Digital Health Trends Journal. 2025;1(1):3-17.
doi:10.34172/dhtj.02
https://dhtjournal.ae

Digital Health Trends Journal

Systematic Review

Application of Artificial Intelligence in Hypertension Diagnosis: A Systematic Review

Ali Valizadeh[1], Zahra Hosseinzadeh[2*], Elnaz Jalilian[3]

[1]School of Medicine, Bushehr University of Medical Sciences, Bushehr, Iran
[2]Department of Health Information Technology, Faculty of Paramedical Sciences, Lorestan University of Medical Sciences, Khorramabad, Iran
[3]Postdoctoral Researcher, Universite Grenoble Alpes, Grenoble, France

***Corresponding author:** Zahra Hosseinzadeh, Email: zahraahoseinzadeh2001@gmail.com

Abstract
High blood pressure (BP) is a major health concern that can lead to various cardiovascular diseases, serious complications, and even death. Although much progress has been made in the diagnosis and treatment of high BP, awareness of this medical condition and its effective control are not fully achieved. Nonetheless, artificial intelligence (AI) has created new opportunities for better detection and management of high BP. Therefore, this review investigated AI-related studies (2015-2025) performed for BP diagnosis. Overall, 48 studies using various AI methods (e.g., machine learning and deep learning) and covering different AI applications (i.e., non-invasive BP monitoring, early detection and classification of different types of high BP, and even disease risk prediction) were reviewed. Most AI models had very good accuracy (71–97%), and some even outperformed human experts. However, there were still problems, such as small sample sizes, lack of external validation, high data variability, and difficulties in understanding and interpreting complex models. Our findings revealed that AI can diagnose high BP more accurately, help people receive treatment more quickly, and even personalize their care. Nevertheless, some issues should be figured out before using AI. Researchers need to understand how these AI models act in actual hospitals and clinics, how they make their decisions, and how easily they can be integrated into the systems that doctors and nurses currently use. Otherwise, it is difficult for individuals to really trust this technology.
Keywords: Artificial intelligence, Hypertension, Diagnosis

Received: July 23, 2025, **Revised:** August 28, 2025, **Accepted:** September 10, 2025, **ePublished:** September 18, 2025

Background

High blood pressure (BP) is considered one of the main health concerns that affects an estimated 1.28 billion people worldwide. In fact, it is recognized as a major risk factor for cardiovascular diseases, serious complications, and even preventable mortality.[1,2]

Considerable progress has been made in the diagnosis and management of high BP, although much work remains to be performed in terms of the awareness, treatment, and control of this disease. Interestingly, only about 14% of patients can manage to keep their systolic BP below 140 mmHg worldwide.[3,4] High BP is extremely difficult to control due to the complexity of the disease itself. Different factors (e.g., genes, the role of environment in gene alteration, residence, financial resources, and daily activities) contribute to high BP.[5-7]

In recent years, artificial intelligence (AI) has entered the medical world as a transformative technology, opening up new horizons for specialists. More precisely, doctors can now view and address complex clinical challenges (e.g., managing high BP) in innovative ways.[8,9] AI, which includes machine learning (ML) and deep learning (DL), has shown promise in pattern recognition, risk assessment, and outcome prediction. This is particularly promising when applied to large amounts of health data, such as information from wearable devices, electronic health records (EHRs), or omics data.[9-11]

In addition, AI is extensively applied in the field of hypertension (HTN), including noninvasive measurement, early diagnosis of various types of the disease, risk prediction, and individualized treatment design. In fact, this technology is laying the foundation for precision and personalized medicine more firmly than ever before.[8,12]

Moreover, the combination of AI with mobile technologies, wearables, and telemedicine enables continuous and remote monitoring and control of BP. This is especially valuable for chronic diseases (e.g., high BP) and can be of great help to patients and doctors.[4,13]

Despite all these advances, AI tools have yet to enter clinical practice in various places and remain largely in the experimental stage. However, there are challenges

in this regard. For example, it is difficult to precisely understand the algorithms. In addition, the models should be tested in real-world conditions, and it should be ensured that their results apply to different populations.[8] Eventually, researchers have been diving into how AI can help diagnose and manage HTN. Gudigar et al found that most studies stuck to just one type of data. Although methods such as electrocardiograms (ECG), photoplethysmography (PPG), or medical images kept coming up, there was not much beyond that. Only a few studies could combine multidimensional data, including bringing together clinical information, signals, and imaging to achieve a more complete view.[14]

Despite the scientific value of this review, there are important aspects that justify the present study. First, the present study exclusively focuses on the diagnosis of HTN. Gudigar et al investigated the secondary effects of HTN (e.g., kidney and heart issues). However, our study concentrates on AI-based methods for actually diagnosing HTN. Nonetheless, there is a large research gap here. The mentioned researchers have indicated that people have not really dug into using multimodal and explainable AI for diagnosing high BP. This systematic review seeks to shine a light on that gap and offer some ideas to help shape the direction of future research.

Further studies are required to obtain data about AI and high BP diagnosis. Accordingly, this study aims to review articles published between 2015 and 2025, focusing on studies that evaluated patients with high BP and how AI could help diagnose this medical condition.

Methods

Search Strategy

This systematic review was conducted according to the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses. Several databases, including PubMed, Scopus, and Web of Science, were searched for studies published between January 2010 and August 2025. To this end, specific keywords were used, such as "Artificial intelligence" OR "machine learning" OR "deep learning" AND "hypertension" OR "high blood pressure" AND "diagnosis" OR "detection" OR "screening".

Inclusion and Exclusion Criteria

Inclusion Criteria

- Studies that used AI methods, such as ML, DL, or neural networks (NNs), to manage high BP
- Articles published in English
- Original peer-reviewed research articles, including clinical trials, observational studies, and retrospective analyses
- Studies that primarily focused on the use of AI for HTN diagnosis

Exclusion Criteria

- Non-English publications
- Review articles, editorials, commentaries, conferences, and letters to the editor
- Studies not specifically focusing on HTN
- AI applications in unrelated medical conditions

Data Extraction

Two independent reviewers reviewed the titles and abstracts. Then, full texts of potentially eligible articles were assessed for inclusion. Disagreements were resolved by a third reviewer. Different data were extracted for each study, including (1) authors and year of publication, (2) study type and population, (3) AI technique(s) used, (4) purpose of AI application (e.g., diagnosis or monitoring), (5) key outcomes and performance measures (e.g., accuracy, sensitivity, and specificity), and (6) limitations and challenges.

Quality Assessment

The methodological quality and risk of bias of the included studies were assessed using the QUADAS-2 tool (Quality Assessment of Diagnostic Accuracy Studies-2). This instrument evaluates four key domains: patient selection, index test, reference standard, and flow and timing. The findings of the quality assessment are qualitatively reported in the Discussion section.

Data Analysis

The data were qualitatively synthesized, and where possible, a comparative analysis was performed based on AI techniques and application domains.

Results

Study Selection

A total of 1,628 records were identified through database searches. After discarding duplicates (n = 485), records that were removed by automation tools (n = 36), and other irrelevant records (n = 289), 818 articles remained, which were screened based on their titles and abstracts. Of this number, 477 were excluded since they did not match the inclusion criteria. The full text of 341 articles was reviewed, and 118 of them were excluded due to irrelevance. Another 175 articles were also removed because they did not include the intended data. Eventually, 48 studies were included in this systematic review (Figure 1).

Overview of Included Studies

Table 1 summarizes the findings of previous studies conducted on the use of AI in hypertension diagnosis.

The selected studies spanned various AI approaches applied to the diagnosis of HTN and related conditions, including essential hypertension, pulmonary hypertension (PH), intracranial HTN, and secondary HTN. In addition, the study types included a combination of retrospective and observational designs, along with secondary analyses of clinical trials, cross-sectional studies, and development or evaluation studies. Moreover, the studies examined different numbers of people, from small groups of less than 100 to large sets of real-world data with over 20,000 people.

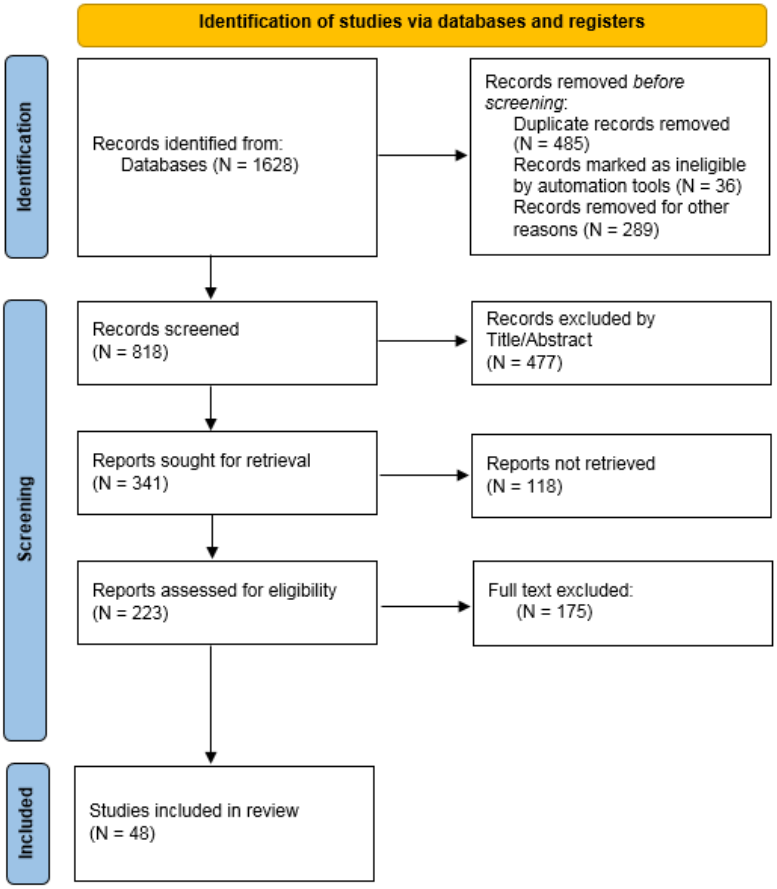

Figure 1. PRISMA Flow Diagram of Study Selection Process. *Note*. PRISMA: Preferred Reporting Framework for Systematic Reviews and Meta-Analyses

For Non-Invasive Hypertension Diagnosis

Overall, 12 studies were found, focusing on non-invasive ways to diagnose HTN (e.g., ECG, PPG, or wearables).[27,31-34,39,41,46,48,49,52,56] Most of them reported pretty impressive accuracy, anywhere from 71% up to 99%. Nonetheless, a closer look revealed that four of these studies had some issues. Either their sample sizes were tiny, or they did not explain their design and timing clearly, making it harder to trust that their results would hold up in the real world.

Pulmonary hypertension: Fifteen studies addressed PH detection using echocardiography, chest X-rays, cardiac magnetic resonance imaging, or phonocardiograms.[15,19,21,23,26-30,34,35,59] Overall, these studies showed very good diagnostic performance (with an area under the curve [AUC] of up to 0.97), and most of them were at low risk of bias as assessed by QUADAS-2. However, 3 studies had limitations in patient selection or lacked external validation, leading to an "unclear" risk in certain domains.

Secondary hypertension: Six of these studies used AI to differentiate between different types of high BP (e.g., to identify secondary HTN or subtypes such as primary aldosteronism and kidney problems).[17,29,38,42,43,58] Most of them were retrospective in nature. Although methodological quality was generally acceptable, 2 studies demonstrated concerns regarding incomplete data and generalizability.

Intracranial hypertension: Two studies investigated intracranial HTN using waveform analysis and PPG-based models.[16,63] Both studies reported promising diagnostic accuracy (>90%), but their small sample sizes limited the certainty of evidence. The QUADAS-2 assessment rated them as low risk in the index test and reference standard but unclear in selecting patients.

Population-level screening and risk prediction: Fifteen studies examined the prediction of HTN risk from large survey or population-based data.[20,36,37,40,41,44-47,51,53-57,61] They represented moderate-to-high diagnostic accuracy, although many of them were considered to have limitations in either patient selection or flow/timing due to reliance on self-reported or single-institution data.

Artificial Intelligence Techniques and Applications

The investigated studies used different AI setups. RFs, support vector machines (SVMs), DTs, and gradient boosting (old-school ML) were utilized in at least 18 studies.[15,19,20,25,29,31,33-36,39-42,44-46] Moreover, DL methods (e.g., convolutional neural networks, residual network, and InceptionV3) and semi-supervised models were employed in 20 studies.[16-18,21,23,26-28,30,32,35,48-56,58,59,61] Additionally, 12 studies included clinical information,

Table 1. Summary of Studies on the Application of AI in Hypertension Diagnosis

Authors/Year	Study Type and Population	AI technique(s) Used in Studies	Purpose of the AI Application	Key Findings	Limitations and Challenges
Priya et al (2021)[15]	A retrospective study of 72 patients, including 42 with pulmonary hypertension (PH) and 30 healthy controls	Radiomics with texture feature extraction from cardiac magnetic resonance imaging (MRI) combined with various machine learning (ML) classifiers (including multilayer perceptron)	Non-invasive diagnosis of PH through cardiac MRI texture analysis	The best model (multilayer perceptron) achieved an AUC of 0.862 (78% ACC) in the primary analysis and an AUC of 0.918 (80% ACC) in the subgroup with preserved left ventricular function.	- Small sample size - Variability in model performance depending on classifier and feature selection methods
Quachtran (2016)[16]	A retrospective study on 60 patients monitored for intracranial pressure over 30-minute intervals	Deep learning (DL) applied to waveform morphology analysis	Diagnosis/detection of intracranial HTN by predicting elevated intracranial pressure	The DL model achieved approximately 92% ACC in detecting intracranial HTN.	- Small sample size - Need for further validation on larger and diverse populations
Wu et al (2013)[17]	A cross-sectional study using real-world electronic health record (EHR) data from11,961 hypertensive patients(2013-2019)	A two-stage DL framework withnatural language processing,label embedding, andattention mechanisms	Provision of supportfor differential diagnosisofsecondary HTN by integrating textual (e.g., symptoms) and numerical (e.g., lab results) data	The model accurately identified common types of secondary HTN by combining text and numerical data from patient records. It outperformed doctors and standard methods. However, it needs further testing in real clinical settings.	- Lack of real-world EHR data or incomplete entries - Lack of testing generalizability to other hospitals or healthcare systems - Need for further validation inprospective clinical settings
Yoshihara et al (2024)[18]	A secondary analysis of a clinical trial involving 7,710 patients with influenza-like symptoms from 64 primary care clinics in Japan	DL with multi-instance convolutional neural networks (CNNs)	Diagnosisof HTN from pharyngeal images	The results showed that DL can accurately detect HTN from throat images, thereby performing better than traditional methods and working well across all age and gender groups.	- Reliance on data from patients with influenza-like symptoms - Lack of general population - Need for further validation in broader clinical settings and among diverse populations
Zhu et al (2020)[19]	An observational study on 275 patients with PH who underwent both echocardiography and right heart catheterization	Nine ML models, including LogitBoost and decision tree (DT)	Classification of PH as pre-capillary or post-capillary using echocardiographic data	The results demonstrated that this method can accurately distinguish between types of PH without the need for invasive procedures, such as cardiac catheterization. However, further studies are needed to validate its effectiveness across different patient populations.	- Lack of external validation - Limited to a specific patient population - Need for further comparison with other noninvasive diagnostic methods
Montagna et al (2022)[20]	An observational study analyzing questionnaire data from 20,206 individuals collected during World HTN Day (2015-2019)	Five ML algorithms were tested, including random forest (RF) with different data balancing techniques	Improvement of HTN risk detection and prediction in population screening	RF revealed balanced sensitivity (SE: 0.818) and specificity (SP: 0.629), better SP than current medical protocols but lower SE; no single best algorithm was found.	- Trade-off between SE and SP - Need for more accurate data and additional features to improve model performance - Difficulty in population-level detection
Zou et al (2020)[21]	A retrospective study on 762 patients (405 PH and 357 controls)	DL (ResNet50, Xception, and Inception V3)	Detection of PH from chest X-rays and prediction of pulmonary artery systolic pressure (PASP)	The best model (Inception V3) achieved high ACC in classification (AUC: 0.970 internal, 0.967 external) and good PASP prediction (MAE: 7.45 internal, 9.95 external); DL outperformed expert radiologist manual classification.	- Need for prospective validation - A slight drop in external test ACC
Li (2021)[22]	A review study including no specific population, as it summarized existing research on portal HTN diagnosis	Various AI techniques (not specified individually) applied in imaging and data analysis for noninvasive diagnosis	Noninvasive diagnosis and risk monitoring of portal HTN and gastroesophageal varices, aiming to personalize patient care	AI shows promise in transforming clinical practice by enabling accurate, noninvasive diagnosis and monitoring, potentially reducing the need for invasive, costly procedures.	- Challenges related to integrating AI in clinical workflows - Need for using large datasets and validation and ensuring personalized, timely care
Kusunose et al (2022)[23]	An observational study with 142 patients having scleroderma or mixed connective tissue disease who underwent 6-minute walk stress echocardiography	A DL model applied to chest X-ray images to predict PH probability	Detection of exercise-induced PH (EIPH) non-invasively using chest X-rays	The DL model probability of PH was significantly higher in patients with EIPH. Adding DL-predicted PH probability improved the predictive ACC for EIPH (AUC increased from 0.65 to 0.74). Gender, resting mean pulmonary artery pressure, and DL model output were independent predictors of EIPH.	- Lack of transparency - Implied challenges, including limited sample size - Need for further validation in broader clinical settings

Table 1. Continued.

Authors/Year	Study Type and Population	AI technique(s) Used in Studies	Purpose of the AI Application	Key Findings	Limitations and Challenges
Juyal et al (2024)[24]	A review article analyzing the role of AI and wearable technology in HTN management; no specific population was studied.	ML algorithms analyzing physiological, lifestyle, and genetic data; adaptive algorithms for personalized treatment optimization	Early and accurate diagnosis of HTN, personalized treatment planning, and real-time blood pressure (BP) monitoring	AI models can detect subtle risk patterns for HTN and optimize treatments based on individual responses, and wearables enable continuous health monitoring for timely intervention.	- Lack of specifying detailed limitations - Implied challenges, including data diversity, integration of multi-source data, and clinical validation of AI models
Jusic et al (2023)[25]	An observational case-control study including 174 participants (89 with essential HTN and 85 controls)	ML – support vector machine (SVM)	Diagnosis of essential HTN by integrating circulating micro ribonucleic acids (miRNAs) and clinical risk factors	The study found several miRNAs that were lower in people with high BP. Using these miRNAs along with clinical risk factors (e.g., gender and smoking), the ML model (SVM) could accurately identify patients with essential HTN. The model displayed high ACC (90%) and balanced SE and SP, implying that it was beneficial at both detecting HTN and ruling it out. However, it still needs to be tested on more patients before it can be used in clinics.	- Need for further validation on independent patient cohorts before clinical use
Han et al (2023)[26]	A retrospective study using 3,255 preoperative chest radiographs, including 1,174 cases with congenital heart disease (CHD) and 2,081 cases without CHD	ResNet18 (pretrained on ImageNet) compared with other models (DenseNet121, MobileNetv2, and MobileViT)	Automatic diagnosis of CHD and PAH associated with CHD (PAH-CHD) from chest X-rays; evaluation if AI can improve radiologists' ACC	The AI accurately detected CHD and reasonably identified PAH. It performed better than radiologists. With AI assistance, radiologists improved their diagnostic ACC.	- Less ACC of PAH-CHD diagnosis compared to CHD - Use of a retrospective design - Need for further validation in clinical settings
Guo et al (2024)[27]	A development and evaluation study using ~6000 labeled and ~169,000 unlabeled phonocardiogram (PCG) recordings; test set included 196 patients	Deep CNNs trained in a semi-supervised manner; GradCAM^{++}for interpretability	Screening and early detection of PH by identifying elevated PASP (≥40 mmHg) using digital stethoscope data	The model could detect PH with 71% SE and 73% SP It performed better at specific chest locations.	- Moderate diagnostic performance (AUC<0.8) Dependent on the quality of PCG recordings - Need for further validation in diverse populations
Diller et al (2022)[28]	An observational study, including 450 pulmonary arterial HTN (PAH) patients, 308 with right ventricular dilation without PAH, 67 healthy controls	DL CNNs using echocardiographic images and estimated right ventricular systolic pressure	Detection of PAH accurately Prediction of patient prognosis (mortality risk)	The DL model detected PAH with 97.6% ACC and predicted patient risk and expert doctors. It also improved diagnosis and prognosis using standard heart ultrasound.	- Lack of transparency in the article - Need for validation in broader and more diverse populations
Buffolo et al (2021)[29]	A retrospective study on 4,059 patients with HTN	ML algorithms	Accurate prediction of primary aldosteronism (PA) and its surgically treatable form (unilateral PA)	The ML model accurately predicted PA and its surgically treatable form with high ACC. It could identify about one-third of patients as low risk, allowing them to safely skip further screening.	- Lack of transparency in the article - Need for further validation in diverse populations
Aras et al (2022)[30]	A retrospective study of 24,470 adults who had electrocardiogram (ECG) and either right heart catheterization or echocardiogram within 90 days	Deep CNNs	Detection of PH and its subtypes using 12-lead ECG data	A DL model could accurately detect PH using only ECG data.	- Need for 12-lead ECG data - Need for further validation in diverse clinical settings
Angelaki et al (2022)[31]	An observational study with 1,091 individuals without cardiovascular disease (CVD), classified as hypertensive or normotensive	The RF ML model using clinical and ECG-derived features	Opportunistic detection of HTN using ECG and clinical data	The ML model effectively detected HTN using ECG and basic body measurements. It could help identify many people with undiagnosed high BP, reducing cardiovascular risk.	- Inclusion of a limited population (patients without CVD) - Need for further validation
Adeleke et al (2024)[32]	A retrospective analysis of 1,723 staff members at Bowen University, Nigeria, between 2018 and 2022	K-means clustering for analyzing BP data	Evaluation of the impact of a workplace HTN screening program using ML	Using ML to check employee health over time was a useful and practical way to find and manage high BP at work.	- Limitation of data (from a single institution) - Unavailability of data for 2020 due to the COVID-19 lockdown - Lack of assessing cost-effectiveness and practical challenges of implementation

Table 1. Continued.

Authors/Year	Study Type and Population	AI technique(s) Used in Studies	Purpose of the AI Application	Key Findings	Limitations and Challenges
Angelaki et al (2025)[33]	An observational prospective study with 1,254 subjects (average age ~60 years), both hypertensive and normotensive, with no CVD	The RF ML model on single-lead ECG data	Detection of arterial HTN using single-lead ECGs as a proof of concept for wearable devices	The findings revealed that even simple ECG signals (e.g., those from a smartwatch) could help identify hidden HTN, supporting early detection and awareness.	- Need for validation with data from actual wearable devices (e.g., smartwatches)
Martinez-Ríos et al (2022)[34]	A study analyzing photoplethysmography (PPG) signals and clinical data for HTN detection (population details not specified)	Wavelet scattering transform (WST) for feature extraction from PPG signals combined with SVM classifiers; Early and late fusion methods for combining data types	Early detection of HTN using non-invasive PPG signals and clinical data	Using WST-extracted PPG features with SVM achieved 71.42% ACC and 76% F1-score for classifying normotension vs. pre-HTN; combining PPG and clinical data via fusion did not improve performance.	- Need for high-quality PPG signals - Lack of detection ACC enhancement with the fusion of data types - Avoidance of DL methods due to complexity and lack of interpretability
Anand et al (2024)[35]	A retrospective analysis of 7,853 patients with heart catheterization and echocardiography data	Gradient boosting ML model	Prediction of PH using echocardiographic data without relying on tricuspid regurgitation velocity	The model accurately identified most patients with PH, correctly detecting 88% of true cases, but was less effective at ruling out those without the condition.	- Moderate SP and NPV - Exclusion of some direct heart pressure measures - Need for further validation
Golino et al (2014)[36]	An observational cross-sectional study including 400 undergraduate students (56.3% women), age 16–63, Brazil	Classification tree (an ML algorithm in R, 15 random trees for each gender)	Prediction of increased BP using anthropometric measures (BMI, WC, HC, and WHR)	Women: Best model (BMI+WC+WHR) SE 80.86%, SP 81.22% in training; dropped to 45.65% and 65.15% in testing Men: Best model (BMI+WC+HC+WHR) SE 72%, SP 86.25% in training; dropped to 58.38% and 69.70% in testing Classification trees outperformed LR.	- Small, non-representative convenience sample (university students) - Imbalanced dataset (few women with HTN) - A decline in predictive performance in the test set, indicating limited generalizability - Cross-sectional design - Lack of validation in broader populations
Seffens et al (2015)[37]	A multicohort clinical study using the Minority Health Genomics and Translational Research Repository Database; self-reported African American participants and related cohorts	NNs (for missing-data imputation) and data mining classification tools (association rule generation)	Imputation of missing phenotype data and classification of HTN case/control status in a large genomic/phenotypic database	NNs improved dataset completeness and increased power for detecting associations between phenotype variables and HTN status. Data mining produced association rules, highlighting links between clinical/phenotypic factors and HTN.	- A focus on data imputation rather than direct clinical diagnosis - Limited external validation - Lack of generalizability beyond African American cohorts - Modest genomic associations despite improved imputation
Ren et al (2019)[38]	A retrospective analysis using electronic health records of 35,332 HTN patients	Hybrid NN (BiLSTM +autoencoder)	Prediction of kidney disease in patients with HTN by integrating textual and numerical EHR data	It achieved 89.7% ACC, outperforming traditional statistical models and baseline neural systems.	- A limited focus on kidney disease as a complication of HTN - Lack of data generalizability beyond the applied dataset - Need for external validation in diverse populations
Nour and Polat (2020)[39]	An observational study; an HTN dataset with 8 personal features (gender, age, height, weight, SBP, DBP, HR, and BMI); 4 classes (normal, pre-HTN, stage-1, and stage-2 HTN)	C4.5 DT, RF, linear discriminant analysis (LDA), and linear SVM	Automatic classification of HTN types using personal demographic and physiological features	It achieved 99.5% ACC with DT and RF, 96.3% with LDA, and 92.7% with SVM; it demonstrated the feasibility of ML in automatic HTN classification.	- Lack of full description of dataset characteristics (size and source) - Possible overfitting given very high ACC - Lack of external/clinical validation - Limited generalizability beyond the dataset
López-Martínez et al (2020)[40]	A retrospective analysis of NHANES data (2007-2016); 24,434 participants (69.7% non-hypertensive and 30.3% hypertensive)	An artificial NN (ANN) classification model	Prediction of HTN based on demographic, lifestyle, and comorbidity factors (gender, race, BMI, age, smoking, kidney disease, and diabetes)	It achieved an AUC of 0.77 (95% CI 0.75–0.79), an SE of 40%, an SP of 87%, and a precision of 57.8%. It outperformed the prior statistical model (AUC of 0.73).	- Imbalanced dataset (more non-hypertensives) - Relatively low sensitivity - Need for external/clinical validation before clinical adoption

Table 1. Continued.

Authors/Year	Study Type and Population	AI technique(s) Used in Studies	Purpose of the AI Application	Key Findings	Limitations and Challenges
Islam et al (2022)[41]	A retrospective, population-level study using harmonized Demographic and Health Surveys data from Bangladesh, Nepal, and India (N = 818,603; 82,748 hypertensive, 10.1%)	DT, RF, gradient boosting machine, XGBoost, logistic regression (LR), and LDA	Prediction of HTN and identification of key associated factors in South Asian populations	XGBoost, GBM, LR, and LDA achieved 90% ACC, 100% recall, and 95% F1-score; RF achieved 89% ACC, and 83% DT. Age and BMI were the strongest predictors.	- Reliance of models mainly on survey/self-reported data - Lack of the inclusion of biochemical markers - Uncertainty in generalizability outside South Asia - Need for real-world validation
Shih et al (2022)[42]	Two retrospective cohorts in Taiwan (970 patients for model development and internal validation and 464 patients for external validation)	LR, RF, XGBoost, and ANNs	Prediction of white-coat HTN and white-coat uncontrolled HTN from outpatient visit data	RF performed best (AUROC = 0.884, SE = 0.619, SP = 0.887, ACC = 0.819, NPV = 0.872). Top predictors: office DBP, office SBP, smoking status, estimated glomerular filtration rate, and fasting glucose	- Need for conducting the study on other populations than Taiwanese cohorts - Limited external validation to one cohort - Need for prospective validation for generalizability
Orozco Torres et al (2022)[43]	A population-based study	Multilayer feed-forward NNs with backpropagation	Diagnosis of HTN using physiological and lifestyle factors (e.g., obesity, diabetes, smoking, stress, physical inactivity, and the like)	It achieved ~90% effectiveness in diagnosing HTN risk from physiological data.	- Lack of detailed information on sample size – Need for external validation - Need for testing on more diverse populations and clinical settings
Fang et al (2023)[44]	An observational study; population details not fully specified in the abstract (Chinese cohort)	Hybrid ML combining K-nearest neighbor (KNN) and LightGBM	Prediction of individual risk of developing HTN within the next five years based on demographic and blood indicators	It achieved ACC > 86% and recall > 92%, demonstrating reliable risk prediction.	- Lack of full description of population and dataset details - Need for external validation - Uncertainty in applicability to non-Chinese populations
Kurniawan et al (2024)[45]	An observational study among Indonesian adults (population details in full text)	DT, RF, gradient boosting, and LR	Development and validation of an HTN risk prediction model using modifiable risk factors	LR achieved the best performance: AUC of 0.829, ACC of 89.6%, recall of 0.896, precision of 0.878, and F1 score of 0.877.	- Need for external validation - Limited generalizability outside the Indonesian population - Lack of fully specified dataset details in the abstract
Sakka et al (2020)[46]	A case study at Petra University, Jordan, with a dataset of 31,500 patients (12,658 hypertensive and 18,842 non-hypertensive)	Synthetic minority oversampling technique (SMOTE) with KNN, DT, and other ML classifiers	Prediction of HTN risk using noninvasive, inexpensive patient data from university health records	SMOTE-KNN achieved 83.9% ACC, 85.1% SP, 83.3% SE, and 89.6% AUC with 10-fold cross-validation (CV).	- Reliance on a single-institution dataset (Petra University) - Uncertainty in generalizability of results to broader populations - Lack of external validation
Islam et al (2023)[47]	A cross-sectional study, Ethiopia; 612 respondents with 27 risk factors	LR, ANN, RF, and XGBoost (Boruta feature selection and SHAP for explainability)	Prediction of patients at risk of developing HTN and identification of key risk factors	XGBoost achieved the best performance: ACC of 88.81%, precision of 89.62%, recall of 97.04%, F1-score of 93.18%, and AUC of 0.894; SHAP analysis highlighted age, BMI, diabetes, salt intake, alcohol, smoking, and history of HTN as key predictors.	- Small sample size (n = 612) - Lack of generalizability of findings to populations other than the Ethiopian population - Need for external validation in larger, diverse cohorts for generalizability
El-Dahshan et al (2023)[48]	A development and validation study; evaluation of two public PPG datasets (PPG-BP and MIMIC-II)	ExHyptNet (EfficientNetB3 for feature extraction, Grad-CAM for explainability, and XGBoost and extremely randomized trees for classification)	Development of an explainable AI model for automatic detection and classification of HTN using PPG signals	The model achieved 100% detection ACC across multiple validation methods (holdout, 10-fold CV, and leave-one-out); it provided interpretable outputs with Grad-CAM.	- Limited evaluation (only public datasets rather than prospective clinical data) - Extremely high performance, raising concern for overfitting - Need for further testing regarding generalizability to real-world settings
Esmaelpoor et al (2020)[49]	A development and evaluation study; PPG data from 200 subjects	Multistage deep NN (CNNs for feature extraction and LSTM for temporal dependencies)	Estimation of SBP and DBP from PPG signals	It met the AAMI standard and achieved Grade A (BHS standard) for both SBP and DBP estimation; it had consistent and accurate predictions.	- Limited dataset evaluation (200 subjects) - Uncertainty in generalizability to larger/diverse populations -Need for validation in real-world/clinical settings
Chen et al (2022)[50]	A development and evaluation study; dataset: MIMIC-II, 1,562 subjects (normal, hypertensive, and hypotensive)	DL: Receptive field parallel attention shrinkage network with the BP range constraint	Cuff-less and continuous estimation of SBP and DBP from PPG signals	Achieved MAE of 2.26/2.15 mmHg (SBP) and 1.63/1.59 mmHg (DBP); it outperformed state-of-the-art methods and was robust against fluctuations.	- Limitation in testing (only on the MIMIC-II dataset) - Lack of validation regarding real-world generalizability and deployment in clinical monitoring systems

Table 1. Continued.

Authors/Year	Study Type and Population	AI technique(s) Used in Studies	Purpose of the AI Application	Key Findings	Limitations and Challenges
Evdochim et al (2022)[51]	A development study; 359 PPG recordings from open-source databases	ML models	HTN detection by analyzing PPG waveform morphology (ANC Test™)	It had an ACC of systolic HTN detection of ~70%, and a maximum of 72.9%.	- Variability in PPG acquisition standards across databases - Relatively small and heterogeneous dataset - Limited generalizability
Soh et al (2020)[52]	A development and evaluation study; ECG data from the MIT-BIH normal sinus rhythm and SHAREE database	Empirical mode decomposition for feature extraction + nonlinear features + k-NN classifier	Detection of HTN and masked HTN using ECG signals	It achieved 97.7% ACC with 10-fold CV.	- Evaluation limited to two datasets - Lack of external clinical validation - Probable variability with real-world ECG acquisition
Rajput et al (2019)[53]	A development study; 139 subjects (SHAREE database, Physionet); ECG signals segmented into 5-minute epochs	Optimal Orthogonal Wavelet Filter Bank + Student's t-test feature selection → Hypertension diagnosis index (HDI)	Automated detection of low-risk vs. high-risk HTN using ECG signals	HDI using LOGE + SFD features effectively discriminated LRHT vs. HRHT; it had encouraging performance for ICU use.	- Small dataset (139 subjects) - Limited population (the HRHT group included patients with comorbid myocardial infarction, stroke, and syncope, probably confounding pure HTN detection) - Need for external validation on larger datasets
Rajput et al (2022)[54]	An experimental study using ballistocardiography (BCG) signals	Continuous wavelet transforms for feature extraction + 2D CNNs	Automated detection of HTN from BCG signals	It achieved 86.14% classification ACC using 10-fold CV.	- Limited dataset/sample details in abstract - Need for external/clinical validation - Lack of testing real-world applicability because of the first use of 2D-CNN with BCG
Huang et al (2021)[55]	A retrospective analysis using PPG and ECG signals from the MIMIC-II dataset	MLP-mixer–based deep NNs (MLP-BP, gMLP-BP, and MLPlstm-BP); Multi-filter multi-channel preprocessing	Estimation of BP (cuff-less) for HTN monitoring and diagnosis	MLPlstm-BP achieved an MAE of 3.52 mmHg (SBP) and 2.13 mmHg (DBP); SD of 5.10 mmHg (SBP), 3.07 mmHg (DBP); it met the highest standards of AAMI and BHS; moreover, it outperformed state-of-the-art models.	- Testing only on the MIMIC-II dataset - Lack of validation of generalizability to diverse real-world populations and wearable devices - Need for prospective clinical trials
Miao et al (2020)[56]	A retrospective analysis using the public MIMIC waveform database (ICU patients) and an independent dataset of arrhythmia patients	A DL model combining ResNet and LSTM	Continuous, cuff-less BP estimation using one-channel ECG signals for HTN detection and monitoring	It achieved an estimation error of 0.07 ± 7.77 mmHg (MAP) and 0.01 ± 6.29 mmHg (DBP); compliant with AAMI standards; Grade A (MAP, DBP) and Grade B (SBP) per BHS standards; validated on independent arrhythmia dataset with similarly low errors.	- Focus on ICU/arrhythmia populations - Need for real-world wearable device validation - Less ACC of SBP estimation (Grade B)
Kublanov et al (2017)[57]	An observational study; 70 participants (30 healthy and 40 with stage II–III HTN); short-term HRV signals	Linear and quadratic discriminant analysis, k-NN, SVM (radial basis function), DT, and Naïve Bayes	Diagnosis of arterial HTN using HRV-derived features	Discriminant analysis achieved the highest ACC; feature selection with non-correlated sets performed better than PCA-based features.	- Small sample size - Limitation of the model to HRV signals - Need for external validation in larger and more diverse populations
Kandil et al (2019)[58]	An observational study; 66 subjects (balanced dataset: normal vs. hypertensive); brain MRA scans	Three-dimensional CNNs for cerebrovascular segmentation + ANN classifiers	Early detection of HTN through analysis of cerebrovascular alterations (diameter and tortuosity)	It achieved 90.9% ACC in distinguishing hypertensive vs. normal subjects; it demonstrated feasibility of predicting HTN before onset.	- Small sample size - Reliance on MRA imaging (costly and not widely available for screening) - Need for external validation for broader clinical utility
Gudigar et al (2019)[59]	An observational study using echocardiography images	Globally weighted local binary pattern (LBP) + entropy features with the SVM classifier	Development of a CAD tool for automated identification of PH from cardiac ultrasound images	It achieved ~92% classification ACC; the proposed method outperformed other LBP variants.	- Lack of detailed data in abstract regarding the dataset and sample size - Limitation in the use of method (echocardiography imaging) - Need for external validation for clinical adoption
Raghavendra et al (2022)[60]	A development study; ultrasound images of patients with HTN	Contourlet and shearlet feature extraction + locality sensitive discriminant analysis for dimension reduction + DT classifier	Automated detection of early structural heart muscle changes induced by HTN from ultrasound	It showed high ACC in discriminating hypertensive structural changes using only 2 features; it was proposed as a CAD tool for clinics.	- Lack of the inclusion of the sample size in the abstract - Testing only on a specific population - Need for validation in diverse datasets

Table 1. Continued.

Authors/Year	Study Type and Population	AI technique(s) Used in Studies	Purpose of the AI Application	Key Findings	Limitations and Challenges
Zhang et al (2020)[61]	A cross-sectional study; 625 subjects, 1,222 high-quality fundus images (Henan, China)	DL NN models	Prediction of HTN, hyperglycemia, dyslipidemia, and CVD risk factors from retinal fundus images	AUC for HTN of 0.766; hyperglycemia of 0.880; dyslipidemia of 0.703; it also predicted hematocrit and MCHC with AUC > 0.7.	- Moderate performance for HTN prediction - Limitation of dataset (only the rural Chinese population) - Uncertainty in generalizability
Abbas and Ibrahim (2020)[62]	A development and evaluation study; 4,270 retinal fundus images (3 public datasets + 1 private)	Dense feature transform layer + trained feature layer + deep residual learning (CNN)	Automated detection of hypertensive retinopathy (HR) from retinal fundus images	It achieved an SE of 93%, an SP of 95%, an ACC of 95%, and an AUC of 0.96 (10-fold CV); it outperformed state-of-the-art methods.	- Focus on retinal HR detection rather than direct BP measurement - Need for validation in prospective clinical settings

Note. SBP: Systolic blood pressure; DBP: Diastolic blood pressure; AUC: Area under the curve; CI: Confidence interval; COVID-19: Coronavirus disease 2019; LSTM: Long short-term memory; MCHC: Mean corpuscular hemoglobin concentration; AI: Artificial intelligence; ICU: Intensive care unit; HR: Heart rate; BMI: Body mass index; WC: Waist circumference; HC: Hip circumference; WHR: Waist-to-hip ratio; HRV: Heart rate variability; ACC: Accuracy; MRA: Magnetic resonance angiogram; ResNet: Residual network; SHAP: SHapley Additive exPlanations; Grad-CAM: Gradient-weighted class activation mapping; MAE: Mean absolute error; GBM: Gradient boosting; XGBoost: Extreme gradient boosting; NPV: Negative predictive value; AAMI: Association for the Advancement of Medical Instrumentation; PCA: Principal component analysis.

images, ECG, phonocardiograms (PCG), or PPG data in mixed ways.[17,19,27,29,34,35,38,42,44,48,49,56]

These models were used to evaluate HTN using non-invasive methods,[27,31-34,39,46-49,52,56] sort out kinds of pulmonary HTN,[27,31-34,39,46-49,52,56] spot early signs with images and wearables,[14,15,19,21,23,26-30,35] and detect through imaging and wearable technologies earlier.[18,24,27,33] They were also utilized to differentiate secondary HTN from essential HTN[17,25,29,38,42,43] and guess the risks or predict how things will proceed.[17,35,36,40,41,44-47]

According to QUADAS-2 (Table 2), most studies had the lowest risk of error in the index and reference standard test sections. Conversely, some lacked external validation; thus, their risk was assessed as "uncertain" in the flow and timing sections.

Diagnostic Performance

Diagnostic performance was generally high across the included studies. The reported accuracy ranged from 71% (34) to over 97%.[28,52] For example, Zou et al[21] achieved AUC values of 0.97 in pulmonary HTN detection from chest X-rays, and Jusic et al[25] reported ~90% accuracy for essential HTN classification using microRNA-based SVM models. In several instances, AI performance exceeded that of human experts; for example, Han et al[26] and Diller et al[28] found that AI assistance improved or outperformed radiologists in image interpretation.

Based on quality assessment, most high-performing studies were categorized as low risk of bias; however, those with small datasets (e.g., Priya et al [15] and Quachtran et al[16]) or retrospective designs were downgraded to "unclear" in patient selection or flow/timing domains.

Limitations and Challenges

Although the results were promising, several methodological issues remained:

Small sample sizes made the results ungeneralizable across more than 10 studies.[15,16,23,36,52,53,57,58,63] In addition, lack of external validation was a common problem, especially in studies that used imaging data and EHRs.[17,19,21,29,31,34,35]

Moreover, data heterogeneity and incomplete information, especially in models built on EHRs, reduced confidence in the results.[17,31,38] Additionally, model interpretability was a barrier for DL models.[27,34,48,49]

The risk of bias and overfitting in data collected from a single institution was also identified as well.[20,32,46]

Overall, the QUADAS-2 evaluation revealed that most studies were of acceptable methodological quality. However, future research should focus on large-scale prospective validation, standardized reporting, and the use of explainable AI to enhance the clinical application of these technologies.

Quality Assessment

According to the obtained data (Table 2), the majority of included studies showed a low risk of bias in the domains of patient selection, index test, and reference standard. Specifically, 38 studies (76%) were rated as low risk across these domains. In contrast, in the domain of flow and timing, 14 studies (28%) were judged as "unclear" due to insufficient reporting. Additionally, some methodological limitations (e.g., small sample size, lack of external validation, or incomplete patient information) raised concerns about potential bias in 8 studies (16%). Even though these problems were present, over two-thirds of the studies were performed well and provided trustworthy information, highlighting the validity of the overall evidence.

Discussion

This review thoroughly delved into how AI is being used to diagnose and handle high BP, along with its different types (e.g., primary, pulmonary, intracranial, and secondary HTN). The studies revealed that AI methods go from basic ML models (e.g., SVMs and RFs) to more advanced DL setups (e.g., convolutional neural networks) and models that mix clinical info with imaging data. The findings confirmed that AI has real potential to improve diagnostic accuracy, early detection, risk stratification,

Table 2. Quality Assessment of Included Studies Using QUADAS-2

Authors/Year	Patient Selection	Index Test	Reference Standard	Flow and Timing	Comments/Justification
Priya et al (2021)[15]	Low	Low	Low	Low	- Small sample (72) - Variation in model performance in terms of the classifier
Quachtran et al (2016)[16]	Low	Low	Low	Low	- Small sample - Need for further validation
Wu et al (2013)[17]	Low	Low	Unclear	Unclear	- Probable missing/incomplete entries due to the use of real-world EHRs - Need for prospective validation
Yoshihara et al (2024)[18]	Low	Low	Low	Low	- Limited study population (influenza-like illness patients)
Zhu et al (2020)[19]	Low	Low	Low	Low	- Lack of external validation - Limited population
Montagna et al (2022)[20]	Low	Low	Unclear	Low	- Population survey - Trade-off sensitivity vs. specificity
Zou et al (2020)[21]	Low	Low	Low	Low	- High internal/external accuracy - Need for prospective validation
Li (2021)[22]	Unclear	Unclear	Unclear	Unclear	- A review-type article - Lack of a specific population - Limited generalizability
Kusunose et al (2022)[23]	Low	Low	Low	Low	- Small sample size (142) - Need for broader validation
Juyal et al (2024)[24]	Unclear	Low	Unclear	Unclear	- Review-type study - Lack of original patient cohort
Jusic et al (2023)[25]	Low	Low	Low	Low	- Limited sample size (174 participants) - Need for independent validation
Han et al (2023)[26]	Low	Low	Low	Low	- Retrospective nature - Less accuracy of PAH-CHD - Need for further clinical validation
Guo et al (2024)[27]	Low	Low	Low	Low	- Moderate diagnostic performance (AUC < 0.8) - PCG quality dependent
Diller et al (2022)[28]	Low	Low	Low	Low	- Need for validation in more diverse populations
Buffolo et al (2021)[29]	Low	Low	Low	Low	- Inclusion of 4,059 patients - Need for further validation
Aras et al (2022)[30]	Low	Low	Low	Low	- Need for 12-lead ECG - Need for diverse clinical validation
Angelaki et al (2022)[31]	Low	Low	Low	Low	- Limitation in the inclusion of subjects (patients without CVD) - Need for further validation
Adeleke et al (2024)[32]	Low	Low	Unclear	Low	- Single institution - Data gaps (COVID-19)
Angelaki et al (2025)[33]	Low	Low	Low	Low	- Prospective proof-of-concept - Need for validation on wearables
Martinez-Ríos et al (2022)[34]	Unclear	Low	Unclear	Unclear	- Lack of population details - Lack of improvement in performance by using the fusion method
Anand et al (2024)[35]	Low	Low	Low	Low	- Moderate specificity - Exclusion of direct heart pressure measures
Golino et al (2014)[36]	Low	Low	Unclear	Low	- Small, non-representative student sample - Limited generalizability
Seffens et al (2015)[37]	Low	Low	Unclear	Low	- Focus on data imputation - Limited external validation
Ren et al (2019)[38]	Low	Low	Low	Low	- Focus on kidney disease - Need for external validation
Nour and Polat (2020)[39]	Unclear	Low	Unclear	Low	- Small dataset - Possible overfitting - Lack of clinical validation
López-Martínez et al F (2020)[40]	Low	Low	Unclear	Low	- Imbalanced dataset - Low sensitivity - Need for external validation
Islam et al (2022)[41]	Low	Low	Unclear	Low	- Population-level survey - Uncertainty in generalizability outside South Asia
Shih et al (2022)[42]	Low	Low	Low	Low	- Inclusion of two cohorts in Taiwan - Limited external validation
Orozco Torres et al (2022)[43]	Low	Low	Unclear	Low	- Population-based - Limited sample details - Need for diverse testing

Table 2. Continued.

Authors/Year	Patient Selection	Index Test	Reference Standard	Flow and Timing	Comments/Justification
Fang et al (2023)[44]	Unclear	Low	Unclear	Low	- Chinese cohort - Need for external validation
Kurniawan et al (2024)[45]	Low	Low	Unclear	Low	- Indonesian adults - Need for external validation
Sakka et al (2020)[46]	Low	Low	Low	Low	- Single institution (Petra University) - Lack of external validation
Islam et al (2023)[47]	Low	Low	Unclear	Low	- Small sample (n = 612) - Ethiopian population - Need for external validation
El-Dahshan et al EA (2023)[48]	Unclear	Low	Low	Low	- Public PPG datasets - Probable overfitting due to extreme accuracy
Esmaelpoor et al (2020)[49]	Low	Low	Low	Low	- Use of fundus images only - Need for prospective clinical validation
Chen et al (2022)[50]	Low	Low	Low	Low	- Small dataset (200) - Uncertainty in generalizability
Evdochim et al (2022)[51]	Low	Low	Low	Low	- Use of MIMIC-II only - Lack of deployment validation in clinical systems
Soh et al (2020)[52]	Low	Low	Unclear	Low	- Small, heterogeneous PPG dataset - Limited generalizability
Rajput et al (2019)[53]	Low	Low	Low	Low	- Inclusion of only two datasets - Lack of external validation
Rajput et al (2022)[54]	Low	Low	Low	Low	- Small dataset - Inclusion of ICU patients - Need for external validation
Huang et al (2021)[55]	Unclear	Low	Low	Low	- BCG signals - Lack of previous data (the first study) - Need for external validation
Miao et al (2020)[56]	Low	Low	Low	Low	- Use of MIMIC-II only - Need for prospective validation
Kublanov et al (2017)[57]	Low	Low	Low	Low	- Inclusion of ICU/arrhythmia population - Less ACC of SBP estimation
Kandil et al (2019)[58]	Low	Low	Low	Low	- Small sample (70) - Use of HRV only - Need for external validation
Gudigar et al (2019)[59]	Low	Low	Low	Low	- Use of fundus images - Moderate performance - Limited generalizability
Raghavendra et al U (2022)[60]	Low	Low	Low	Low	- Small sample (66) - Expensiveness of MRA imaging - Need for external validation
Zhang et al (2020)[61]	Low	Low	Low	Low	- Use of echocardiography only - Unclear sample size - Need for external validation
Abbas and Ibrahim (2020)[62]	Unclear	Low	Low	Low	- Use of ultrasound images - Unclear sample size - Need for diverse validation

Note. SBP: Systolic blood pressure; PCG: Phonocardiogram; PPG: Photoplethysmography; ECG: Electrocardiogram; COVID-19: Coronavirus disease 2019; QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies (version 2); PAH-CHD: Pulmonary arterial hypertension associated with congenital heart disease; CVD: Cardiovascular disease; MIMIC: Multiparameter intelligent monitoring in intensive care; ICU: Intensive care unit; HRV: Heart rate variability; MRA: Magnetic resonance angiogram.

and prognosis prediction in hypertensive disorders. However, when the quality of studies was assessed using the QUADAS-2 tool, it was found that the reliability and generalizability of results largely relied on the methodology and design of the studies themselves.

Diagnostic Accuracy and Performance

Most AI models demonstrated high diagnostic performance, with DL models for PH diagnosis from chest X-rays achieving AUC values up to 0.97,[21] occasionally surpassing expert radiologists.[26,28] Similarly, SVM-based models using circulating microRNAs reached 90% accuracy in diagnosing essential HTN.[25] These high-performing studies were generally rated as low risk of bias in QUADAS-2, particularly in the domains of index test and reference standard, strengthening confidence in their results.

The use of multimodal data fusion, combining physiological signals (e.g., PPG, PCG, ECG, and imaging), further enhanced diagnostic potential in

several studies.[27,30,34] However, the benefits of fusion were not consistent across all investigations. For example, Martinez-Ríos et al reported that the accuracy of BP detection using PPG was only 71%, and the QUADAS-2 evaluation rated the study as "uncertain" in terms of patient selection and flow/timing, as the sampling method was not adequately described.[34] This indicates that poor design can reduce the benefits of multimodal approaches. Likewise, Priya et al and Quachtran et al obtained relatively high accuracies of over 80% and 90%, respectively, but both were conducted on very small samples (72 and 60 subjects). This issue reduced the external validity of their results, even though they were somewhat biased in other parts of QUADAS-.[15,16]

Applications Across Hypertension Subtypes

The reviewed studies covered a wide range of HTN subtypes, demonstrating AI adaptability. The diagnosis of PH was a major focus, and AI models used imaging data (e.g., echocardiography, cardiac magnetic resonance imaging, chest X-ray, and electrocardiogram) to noninvasively classify PH subtypes and predict pulmonary artery systolic pressure.[15,19,21,23,28,35] A number of these imaging-based models have shown extremely high diagnostic accuracy, with AUC values above 0.90.[19,21,28] suggesting that these models have the potential to reduce the reliance on invasive procedures (e.g., right heart catheterization), thereby improving patient safety and comfort.[19,21,28] Nonetheless, some studies lacked external validation,[19,23] highlighting that despite high accuracy, reproducibility across diverse populations remains unproven.

For secondary hypertension, DL models incorporating natural language processing on EHRs demonstrated accurate differential diagnosis by integrating textual clinical notes with numerical data, in some cases even outperforming clinicians in retrospective analysis.[17,38] These findings underscore the potential of AI in managing complex clinical scenarios, where multifaceted data sources must be synthesized. However, quality assessment using QUADAS-2 highlighted unclear bias in patient selection and flow/timing due to incomplete or heterogeneous datasets,[17] underlining a major challenge of applying AI in real-world clinical data, where missing or inconsistent entries can compromise reliability.

Furthermore, studies using data from wearable technologies and simplified ECG signals have shown that AI can enable early detection and continuous monitoring, providing non-invasive, real-time assessment of high BP risk, even outside of traditional clinical settings.[24,31,33] Some studies were prospective,[33] which gave them higher methodological power. However, others were retrospective and limited to single-institution cohorts,[31] which made their results less generalizable. Based on QUADAS-2 assessments, quite a few studies had an uncertain risk for external validation. Thus, even though the results align with current trends in personalized medicine, where treatments are customized using algorithms based on patient traits, these methods still need more testing. They should be tested with larger groups of people and in different locations.

Limitations and Challenges

Although these results are promising, there are some very significant limitations that hinder their application in common clinical practice. By the QUADAS-2 assessment, around 76% of the studies contained a low risk of bias in the index test, reference standard, and patient selection. Nonetheless, 28% were deemed suboptimal due to inadequate reporting, unclear timing, or lack of repeated measurement, and 16% had miscellaneous problems (e.g., very small sample sizes, lack of external validation, or missing patient information), indicating that despite good diagnostic performance, evidence is still limited due to methodological shortcomings.

A common problem was the small size of the dataset,[15,16,58] causing the models to "overfit" too much to existing data and not easily perform well on new data. Similarly, the lack of external validation was a major problem,[19,29,31,34] limiting our confidence in the performance of the models across populations and healthcare systems. It should be noted that data heterogeneity, especially when EHRs are incomplete or inconsistent, reduces confidence in the model results and affects their reliability.[17,31] Furthermore, the risk of selection bias was high when research was conducted on data from only one institution.[29,32,46] A critical barrier was the interpretability of DL models, where "black-box" decision-making hinders clinician trust and regulatory approval.[27,34,48] Beyond methodological concerns, successful integration of AI into clinical workflows will also require addressing ethical and practical issues (e.g., data privacy, bias mitigation, and cost-effectiveness).

Future Implications

To improve the generalizability and reproducibility of AI models in HTN diagnosis, future research should prioritize prospective, multi-center validation studies with large and diverse populations. Moreover, standardized and transparent reporting of patient selection, study flow, and timing is essential for reducing the "unclear" bias observed in nearly one-third of the included studies. Furthermore, it is better to increase the transparency of models using explainable AI methods (e.g., GradCAM^{++}) and other interpretable tools in order to gain clinician trust, facilitate regulatory approvals, and simplify their integration into clinical systems.[27,48]

In addition, expanding data sources beyond conventional medical images and physiological signals (e.g., the use of multiomics, wearable data, and longitudinal monitoring) can help us more accurately categorize risks, diagnose diseases earlier, and optimize personalized treatments.[24,27,30,33,34] For example, when multimodal signals (e.g., PPG, ECG, PCG, and medical images) are combined, diagnostic accuracy can be greatly

improved. However, the actual impact of this combination largely depends on the methodological rigor and quality of the input data.[27,30,34] Similarly, studies on wearable ECGs demonstrate that this technology can enable non-invasive, real-time monitoring, but external validation in larger populations is still needed.[24,33,34]

Ultimately, a collaborative effort from clinicians, programmers, and regulators is required for AI to be truly useful and safe in diagnosing high BP. When there is a collaboration among the experts, it is possible to set quality standards, reduce the risk of data bias from a particular center, and ensure that the technology is used ethically, safely, and fairly. Accordingly, the existing models can become powerful clinical tools that can actually prevent complications and deaths from high BP worldwide.

Conclusion

The findings of this article demonstrated how AI could revolutionize the way we diagnose and manage high BP, adding incredible precision to various treatment options. Of course, several challenges remain, including properly validating the data, accurately understanding the results, and understanding how to integrate this technology into healthcare systems. However, the increasing advances make the future look promising. Indeed, treatments become more personalized and tailored where early detection is easier, ultimately reducing high BP and its complications around the world.

Authors' Contribution

Conceptualization: Ali Valizadeh, Zahra Hosseinzadeh.
Data curation: Zahra Hosseinzadeh, Elnaz Jalilian.
Formal analysis: Zahra Hosseinzadeh.
Funding acquisition: Ali Valizadeh.
Investigation: Zahra Hosseinzadeh, Ali Valizadeh.
Methodology: Zahra Hosseinzadeh.
Project administration: Ali Valizadeh, Zahra Hosseinzadeh, Elnaz Jalilian.
Resources: Zahra Hosseinzadeh
Software: Ali Valizadeh, Elnaz Jalilian.
Supervision: Ali Valizadeh.
Validation: Ali Valizadeh.
Visualization: Elnaz Jalilian.
Writing–original draft: Zahra Hosseinzadeh, Elnaz Jalilian.
Writing–review & editing: Ali Valizadeh, Elnaz Jalilian.

Competing Interests

None.

Data Availability Statement

All data generated or analyzed during this study are included in this article.

Ethical Approval

No formal approval of the ethics committee was required, as this work involved a review of previously published studies. All included studies were accurately cited, and the review adhered to principles of transparency, reproducibility, and avoidance of data misrepresentation.

Funding

This study was self-funded.

References

1. Zhou B, Perel P, Mensah GA, Ezzati M. Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension. Nat Rev Cardiol. 2021;18(11):785-802. doi: 10.1038/s41569-021-00559-8.
2. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in blood pressure from 1975 to 2015: a pooled analysis of 1479 population-based measurement studies with 19·1 million participants. Lancet. 2017;389(10064):37-55. doi: 10.1016/s0140-6736(16)31919-5.
3. Egan BM, Kjeldsen SE, Grassi G, Esler M, Mancia G. The global burden of hypertension exceeds 1.4 billion people: should a systolic blood pressure target below 130 become the universal standard? J Hypertens. 2019;37(6):1148-53. doi: 10.1097/hjh.0000000000002021.
4. Chaikijurajai T, Laffin LJ, Tang WHW. Artificial intelligence and hypertension: recent advances and future outlook. Am J Hypertens. 2020;33(11):967-74. doi: 10.1093/ajh/hpaa102.
5. Kim HL, Lee EM, Ahn SY, Kim KI, Kim HC, Kim JH, et al. The 2022 focused update of the 2018 Korean Hypertension Society Guidelines for the management of hypertension. Clin Hypertens. 2023;29(1):11. doi: 10.1186/s40885-023-00234-9.
6. Whelton PK, Carey RM, Aronow WS, Casey DE Jr, Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol. 2018;71(19):e127-248. doi: 10.1016/j.jacc.2017.11.006.
7. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension. Eur Heart J. 2018;39(33):3021-104. doi: 10.1093/eurheartj/ehy339.
8. Cho JS, Park JH. Application of artificial intelligence in hypertension. Clin Hypertens. 2024;30(1):11. doi: 10.1186/s40885-024-00266-9.
9. Krittanawong C, Zhang H, Wang Z, Aydar M, Kitai T. Artificial intelligence in precision cardiovascular medicine. J Am Coll Cardiol. 2017;69(21):2657-64. doi: 10.1016/j.jacc.2017.03.571.
10. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, et al. Artificial intelligence in cardiology. J Am Coll Cardiol. 2018;71(23):2668-79. doi: 10.1016/j.jacc.2018.03.521.
11. Krittanawong C, Bomback AS, Baber U, Bangalore S, Messerli FH, Wilson Tang WH. Future direction for using artificial intelligence to predict and manage hypertension. Curr Hypertens Rep. 2018;20(9):75. doi: 10.1007/s11906-018-0875-x.
12. Kotchen TA, Cowley AW Jr, Liang M. Ushering hypertension into a new era of precision medicine. Jama. 2016;315(4):343-4. doi: 10.1001/jama.2015.18359.
13. Tsoi K, Yiu K, Lee H, Cheng HM, Wang TD, Tay JC, et al. Applications of artificial intelligence for hypertension management. J Clin Hypertens (Greenwich). 2021;23(3):568-74. doi: 10.1111/jch.14180.
14. Gudigar A, Kadri NA, Raghavendra U, Samanth J, Maithri M, Inamdar MA, et al. Automatic identification of hypertension and assessment of its secondary effects using artificial intelligence: a systematic review (2013-2023). Comput Biol Med. 2024;172:108207. doi: 10.1016/j.compbiomed.2024.108207.
15. Priya S, Aggarwal T, Ward C, Bathla G, Jacob M, Gerke A, et al. Radiomics detection of pulmonary hypertension via texture-based assessments of cardiac MRI: a machine-learning model comparison-cardiac MRI radiomics in pulmonary hypertension. J Clin Med. 2021;10(9):1921. doi: 10.3390/

jcm10091921.
16. Quachtran B, Hamilton R, Scalzo F. Detection of intracranial hypertension using deep learning. In: 2016 23rd International Conference on Pattern Recognition (ICPR). Cancun, Mexico: IEEE; 2016. p. 2491-6. doi: 10.1109/icpr.2016.7900010.
17. Wu L, Huang L, Li M, Xiong Z, Liu D, Liu Y, et al. Differential diagnosis of secondary hypertension based on deep learning. Artif Intell Med. 2023;141:102554. doi: 10.1016/j.artmed.2023.102554.
18. Yoshihara H, Tsugawa Y, Fukuda M, Okiyama S, Nakayama T. Detection of hypertension from pharyngeal images using deep learning algorithm in primary care settings in Japan. BMJ Health Care Inform. 2024;31(1). doi: 10.1136/bmjhci-2023-100824.
19. Zhu F, Xu D, Liu Y, Lou K, He Z, Zhang H, et al. Machine learning for the diagnosis of pulmonary hypertension. Kardiologiia. 2020;60(6):953. doi: 10.18087/cardio.2020.6.n953.
20. Montagna S, Pengo MF, Ferretti S, Borghi C, Ferri C, Grassi G, et al. Machine learning in hypertension detection: a study on World Hypertension Day data. J Med Syst. 2022;47(1):1. doi: 10.1007/s10916-022-01900-5.
21. Zou XL, Ren Y, Feng DY, He XQ, Guo YF, Yang HL, et al. A promising approach for screening pulmonary hypertension based on frontal chest radiographs using deep learning: a retrospective study. PLoS One. 2020;15(7):e0236378. doi: 10.1371/journal.pone.0236378.
22. Li X, Kang N, Qi X, Huang Y. Artificial intelligence in the diagnosis of cirrhosis and portal hypertension. J Med Ultrason (2001). 2022;49(3):371-9. doi: 10.1007/s10396-021-01153-8.
23. Kusunose K, Hirata Y, Yamaguchi N, Kosaka Y, Tsuji T, Kotoku J, et al. Deep learning for detection of exercise-induced pulmonary hypertension using chest X-ray images. Front Cardiovasc Med. 2022;9:891703. doi: 10.3389/fcvm.2022.891703.
24. Juyal A, Bisht S, Singh MF. Smart solutions in hypertension diagnosis and management: a deep dive into artificial intelligence and modern wearables for blood pressure monitoring. Blood Press Monit. 2024;29(5):260-71. doi: 10.1097/mbp.0000000000000711.
25. Jusic A, Junuzovic I, Hujdurovic A, Zhang L, Vausort M, Devaux Y. A machine learning model based on microRNAs for the diagnosis of essential hypertension. Noncoding RNA. 2023;9(6):64. doi: 10.3390/ncrna9060064.
26. Han PL, Jiang L, Cheng JL, Shi K, Huang S, Jiang Y, et al. Artificial intelligence-assisted diagnosis of congenital heart disease and associated pulmonary arterial hypertension from chest radiographs: a multi-reader multi-case study. Eur J Radiol. 2024;171:111277. doi: 10.1016/j.ejrad.2023.111277.
27. Guo L, Khobragade N, Kieu S, Ilyas S, Nicely PN, Asiedu EK, et al. Development and evaluation of a deep learning-based pulmonary hypertension screening algorithm using a digital stethoscope. J Am Heart Assoc. 2025;14(3):e036882. doi: 10.1161/jaha.124.036882.
28. Diller GP, Benesch Vidal ML, Kempny A, Kubota K, Li W, Dimopoulos K, et al. A framework of deep learning networks provides expert-level accuracy for the detection and prognostication of pulmonary arterial hypertension. Eur Heart J Cardiovasc Imaging. 2022;23(11):1447-56. doi: 10.1093/ehjci/jeac147.
29. Buffolo F, Burrello J, Burrello A, Heinrich D, Adolf C, Müller LM, et al. Clinical score and machine learning-based model to predict diagnosis of primary aldosteronism in arterial hypertension. Hypertension. 2021;78(5):1595-604. doi: 10.1161/hypertensionaha.121.17444.
30. Aras MA, Abreau S, Mills H, Radhakrishnan L, Klein L, Mantri N, et al. Electrocardiogram detection of pulmonary hypertension using deep learning. J Card Fail. 2023;29(7):1017-28. doi: 10.1016/j.cardfail.2022.12.016.
31. Angelaki E, Barmparis GD, Kochiadakis G, Maragkoudakis S, Savva E, Kampanieris E, et al. Artificial intelligence-based opportunistic screening for the detection of arterial hypertension through ECG signals. J Hypertens. 2022;40(12):2494-501. doi: 10.1097/hjh.0000000000003286.
32. Adeleke O, Adebayo S, Aworinde H, Adeleke O, Adeniyi AE, Aroba OJ. Machine learning evaluation of a hypertension screening program in a university workforce over five years. Sci Rep. 2024;14(1):30255. doi: 10.1038/s41598-024-74360-1.
33. Angelaki E, Barmparis GD, Fragkiadakis K, Maragkoudakis S, Zacharis E, Plevritaki A, et al. Diagnostic performance of single-lead electrocardiograms for arterial hypertension diagnosis: a machine learning approach. J Hum Hypertens. 2025;39(1):58-65. doi: 10.1038/s41371-024-00969-4.
34. Martinez-Ríos E, Montesinos L, Alfaro-Ponce M. A machine learning approach for hypertension detection based on photoplethysmography and clinical data. Comput Biol Med. 2022;145:105479. doi: 10.1016/j.compbiomed.2022.105479.
35. Anand V, Weston AD, Scott CG, Kane GC, Pellikka PA, Carter RE. Machine learning for diagnosis of pulmonary hypertension by echocardiography. Mayo Clin Proc. 2024;99(2):260-70. doi: 10.1016/j.mayocp.2023.05.006.
36. Golino HF, de Brito Amaral LS, Duarte SF, Gomes CM, de Jesus Soares de, Dos Reis LA, et al. Predicting increased blood pressure using machine learning. J Obes. 2014;2014:637635. doi: 10.1155/2014/637635.
37. Seffens W, Evans C, Taylor H. Machine learning data imputation and classification in a multicohort hypertension clinical study. Bioinform Biol Insights. 2015;9(Suppl 3):43-54. doi: 10.4137/bbi.S29473.
38. Ren Y, Fei H, Liang X, Ji D, Cheng M. A hybrid neural network model for predicting kidney disease in hypertension patients based on electronic health records. BMC Med Inform Decis Mak. 2019;19(Suppl 2):51. doi: 10.1186/s12911-019-0765-4.
39. Nour M, Polat K. Automatic classification of hypertension types based on personal features by machine learning algorithms. Math Probl Eng. 2020;2020(1):2742781. doi: 10.1155/2020/2742781.
40. López-Martínez F, Núñez-Valdez ER, Crespo RG, García-Díaz V. An artificial neural network approach for predicting hypertension using NHANES data. Sci Rep. 2020;10(1):10620. doi: 10.1038/s41598-020-67640-z.
41. Islam SM, Talukder A, Awal MA, Siddiqui MM, Ahamad MM, Ahammed B, et al. Machine learning approaches for predicting hypertension and its associated factors using population-level data from three South Asian countries. Front Cardiovasc Med. 2022;9:839379. doi: 10.3389/fcvm.2022.839379.
42. Shih LC, Wang YC, Hung MH, Cheng H, Shiao YC, Tseng YH, et al. Prediction of white-coat hypertension and white-coat uncontrolled hypertension using machine learning algorithm. Eur Heart J Digit Health. 2022;3(4):559-69. doi: 10.1093/ehjdh/ztac066.
43. Orozco Torres JA, Medina Santiago A, Villegas Izaguirre JM, Amador García M, Delgado Hernández A. Hypertension diagnosis with backpropagation neural networks for sustainability in public health. Sensors (Basel). 2022;22(14):5272. doi: 10.3390/s22145272.
44. Fang M, Chen Y, Xue R, Wang H, Chakraborty N, Su T, et al. A hybrid machine learning approach for hypertension risk prediction. Neural Comput Appl. 2023;35(20):14487-97. doi: 10.1007/s00521-021-06060-0.
45. Kurniawan R, Utomo B, Siregar KN, Ramli K, Besral B, Suhatril RJ, et al. Hypertension prediction using machine learning algorithm among Indonesian adults. IAES Int J Artif Intell. 2023;12(2):776-84. doi: 10.11591/ijai.v12.i2.pp776-784.
46. Sakka Y, Qarashai D, Altarawneh A. Predicting hypertension using machine learning: a case study at Petra university. Int J Adv Comput Sci Appl. 2023;14(3):586-91.

47. Islam MM, Alam MJ, Maniruzzaman M, Ahmed N, Ali MS, Rahman MJ, et al. Predicting the risk of hypertension using machine learning algorithms: a cross sectional study in Ethiopia. PLoS One. 2023;18(8):e0289613. doi: 10.1371/journal.pone.0289613.
48. El-Dahshan EA, Bassiouni MM, Khare SK, Tan RS, Rajendra Acharya U. ExHyptNet: an explainable diagnosis of hypertension using EfficientNet with PPG signals. Expert Syst Appl. 2024;239:122388. doi: 10.1016/j.eswa.2023.122388.
49. Esmaelpoor J, Moradi MH, Kadkhodamohammadi A. A multistage deep neural network model for blood pressure estimation using photoplethysmogram signals. Comput Biol Med. 2020;120:103719. doi: 10.1016/j.compbiomed.2020.103719.
50. Chen Y, Zhang D, Karimi HR, Deng C, Yin W. A new deep learning framework based on blood pressure range constraint for continuous cuffless BP estimation. Neural Netw. 2022;152:181-90. doi: 10.1016/j.neunet.2022.04.017.
51. Evdochim L, Dobrescu D, Halichidis S, Dobrescu L, Stanciu S. Hypertension detection based on photoplethysmography signal morphology and machine learning techniques. Appl Sci. 2022;12(16):8380. doi: 10.3390/app12168380.
52. Soh DC, Ng EY, Jahmunah V, Oh SL, San TR, Acharya UR. A computational intelligence tool for the detection of hypertension using empirical mode decomposition. Comput Biol Med. 2020;118:103630. doi: 10.1016/j.compbiomed.2020.103630.
53. Rajput JS, Sharma M, Acharya UR. Hypertension diagnosis index for discrimination of high-risk hypertension ECG signals using optimal orthogonal wavelet filter bank. Int J Environ Res Public Health. 2019;16(21):4068. doi: 10.3390/ijerph16214068.
54. Rajput JS, Sharma M, Kumar TS, Acharya UR. Automated detection of hypertension using continuous wavelet transform and a deep neural network with ballistocardiography signals. Int J Environ Res Public Health. 2022;19(7):4014. doi: 10.3390/ijerph19074014.
55. Huang B, Chen W, Lin CL, Juang CF, Wang J. MLP-BP: a novel framework for cuffless blood pressure measurement with PPG and ECG signals based on MLP-Mixer neural networks. Biomed Signal Process Control. 2022;73:103404. doi: 10.1016/j.bspc.2021.103404.
56. Miao F, Wen B, Hu Z, Fortino G, Wang XP, Liu ZD, et al. Continuous blood pressure measurement from one-channel electrocardiogram signal using deep-learning techniques. Artif Intell Med. 2020;108:101919. doi: 10.1016/j.artmed.2020.101919.
57. Kublanov VS, Dolganov AY, Belo D, Gamboa H. Comparison of machine learning methods for the arterial hypertension diagnostics. Appl Bionics Biomech. 2017;2017:5985479. doi: 10.1155/2017/5985479.
58. Kandil H, Soliman A, Taher F, Ghazal M, Khalil A, Giridharan G, et al. A novel computer-aided diagnosis system for the early detection of hypertension based on cerebrovascular alterations. Neuroimage Clin. 2020;25:102107. doi: 10.1016/j.nicl.2019.102107.
59. Gudigar A, Raghavendra U, Devasia T, Nayak K, Danish SM, Kamath G, et al. Global weighted LBP based entropy features for the assessment of pulmonary hypertension. Pattern Recognit Lett. 2019;125:35-41. doi: 10.1016/j.patrec.2019.03.027.
60. Raghavendra U, En Wei JK, Gudigar A, Shetty A, Samanth J, Paramasivam G, et al. Automated diagnosis and assessment of cardiac structural alteration in hypertension ultrasound images. Contrast Media Mol Imaging. 2022;2022:5616939. doi: 10.1155/2022/5616939.
61. Zhang L, Yuan M, An Z, Zhao X, Wu H, Li H, et al. 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. 2020;15(5):e0233166. doi: 10.1371/journal.pone.0233166.
62. Abbas Q, Ibrahim ME. DenseHyper: an automatic recognition system for detection of hypertensive retinopathy using dense features transform and deep-residual learning. Multimed Tools Appl. 2020;79(41):31595-623. doi: 10.1007/s11042-020-09630-x.
63. Abdul-Rahman A, Morgan W, Yu DY. A machine learning approach in the non-invasive prediction of intracranial pressure using modified photoplethysmography. PLoS One. 2022;17(9):e0275417. doi: 10.1371/journal.pone.0275417.