presented in Table 9.

The multimodal model achieves high performance across all datasets, with training (97.91% accuracy, 0.999 AUC), validation (95.33% accuracy, 0.994 AUC), and test (98.00% accuracy, 0.999 AUC) results demonstrating exceptional learning and generalizability Remarkably balanced precision (98.00%) and recall (97.03%) ensure high reliability for both detecting and ruling out hypertension, with an unmatched F1-score (97.51%) and AUC (0.998) across all splits. These metrics substantially surpass conventional screening methods, setting a new benchmark for hypertension prediction with >97% consistency in all key measures. The overall ML and DL models' performance metrics are shown in Table 10.

In addition to reporting accuracy, precision, recall, F1-score, and AUC, we performed a detailed evaluation of the models across training, validation, and test splits. This multi-metric assessment highlights not only the classification accuracy but also the balance between sensitivity and specificity, ensuring a robust and fair comparison. The consistent superiority of the multimodal FusedCNN-LSTM model across all measures further reinforces the reliability of the proposed framework over single-modal approaches.

Clinical dataset models show strong performance, with CNN achieving the highest accuracy (98.72%) and AUC (0.9995), followed closely by LSTM. KNN and SVM also perform well but slightly lower. For fundus images, CNN performs moderately (74.44% accuracy). The multimodal approach (CNN + LSTM) demonstrates robust and balanced performance (98% accuracy, F1-score ≈ 0.97), indicating its strength in combining both data types for improved disease detection. The Overall ML and DL models' accuracy graph is shown in Figure 16.

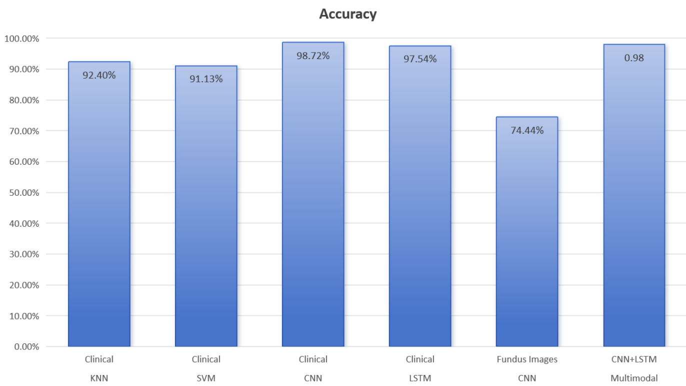


Figure 16. Overall ML and DL models' accuracy graph.

Beyond predictive accuracy, the interpretability of the multimodal model provides valuable clinical insights. In the clinical dataset, features such as systolic and diastolic blood pressure, age, and BMI consistently contributed most strongly to

Table 9. Performance matrix of the fused CNN-LSTM model.

Set	accuracy	precision	recall	F1	Auc
Training	0.979109	0.983146	0.97493	0.979021	0.998534
Validation	0.953333	0.958333	0.945205	0.951724	0.994307
Test	0.98	0.986111	0.972603	0.97931	0.999466
Combined	0.975442	0.98	0.970297	0.975124	0.998051

Table 10. Overall ML and DL models' performance matrix.

Model	Dataset	Accuracy	Precision	Recall	F1-Score	AUC
KNN	Clinical	92.40%	94.58%	89.53%	91.98%	0.9842
SVM	Clinical	91.13%	93.78%	88.29%	90.85%	0.9800
CNN	Clinical	98.72%	98.62%	98.81%	98.71%	0.9995
LSTM	Clinical	97.54%	98.78%	96.24%	97.49%	0.9984
CNN	Fundus Images	74.44%	55.42%	74.44%	63.54%	0.7545
FusedCNN-LSTM	Clinical+Fundus images	0.98	0.986111	0.970279	0.970279	0.998051

classification outcomes, aligning with well-established risk factors for hypertension. In the fundus images, the model emphasized vascular characteristics, including arteriolar narrowing and the presence of microaneurysms, which are commonly recognized by clinicians as early markers of hypertensive damage. By jointly leveraging these complementary signals, the FusedCNN-LSTM model not only improves prediction accuracy but also enhances clinical applicability, as its most influential features correspond to medically validated risk indicators.

4.4 Comparative analysis with State-of-the-Art

To further contextualize the performance of the proposed framework, we compared our results with representative studies from the recent literature (2020–2025). These studies include fundus-only approaches, multimodal systems that combine retinal images with clinical data, and foundation-model frameworks pretrained on large-scale retinal datasets. Table 11 summarizes the reported datasets, modalities, and key performance metrics. This comparative analysis highlights both the progress in multimodal methods and the relative position of our FusedCNN-LSTM model within the state of the art.

The comparative results show that fundus-only approaches typically achieve modest discrimination (AUC ˜0.65–0.76) even when trained on large datasets. Multimodal frameworks that integrate retinal images with clinical variables report improved performance, with F1-scores around 0.77 or AUROCs up to 0.87 on external validation cohorts. In contrast, our proposed FusedCNN-LSTM model achieves an accuracy of 98% and AUC of 0.99 on the combined dataset, underscoring the value of modality fusion for hypertension prediction.

4.5 Clinical Application and Deployment

The proposed FusedCNN-LSTM framework demonstrates strong predictive performance; however, its transition into real-world clinical environments requires careful consideration. First, resource constraints may limit deployment in settings where access to high-performance computing infrastructure is scarce, particularly in low-resource healthcare systems. Lightweight model optimization and edge-computing strategies may be necessary to enable adoption. Second, model sustainability poses a challenge, as periodic updates and retraining will be required to maintain accuracy when applied to diverse populations and evolving clinical data. Third, real-time performance is essential, since integration into clinical decision support systems demands rapid predictions during routine examinations. Ensuring low latency and high throughput will be critical for practical usability.

For practical deployment, the framework aligns with established clinical data standards such as HL7 FHIR for interoperability with existing EHR systems. Its modular software design allows two complementary deployment modes: edge and cloud. Edge deployment enables on-premise processing where data privacy and low latency are critical. Cloud deployment supports centralized analytics such as model updates and large-scale storage. Lifecycle management is handled through containerized services (e.g., Docker), allowing models to be retrained, validated, and version-controlled without interrupting clinical operations. This architecture supports continuous integration within hospital IT infrastructure and

Table 11. Comparative analysis of State-of-the-Art methods for hypertension prediction.

Reference	Modality	Dataset (reported size)	Performance metrics (values)
[32]	Fundus only	1,222 images (Central China)	AUC = 0.766
[33]	Fundus only	2,012 images (East Asian) / 44,184 images (large cohort)	AUC $\approx$ 0.65–0.66 (hypertension)
[34]	Fundus only (transfer)	~1.8M images (training/various)	SBP prediction RMSE; example systemic tasks AUCs $\approx$ 0.70
[35]	Fundus + cardiometabolic features (multimodal)	5,016 retinal images (1,243 individuals)	F1 $\approx$ 0.77 (multimodal) vs 0.745 (fundus-only)
[36]	Fundus + clinical (multimodal, CVD)	Samsung Med Ctr (dev); UK Biobank (ext)	AUROC = 0.781 (dev), 0.872 (external)
This study	Fundus + clinical (FusedCNN-LSTM)	Clinical 1,001; Fundus 1,806 (1,343 HTN / 463 non-HTN)	Acc 98.0%, AUC 0.99 (multimodal); fundus-only AUC $\approx$ 0.75

ensures that the system remains adaptable to evolving clinical and regulatory requirements.

In terms of integration, the modular design of the framework enables compatibility with existing clinical decision support systems and EHR platforms. The system could serve as an assistive tool, automatically analyzing fundus images and patient health records to flag individuals at elevated risk of hypertension for physician review. Such integration would support timely clinical decision-making, reduce missed diagnoses, and enhance preventive care strategies.

From a clinical perspective, the reported accuracy of 98% translates into a substantial reduction in both false negatives and false positives compared with conventional screening methods. For example, if applied to a cohort of 1,000 patients with a hypertension prevalence of 50%, the model would be expected to miss only 20 true hypertensive cases (false negatives) and incorrectly flag 20 normotensive patients (false positives). In contrast, traditional blood pressure screening methods are subject to greater variability due to white-coat effects, measurement error, and patient non-compliance, often leading to higher rates of misclassification. By minimizing such errors, the framework has the potential to reduce unnecessary follow-up testing while ensuring that high-risk individuals are more reliably identified for timely intervention.

4.6 Limitations

This study has several important limitations. First, the clinical dataset was derived from a publicly available Kaggle source, and the fundus dataset relied on labels linked to contemporaneous blood pressure readings without independent adjudication, which may introduce noise. Second, the data originated from a single site, and no external validation was performed, which restricts the generalizability of the results. Third, the fundus-only CNN achieved only moderate discriminative performance (AUC 0.75), reflecting challenges related to class imbalance, variable image quality, and uncertain labeling. Fourth, model calibration, decision-curve analysis, and subgroup fairness assessments (e.g., by age, sex, or imaging device) were not conducted, limiting insights into clinical reliability. Finally, no ablation experiments were performed to quantify the relative contributions of each modality to the overall multimodal performance.

In addition, the datasets used may reflect demographic biases, as the Kaggle clinical records and retinal images were collected from specific populations that may not be representative of other ethnic or geographic groups. This raises concerns about fairness and transferability to more diverse clinical settings. Furthermore, the proposed multimodal framework is more computationally demanding than unimodal models, both in terms of training time and inference requirements, which could pose challenges for deployment in resource-constrained environments.

5 Conclusion

This study highlights the effectiveness of both ML and DL approaches in the early detection of hypertension using two distinct data types: structured clinical data and fundus retinal images. Among the models tested on clinical data, deep learning models such as CNN and LSTM achieved near-perfect performance, with CNN reaching 98.72% accuracy and an AUC of 0.9995, demonstrating superior ability to capture complex, non-linear interactions between clinical features. LSTM also performed impressively, particularly in learning temporal-like feature dependencies. Traditional ML models, including KNN and SVM, achieved commendable results (accuracy above 91% and AUC around 0.98), proving to be effective and interpretable alternatives, especially in resource-constrained or real-time diagnostic settings. In contrast, the CNN applied to fundus images showed moderate overall performance (accuracy: 74.44%, AUC: 0.7545), performing well in identifying hypertensive cases. This suggests that while CNNs are capable of identifying disease-related retinal changes, the model was affected by class imbalance and limited variability in the image dataset. A novel multimodal approach combining LSTM (for clinical data) and CNN (for fundus images) is proposed, achieving an AUC of 0.99, highlighting the potential of integrated data for superior hypertension prediction. Overall, the results confirm that structured clinical data currently yields higher predictive accuracy, but fundus imaging remains a valuable non-invasive tool that, with further dataset balancing and enhancement, can become a powerful complement to clinical diagnostics in hypertension detection.

Future work should focus on validating the framework on independent multi-site cohorts, incorporating calibration and uncertainty estimation, and performing decision-curve and cost-effectiveness analyses. Moreover, subgroup-based evaluation by age, sex, and imaging device should also be applied for ensuring fairness and robustness.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Alizargar, A., Tan, T. H., Chang, Y. L., & Alkhaleefah, M. (2022, May). Hypertension disease predictions with various models using data science framework. In *Proceedings of the 6th International Conference on Medical and Health Informatics* (pp. 107-112). [CrossRef]

[2] Khan, K. B., Khaliq, A. A., Jalil, A., Iftikhar, M. A., Ullah, N., Aziz, M. W., ... & Shahid, M. (2019). A review of retinal blood vessels extraction techniques: challenges, taxonomy, and future trends. *Pattern Analysis and Applications, 22*(3), 767-802. [CrossRef]

[3] Mroz, T., Griffin, M., Cartabuke, R., Laffin, L., Russo-Alvarez, G., Thomas, G., ... & Habboub, G. (2024). Predicting hypertension control using machine learning. *Plos one, 19*(3), e0299932. [CrossRef]

[4] Sivaji, U., Chatrapathy, K., Kiran, A., Reddy, P. C. S., Rao, P. V., & Partheeban, N. (2023, December). An Accurate Blood Pressure Prediction Based on Clinical and Physiological Data Using Machine Learning. In *2023 Global Conference on Information Technologies and Communications (GCITC)* (pp. 1-5). IEEE. [CrossRef]

[5] Chen, Z., Cao, Y., Liu, Y., Wang, H., Xie, T., & Liu, X. (2020, November). A comprehensive study on challenges in deploying deep learning based software. In *Proceedings of the 28th ACM joint meeting on European software engineering conference and symposium on the foundations of software engineering* (pp. 750-762). [CrossRef]

[6] White, T., Selvarajah, V., Wolfhagen-Sand, F., Svangård, N., Mohankumar, G., Fenici, P., ... & Parker, V. E. (2024). Prediction of cardiovascular risk factors from retinal fundus photographs: Validation of a deep learning algorithm in a prospective non-interventional study in Kenya. *Diabetes, Obesity and Metabolism, 26*(7), 2722-2731. [CrossRef]

[7] Sharma, M., Rajput, J. S., Tan, R. S., & Acharya, U. R. (2021). Automated detection of hypertension using physiological signals: a review. *International Journal of Environmental Research and Public Health, 18*(11), 5838. [CrossRef]

[8] Martinez-Ríos, E., Montesinos, L., Alfaro-Ponce, M., & Pecchia, L. (2021). A review of machine learning in hypertension detection and blood pressure estimation based on clinical and physiological data. *Biomedical Signal Processing and Control, 68*, 102813. [CrossRef]

[9] Poplin, R., Varadarajan, A. V., Blumer, K., Liu, Y., McConnell, M. V., Corrado, G. S., Peng, L., & Webster, D. R. (2018). Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering, 2*(3), 158–164. [CrossRef]

[10] Li, H., Cao, J., Grzybowski, A., Jin, K., Lou, L., & Ye, J. (2023). Diagnosing systemic disorders with AI algorithms based on ocular images. *Healthcare, 11*(12), 1739. [CrossRef]

[11] American College of Cardiology. (2017). Guideline

for the prevention, detection, evaluation, and management of high blood pressure in adults. *J. Amer. College Cardiol*, *71*(19), 4-28.

[12] An, Y., Huang, N., Chen, X., Wu, F., & Wang, J. (2019). High-risk prediction of cardiovascular diseases via attention-based deep neural networks. *IEEE/ACM transactions on computational biology and bioinformatics*, *18*(3), 1093-1105. [CrossRef]

[13] LaFreniere, D., Zulkernine, F., Barber, D., & Martin, K. (2016, December). Using machine learning to predict hypertension from a clinical dataset. In *2016 IEEE symposium series on computational intelligence (SSCI)* (pp. 1-7). IEEE. [CrossRef]

[14] Ye, C., Fu, T., Hao, S., Zhang, Y., Wang, O., Jin, B., ... & Ling, X. (2018). Prediction of incident hypertension within the next year: prospective study using statewide electronic health records and machine learning. *Journal of medical Internet research*, *20*(1), e22. [CrossRef]

[15] Hwang, S. H., Lee, H., Lee, J. H., Lee, M., Koyanagi, A., Smith, L., ... & Lee, J. (2024). Machine Learning–Based Prediction for Incident Hypertension Based on Regular Health Checkup Data: Derivation and Validation in 2 Independent Nationwide Cohorts in South Korea and Japan. *Journal of Medical Internet Research*, *26*, e52794. [CrossRef]

[16] Asadullah, M., Hossain, M. M., Rahaman, S., Amin, M. S., Sumy, M. S. A., Parh, M. Y. A., & Hossain, M. A. (2023). Evaluation of machine learning techniques for hypertension risk prediction based on medical data in Bangladesh. *Indonesian Journal of Electrical Engineering and Computer Science*, *31*(3), 1794-1802. [CrossRef]

[17] Binu, D. (2025). Comprehensive framework for ocular disease detection: utilizing Gegenbauer graph neural networks and fundus image data fusion techniques for enhanced classification of diverse ocular conditions. *Biomedical Signal Processing and Control*, *110*, 108275. [CrossRef]

[18] Saleh, H., Younis, E. M., Sahal, R., & Ali, A. A. (2021). Predicting systolic blood pressure in real-time using streaming data and deep learning. *Mobile Networks and Applications*, *26*(1), 326-335. [CrossRef]

[19] Liang, Y., Chen, Z., Ward, R., & Elgendi, M. (2018). Photoplethysmography and deep learning: Enhancing hypertension risk stratification. *Biosensors*, *8*(4), 101. [CrossRef]

[20] Balasubramaniam, S., Kadry, S., & Kumar, K. S. (2024). Osprey Gannet optimization enabled CNN based Transfer learning for optic disc detection and cardiovascular risk prediction using retinal fundus images. *Biomedical Signal Processing and Control*, *93*, 106177. [CrossRef]

[21] Sunil, D. M., & Joshi, P. V. (2025, March). Early Prediction of Chronic Diseases Through Multimodal Data Fusion in Machine Learning: A Survey. In *2025 7th International Conference on Intelligent Sustainable Systems (ICISS)* (pp. 1108-1115). IEEE. [CrossRef]

[22] Abbas, S., Sampedro, G. A., Krichen, M., Alamro, M. A., Mihoub, A., & Kulhanek, R. (2024). Effective hypertension detection using predictive feature engineering and deep learning. *IEEE Access*, *12*, 89055–89068. [CrossRef]

[23] Suvon, M. N., Tripathi, P. C., Alabed, S., Swift, A. J., & Lu, H. (2022, December). Multimodal learning for predicting mortality in patients with pulmonary arterial hypertension. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 2704-2710). IEEE. [CrossRef]

[24] Iao, W. C., Zhang, W., Wang, X., Wu, Y., Lin, D., & Lin, H. (2023). Deep learning algorithms for screening and diagnosis of systemic diseases based on ophthalmic manifestations: A systematic review. *Diagnostics*, *13*(5), 900. [CrossRef]

[25] Barriada, R. G., & Masip, D. (2023). An overview of deep-learning-based methods for cardiovascular risk assessment with retinal images. *Diagnostics*, *13*(1), 68. [CrossRef]

[26] Baharoon, M., Almatar, H., Alduhayan, R., Aldebasi, T., Alahmadi, B., Bokhari, Y., ... & Aljouie, A. (2024). HyMNet: A multimodal deep learning system for hypertension prediction using fundus images and cardiometabolic risk factors. *Bioengineering*, *11*(11), 1080. [CrossRef]

[27] Dai, G., Zhang, C., & He, W. (2019). Screening of diabetes and hypertension based on retinal fundus photographs using deep learning. *medRxiv*, 2019-12. [CrossRef]

[28] Lee, Y. C., Cha, J., Shim, I., Park, W. Y., Kang, S. W., Lim, D. H., & Won, H. H. (2023). Multimodal deep learning of fundus abnormalities and traditional risk factors for cardiovascular risk prediction. *npj Digital Medicine*, *6*(1), 14. [CrossRef]

[29] Al-Absi, H. R. H., Islam, M. T., Refaee, M. A., Chowdhury, M. E. H., & Alam, T. (2022). Cardiovascular disease diagnosis from DXA scan and retinal images using deep learning. *Sensors*, *22*(12), 4310. [CrossRef]

[30] Khan, K. B., Siddique, M. S., Ahmad, M., & Mazzara, M. (2020). A hybrid unsupervised approach for retinal vessel segmentation. *BioMed Research International*, *2020*(1), 8365783. [CrossRef]

[31] Huang, Y., Wittmann, B., Demler, O., Menze, B., & Davoudi, N. (2024, October). Predicting stroke through retinal graphs and multimodal self-supervised learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 223-234). Cham: Springer Nature Switzerland. [CrossRef]

[32] Khan, K. B., Khaliq, A., & Shahid, M. (2017). A novel fast GLM approach for retinal vascular segmentation and denoising. *Journal of Information Science and Engineering*, *33*(6), 1611–1627. [CrossRef]

[33] Qin, G., Xu, N., & Xu, J. (2025, August). Multimodal Deep Learning for Diabetic Retinopathy: A Survey. In *2025 IEEE 8th International Conference on Multimedia Information Processing and Retrieval (MIPR)* (pp. 151-157). IEEE. [CrossRef]

[34] Hao, R., Xiang, Y., Du, J., He, Q., Hu, J., & Xu, T. (2025). A Hybrid CNN-Transformer Model for Heart Disease Prediction Using Life History Data. *arXiv preprint arXiv:2503.02124*.

[35] Zhang, L., Yuan, M., An, Z., Zhao, X., Wu, H., Li, H., Wang, Y., Sun, D., Liu, X., & Chen, Y. (2020). Prediction of hypertension, hyperglycemia and dyslipidemia from retinal fundus photographs via deep learning: A cross-sectional study of chronic diseases in central China. *PLOS ONE, 15*(5), e0233166. [CrossRef]

[36] Abdollahi, M., Jafarizadeh, A., Ghafouri-Asbagh, A., Sobhi, N., Pourmoghtader, K., Pedrammehr, S., ... & Acharya, U. R. (2024). Artificial intelligence in assessing cardiovascular diseases and risk factors via retinal fundus images: A review of the last decade. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 14*(6), e1560. [CrossRef]

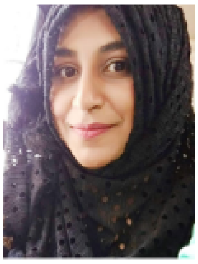
Wajiha Farooq is a lecturer at COMSATS University Islamabad, Sahiwal Campus, in the Department of Computer Science. Her research interests include underwater wireless sensor networks (UWSNs), machine learning, deep learning, and edge computing. She is currently exploring AI-based solutions for real-world sensing and classification problems. (Email: wajihafarooq980@gmail.com)

Aamir Ali is a dedicated researcher in artificial intelligence, healthcare analytics, and IoT environments. His expertise includes machine and deep learning applications in medical diagnosis, speech recognition, and facial emotion recognition. He has also worked on IoT-focused projects such as malware classification, anomaly detection, and cyberattack prediction. Aamir is passionate about developing AI-driven solutions for early disease detection and improving patient care. He actively contributes to interdisciplinary research across clinical and image data domains. (Email: amirali4436823@gmail.com)

Hafza Mehreen Fatima is a Lecturer in the Department of Computer Science at COMSATS University Islamabad, Sahiwal Campus. Her academic interests include data mining, software metrics, agile methodologies, software testing, and big data analytics. Alongside her teaching responsibilities, she actively engages in academic research, focusing on data-driven approaches to improve software quality and development processes. (Email: mehrenfatima136@gmail.com)

Wasif Rafiq is a dedicated researcher specializing in AI, healthcare analytics, and IoT-based systems. His work focuses on leveraging machine learning and deep learning for medical diagnostics and cybersecurity. Passionate about innovation, he aims to develop intelligent tools for early disease detection and improved healthcare delivery. (Email: malikwasif044@gmail.com)

Noor-e-Zainab is a researcher focused on AI, healthcare analytics, and IoT, aiming to develop smart tools for early disease detection and better healthcare through machine learning and deep learning. Her work also addresses cybersecurity challenges in medical systems. She is committed to advancing innovative, data-driven solutions for real-world clinical applications. (Email: nzainab760@gmail.com)



Misbah Ali is a PhD scholar at COMSATS university Islamabad, with research interests in Machine Learning, Deep Learning, Generative Artificial Intelligence, and Software Engineering. Her work focuses on the development of secure, intelligent systems across domains such as healthcare, education, and industrial cybersecurity. She has authored multiple peer-reviewed publications and presented her research at international conferences. She also contributes to the academic community as a reviewer for several reputed journals. (Email: talktomisbah.ali@gmail.com)