electronics

Article

Machine Learning-Based Blood Pressure Prediction Using Cardiovascular Disease Data: A Comprehensive Comparative Study

Irina Naskinova [1], Mikhail Kolev [1,2,*], Dilyana Karova [3] and Mariyan Milev [1,3]

[1] Department of Mathematics, University of Architecture, Civil Engineering and Geodesy, 1 Hristo Smirnenski Blvd., 1164 Sofia, Bulgaria; inaskinova_fte@uacg.bg (I.N.); m.milev@feb.uni-sofia.bg (M.M.)
[2] Department of Applied Computer Science and Mathematical Modelling, Faculty of Mathematics and Computer Science, University of Warmia and Mazury, 10-710 Olsztyn, Poland
[3] Department of Statistics and Econometrics, Faculty of Economics and Business Administration, Sofia University "St. Kliment Ohridski", 125 Tsarigradsko Shosse Blvd., bl. 3, 1113 Sofia, Bulgaria; dilyana.nikolova@feb.uni-sofia.bg
* Correspondence: mkolev_fte@uacg.bg or kolev@matman.uwm.edu.pl

Abstract

Hypertension remains one of the most pressing public health challenges worldwide, affecting more than one billion individuals and serving as a principal risk factor for cardiovascular morbidity and mortality. Whilst blood pressure measurement constitutes a routine component of clinical practice, the capacity to predict blood pressure values from readily obtainable patient characteristics could substantially enhance preventive care strategies and facilitate timely intervention. The present study examines whether machine learning methodologies can reliably forecast blood pressure measurements utilizing cardiovascular risk factors in conjunction with demographic and anthropometric data. We have analyzed data from 68,616 individuals following rigorous quality assessment of 70,000 patient records obtained from Kaggle's cardiovascular disease repository. Beyond the 10 original variables, we engineered additional features encompassing demographic patterns, body composition indices, clinical risk indicators, and their interactions. Nine distinct predictive models were systematically evaluated, spanning from elementary baseline approaches through to sophisticated gradient boosting ensembles. CatBoost demonstrated superior performance, yielding systolic blood pressure predictions with a root mean squared error (RMSE) of 14.37 mmHg and coefficient of determination (R^2) of 0.265, alongside diastolic blood pressure predictions with RMSE of 8.57 mmHg and R^2 of 0.187. These modest explained variance values—substantially below unity—reveal a fundamental limitation: blood pressure proves remarkably resistant to prediction from the demographic, anthropometric, and clinical variables typically available in epidemiological datasets. These findings illuminate a sobering reality regarding blood pressure prediction from routinely collected clinical data. The observation that standard variables account for merely one-quarter of blood pressure variance should temper expectations for machine learning applications within this domain, whilst simultaneously underscoring the necessity for richer data sources or novel biomarkers to achieve clinically meaningful predictive accuracy.

Keywords: blood pressure prediction; hypertension; machine learning; cardiovascular disease; feature engineering; clinical decision support

Academic Editor: Hung-Yu Chien

Received: 14 November 2025
Revised: 22 December 2025
Accepted: 5 January 2026
Published: 10 January 2026

1. Introduction

1.1. The Challenge of Blood Pressure Management

Elevated blood pressure constitutes a silent yet profound threat to public health, affecting millions of individuals annually and contributing substantially to cardiovascular morbidity [1–3]. The World Health Organization estimates that approximately 1.28 billion adults worldwide suffer from hypertension, establishing it as a principal risk factor for heart disease, stroke, and premature mortality [4–6]. Notwithstanding extensive clinical research and the availability of efficacious pharmacological interventions, early detection of hypertension remains a considerable challenge, frequently resulting in severe health complications that might otherwise have been prevented through timely intervention.

Contemporary clinical practice typically relies upon periodic blood pressure assessments conducted during outpatient consultations—an approach characterized by notable limitations. Blood pressure exhibits considerable diurnal variation, and phenomena such as 'white coat hypertension' may substantially distort readings obtained in clinical settings [4,7]. Moreover, a significant proportion of the population does not attend regular medical appointments, thereby complicating timely diagnosis. This raises the pertinent question of whether it might be feasible to predict blood pressure utilizing existing medical records encompassing demographic, anthropometric, and clinical variables.

Recent advances in artificial intelligence have demonstrated considerable promise across diverse medical prediction tasks [8,9]. Investigators have successfully employed machine learning methodologies to diagnose various pathological conditions, forecast treatment responses, and stratify patient risk. Nevertheless, the specific application of machine learning to blood pressure prediction has received comparatively limited attention, particularly regarding systematic comparisons of different algorithmic approaches evaluated under consistent experimental conditions. This gap presents an opportunity for rigorous investigation.

1.2. What Others Have Tried

Research examining machine learning applications for hypertension risk assessment has yielded heterogeneous outcomes [10–13]. Certain investigations have concentrated on discriminating between hypertensive and normotensive patients through ensemble methods, achieving moderate classification accuracy in the range of 70–80% [14,15]. Neural networks have additionally been employed to forecast blood pressure trajectories from continuous monitoring data, with reported prediction errors typically ranging from 8 to 12 mmHg.

Notwithstanding these efforts, several notable gaps persist within the extant literature [16]. The binary classification of hypertension, whilst methodologically straightforward, fails to provide the nuanced blood pressure values that clinicians require for informed therapeutic decision-making. Furthermore, systematic comparisons amongst diverse machine learning approaches-spanning elementary linear models through to sophisticated neural architectures, utilizing consistent datasets and evaluation protocols, remain conspicuously absent. This deficiency renders it challenging to ascertain which methodological approaches are most efficacious for this particular predictive task.

1.3. What We Set Out to Do

The present study pursued four principal objectives:

Firstly, we endeavored to develop a comprehensive set of predictive features extending beyond elementary demographic variables. We constructed derived variables reflecting clinically meaningful patterns, including body mass index, composite cardiovascular risk

scores, metabolic syndrome indicators, and mathematical interaction terms designed to capture potential synergistic effects amongst risk factors.

Secondly, we systematically evaluated an extensive array of machine learning algorithms, encompassing simple baseline predictors, classical statistical approaches such as linear and ridge regression, tree-based ensemble methods including Random Forest and gradient boosting variants (LightGBM, XGBoost, CatBoost), and more sophisticated methodologies. This comprehensive evaluation framework enabled rigorous assessment of relative algorithmic performance.

Thirdly, we assessed model performance employing both conventional statistical metrics and clinically relevant thresholds, recognizing that practitioners frequently prioritize predictions falling within ± 5 or ± 10 mmHg of actual measurements—tolerances reflecting meaningful clinical accuracy.

Finally, we sought to provide practical guidance for future investigators and clinicians regarding model selection, realistic accuracy expectations, and considerations pertinent to real-world healthcare deployment.

1.4. Why This Matters

The capacity to accurately predict blood pressure holds potential to transform multiple facets of cardiovascular care. Primary care settings could proactively screen patients for elevated risk during routine consultations, even in the absence of contemporaneous blood pressure measurements. Telehealth providers could more effectively identify individuals warranting urgent face-to-face evaluation, whilst public health programs could direct resources towards high-risk communities for targeted intervention.

The fundamental challenge lies in determining whether machine learning can yield predictions of sufficient accuracy for clinical utility, whilst maintaining interpretability adequate to engender trust amongst healthcare professionals. Achieving this balance is essential for integrating advanced predictive analytics into routine clinical workflows and ultimately improving cardiovascular outcomes.

2. Methods

2.1. The Data

2.1.1. Where It Came from

The present investigation utilizes a publicly available dataset from Kaggle comprising health information for 70,000 individuals [17]. This dataset encompasses demographic attributes including age and sex, anthropometric measurements (height and weight), cardiovascular parameters (systolic and diastolic blood pressure), metabolic indicators (cholesterol and glucose levels), and lifestyle factors (smoking status, alcohol consumption, and physical activity). Additionally, it includes cardiovascular disease diagnosis status for each participant.

Originally compiled for cardiovascular disease classification research, we repurposed this dataset for blood pressure prediction. Whilst specific details regarding measurement protocols and population demographics are unavailable, the dataset's substantial size and comprehensive coverage of established cardiovascular risk factors render it suitable for our comparative methodological analysis.

2.1.2. Cleaning and Quality Checks

Clinical datasets frequently contain erroneous or physiologically implausible values, necessitating rigorous quality control procedures informed by established physiological parameters and clinical expertise.

For blood pressure measurements, we excluded values falling outside clinically plausible ranges: systolic pressures below 60 mmHg or exceeding 260 mmHg, and diastolic pressures below 40 mmHg or exceeding 180 mmHg. We additionally removed records where systolic pressure was documented as lower than diastolic pressure—a physiologically impossible scenario indicative of measurement or recording error.

Regarding anthropometric data, we excluded extreme values unlikely to represent genuine measurements: heights below 120 cm or exceeding 250 cm, and weights below 30 kg or exceeding 200 kg. Whilst rare individuals may exist outside these parameters, such outliers more plausibly represent data entry errors than genuine physiological values.

This exclusion process resulted in removal of 1325 blood pressure records (1.89% of the sample) and 59 anthropometric records (0.08%), yielding 68,616 high-quality records for analysis—a retention rate of 98%. This rigorous screening enhances confidence that subsequent models are founded upon genuine physiological relationships rather than measurement artefacts.

The cleaned dataset was partitioned into training (64%, n = 43,913), validation (16%, n = 10,979), and test (20%, n = 13,724) sets employing stratified sampling to ensure comparable blood pressure distributions across all three partitions.

Table 1 summarizes the key characteristics of the cleaned dataset.

Table 1. Dataset characteristics after quality screening. Values represent the cleaned dataset of 68,616 individuals retained from the original 70,000 records.

Characteristic	Value
Total samples (after cleaning)	68,616
Training set (64%)	43,913
Validation set (16%)	10,979
Test set (20%)	13,724
Total features (after engineering)	39
Missing values	0 (0%)
Age range	29.6–64.9 years
Gender distribution	Female: 66.4%, Male: 33.6%
Cardiovascular disease prevalence	50.0%

Note: Continuous variables presented as mean ± standard deviation; categorical variables as count (percentage). BP = blood pressure; BMI = body mass index.

The dataset was partitioned into three subsets to facilitate robust model development and unbiased performance evaluation, as summarized in Table 2.

Table 2. Dataset partitioning for model development and evaluation.

Partition	Samples	Percentage	Purpose
Training	43,913	64%	Model training
Validation	10,979	16%	Hyperparameter tuning, early stopping
Test	13,724	20%	Final performance evaluation
Total	68,616	100%	

Note: Stratified sampling was used to ensure representative distribution of blood pressure values across partitions. Validation set used for hyperparameter tuning; test set held out for final performance evaluation.

Table 3 details the specific exclusion criteria applied during data quality screening, along with the number of records affected by each criterion.

Table 3. Data quality exclusion criteria and sample attrition. Records were excluded based on physiological plausibility thresholds derived from clinical guidelines.

Exclusion Reason	Count	Percentage
SBP out of range [60, 260] mmHg	228	0.33%
DBP out of range [40, 180] mmHg	1015	1.45%
SBP $\leq$ DBP (physiologically impossible)	1236	1.77%
Height out of range [120, 250] cm	52	0.07%
Weight out of range [30, 200] kg	7	0.01%
Total unique exclusions	1384	1.98%
Final clean samples	68,616	98.02%

Note: Exclusions applied sequentially. Some records met multiple exclusion criteria; counts shown are for initial detection. SBP = systolic blood pressure; DBP = diastolic blood pressure.

2.2. Creating Predictive Features

Raw clinical data rarely presents in an optimal format for machine learning applications. Age expressed in days, as recorded in this dataset, conveys less clinical meaning than age in years. Similarly, weight and height considered independently provide less insight than their derived ratio, body mass index. Clinical evidence suggests that cardiovascular risk factors may interact synergistically; for instance, the combined effect of obesity and advancing age on blood pressure may exceed the sum of their individual contributions.

We constructed engineered features organized into six conceptual categories:

2.2.1. Demographic Features (3 Features)

To enhance clinical interpretability, we transformed age from days to years and established age decade groupings (30s, 40s, 50s, 60s) alongside broader categorical age bands. This approach enables models to capture potential non-linear age effects, particularly the accelerated cardiovascular risk observed beyond the fifth decade of life.

Anthropometric Features (6 features)

Beyond converting height to meters, we calculated several derived anthropometric indices:

Body Mass Index (BMI): weight in kilograms divided by height in meters squared, representing the standard clinical measure of adiposity.

BMI categories: employing World Health Organization classifications (underweight, normal weight, overweight, obese).

BMI z-scores: standardized BMI values, facilitating population-level comparisons.

Body surface area: calculated using the Du Bois formula, a parameter of relevance to cardiovascular physiology.

Ponderal index: an alternative weight-to-height ratio less commonly employed than BMI but potentially informative for certain body compositions.

2.2.2. Clinical Features (Excluding BP-Derived Variables)

A critical methodological decision merits emphasis: we explicitly excluded features derived from blood pressure values, specifically mean arterial pressure (MAP = (SBP + 2 $\times$ DBP)/3), pulse pressure (PP = SBP $-$ DBP), and hypertension stage classifications. Including such features would constitute target leakage, artificially inflating model performance by allowing predictors that mathematically encode the very values being predicted. The retained clinical features include:

Cardiovascular disease diagnosis: binary indicator of previously diagnosed cardiovascular disease.

Elevated cholesterol indicator: binary flag denoting above-normal cholesterol levels.

Elevated glucose indicator: binary flag denoting above-normal glucose levels.

High cholesterol and glucose indicators: binary flags for elevated values

Risk factor count: aggregate count of present risk factors encompassing smoking, alcohol consumption, elevated cholesterol, elevated glucose, and physical inactivity.

2.2.3. Interaction Features (8 Features)

Clinical observations indicate that risk factors do not merely accumulate additively; rather, they may interact multiplicatively. To explore such phenomena, we constructed interaction terms through multiplication of clinically relevant variable pairs:

- Age with BMI: adiposity may exert differential effects across the lifespan.
- Age with sex: cardiovascular risk trajectories differ between males and females.
- BMI with smoking status: two major modifiable risk factors with potential synergistic effects.
- Analogous interactions with physical activity levels and cholesterol status.

2.2.4. Polynomial Features (3 Features)

To accommodate potential non-linear relationships, we incorporated polynomial terms (squared values) for age, BMI, and their interaction. This approach enables models to capture accelerating effects—for instance, the relationship between age and blood pressure may steepen in older individuals.

2.2.5. Risk Scores (2 Features)

We constructed composite risk scores aggregating multiple individual risk factors:

Cardiovascular risk score: integrating age, BMI, cholesterol, glucose, smoking status, and physical inactivity into a unified risk metric.

Metabolic syndrome score: enumerating components of metabolic syndrome (elevated glucose, elevated cholesterol, obesity), with blood pressure explicitly excluded to prevent target leakage.

To examine associations between predictor features and blood pressure outcomes, we calculated Pearson correlation coefficients as presented in Table 4.

Table 4. Pearson correlation coefficients between predictor features and blood pressure outcomes. Only features without blood pressure derivation are included to prevent target leakage.

Feature	Systolic BP Correlation	Diastolic BP Correlation
Cardiovascular disease	0.428	0.340
Age (years)	0.239	0.125
Body weight	0.271	0.253
Body mass index	0.267	0.240
Age × BMI interaction	0.321	0.270
High cholesterol indicator	0.152	0.128
High glucose indicator	0.098	0.089

Note: All correlations significant at $p < 0.001$. BP-derived features (MAP, pulse pressure, hypertension stage) explicitly excluded from analysis.

Cardiovascular disease diagnosis emerged as the strongest predictor, reflecting its clinical basis in blood pressure assessment criteria. This finding, whilst anticipated, underscores the inherent challenge of predicting blood pressure from truly independent features.

The complete feature set commences with the ten original variables from the cardiovascular disease dataset, described in Table 5.

Table 5. Original features from the cardiovascular disease dataset (10 features). These baseline variables represent raw demographic, anthropometric, lifestyle, and clinical measurements collected during routine medical examinations.

#	Feature Name	Category	Formula/Description	Data Type
1	gender	Original	Biological sex (1 = female, 2 = male)	Categorical
2	height	Original	Height in centimeters	Continuous
3	weight	Original	Weight in kilograms	Continuous
4	cholesterol	Original	Cholesterol level (1 = normal, 2 = above normal, 3 = well above)	Ordinal
5	gluc	Original	Glucose level (1 = normal, 2 = above normal, 3 = well above)	Ordinal
6	cardio	Original	Cardiovascular disease diagnosis (0/1)	Binary
7	age	Original	Age in days (converted to years for analysis)	Continuous
8	smoke	Original	Smoking status (0 = non-smoker, 1 = smoker)	Binary
9	alco	Original	Alcohol consumption (0 = no, 1 = yes)	Binary
10	active	Original	Physical activity (0 = inactive, 1 = active)	Binary

Note: Raw features from the Kaggle cardiovascular disease dataset. Categorical variables encoded using one-hot encoding for model compatibility.

Extending the original variables, we derived 23 engineered features through domain-specific transformations, as detailed in Table 6.

Table 6. Engineered features derived through domain-specific transformations (23 features). These include age normalization, BMI calculation, categorical binning based on clinical thresholds, and interaction terms capturing combined risk factor effects.

#	Feature Name	Category	Formula/Description	Data Type
11	age_years	Demographic	age_days/365.25	Continuous
12	age_decade	Demographic	floor(age_years/10) × 10	Categorical
13	age_group	Demographic	Clinical age categories (<30, 30–40, . . ., 70+)	Categorical
14	height_m	Anthropometric	height_cm/100	Continuous
15	bmi	Anthropometric	weight/height_m^2	Continuous
16	bmi_category	Anthropometric	WHO classification (underweight, normal, overweight, obese)	Categorical
17	bmi_z_score	Anthropometric	(bmi − mean)/std	Continuous
18	bsa	Anthropometric	(height × weight/3600)ˆ(1/2) (Mosteller formula)	Continuous
19	bsa_dubois	Anthropometric	0.007184 × hˆ0.725 × wˆ0.425	Continuous
20	ponderal_index	Anthropometric	weight/height_m^3	Continuous
21	waist_to_height_proxy	Anthropometric	bmi/(height_m × 100)	Continuous
22	n_lifestyle_risk_factors	Lifestyle	smoke + alco + (1 − active)	Categorical
23	high_cholesterol	Clinical	cholesterol $\geq$ 2	Binary
24	high_glucose	Clinical	glucose $\geq$ 2	Binary
25	cholesterol_glucose_ interaction	Clinical	cholesterol × glucose	Categorical
26	n_metabolic_risk_factors	Clinical	high_cholesterol + high_glucose + (bmi $\geq$ 30)	Categorical
27	metabolic_syndrome_ score_no_bp	Risk Score	Count of metabolic syndrome components (excluding BP)	Categorical
28	cvd_risk_score	Risk Score	Weighted composite of risk factors (no BP)	Continuous
29	age_bmi	Interaction	age_years × bmi	Continuous
30	age_gender	Interaction	age_years × gender	Continuous

Table 6. *Cont.*

#	Feature Name	Category	Formula/Description	Data Type
31	age_cholesterol	Interaction	age_years × cholesterol	Continuous
32	age_smoke	Interaction	age_years × smoke	Continuous
33	age_active	Interaction	age_years × active	Continuous

Note: Features calculated using standard clinical formulas. BMI = weight (kg)/height (m)^2. Age converted from days to years using 365.25 days/year.

To capture potential synergistic effects amongst risk factors, we constructed interaction and polynomial features as presented in Table 7.

Table 7. Interaction and polynomial features designed to model non-linear relationships (6 features). These composite variables capture synergistic effects between risk factors that individual predictors cannot represent.

#	Feature Name	Category	Formula/Description	Data Type
34	bmi_gender	Interaction	bmi × gender	Continuous
35	bmi_cholesterol	Interaction	bmi × cholesterol	Continuous
36	bmi_smoke	Interaction	bmi × smoke	Continuous
37	age_sq	Polynomial	age_years^2	Continuous
38	bmi_sq	Polynomial	bmi^2	Continuous
39	age_bmi_sq	Polynomial	age_years × bmi^2	Continuous

Note: Interaction terms capture synergistic effects between risk factors. Polynomial features were standardized after creation to prevent numerical instability.

Table 8 provides a summary of all feature categories, confirming that none of the 75 input features are derived from blood pressure measurements, thereby avoiding target leakage.

Table 8. Feature category summary showing total feature count by category. All features are BP-independent, meaning no blood pressure measurements were used in their calculation.

Category	Feature Count	BP-Derived
Original	10	No
Demographic	3	No
Anthropometric	8	No
Lifestyle	1	No
Clinical (Safe)	4	No
Interaction	8	No
Polynomial	3	No
Risk Scores (Safe)	2	No
One-hot encoded categorical	42	No
Total	75	No

Note: One-hot encoded features expand categorical variables into binary columns. Total of 39 base features plus 36 one-hot encoded columns equals 75 total input features.

2.3. Data Preprocessing

Following feature engineering, we employed standard preprocessing techniques to prepare the data for machine learning algorithms. We distinguished between continuous features (age, BMI, laboratory values) and categorical features (sex, cholesterol categories).

Continuous features underwent median imputation (notwithstanding the absence of missing values in our cleaned dataset) followed by standardization to ensure uniform scaling [18,19] across all features, thereby preventing variables with larger numeric ranges from disproportionately influencing model training.

Categorical features underwent one-hot encoding, transforming variables such as cholesterol level (with three ordinal categories: normal, above normal, well above normal) into separate binary indicator variables.

Crucially, all preprocessing transformations were fitted exclusively to the training set prior to application to validation and test sets, thereby precluding information leakage that could compromise model evaluation integrity.

2.4. The Algorithms We Tested

We evaluated nine models spanning the spectrum from elementary baseline approaches to sophisticated ensemble methods.

2.4.1. Baseline Models

Every predictive modelling study requires baseline comparisons—elementary approaches establishing minimum performance thresholds against which more sophisticated methods can be evaluated.

Global Mean Baseline: For each patient, predict the mean systolic and diastolic pressure derived from the training set. This baseline addresses the fundamental question: can any model surpass simply predicting typical population blood pressure values?

Global Median Baseline: Analogous to the mean baseline, but employing median values. Medians demonstrate greater robustness to outliers than means, potentially providing marginally improved baseline performance.

2.4.2. Classical Regression Models

Linear Regression [20]: The foundational approach to statistical prediction. Linear regression identifies the optimal hyperplane relationship between features and blood pressure. Despite its simplicity, linear regression frequently performs surprisingly well and yields interpretable coefficients quantifying each feature's contribution to the prediction.

Ridge Regression: Linear regression incorporating L2 regularisation, imposing a penalty on large coefficients. This shrinkage mitigates overfitting, particularly important when features exhibit collinearity. We employed regularisation strength $\alpha = 1.0$.

2.4.3. Tree-Based Ensemble Models

Random Forest [20]: Constructs an ensemble of decision trees using random subsets of features and observations, subsequently averaging their predictions. Random Forest naturally accommodates non-linear relationships and feature interactions, requires minimal hyperparameter tuning, and provides feature importance rankings. We employed 100 trees with default hyperparameters.

LightGBM (Light Gradient Boosting Machine [21]): A contemporary gradient boosting algorithm optimized for computational efficiency and predictive accuracy. Unlike Random Forest's parallel tree construction, LightGBM builds trees sequentially, with each tree correcting its predecessor's errors. We have optimized hyperparameters using Optuna 3.3.0, a Bayesian optimization framework, conducting 30 trials to identify optimal configurations [22,23].

- Learning rate: 0.0998
- Number of leaves: 275
- Tree depth: 14
- Minimum samples per leaf: 18
- Feature and sample subsampling rates: 0.774 and 0.784
- L1 and L2 regularization: 4.623 and 4.297
- Maximum iterations: 1000 with early stopping criterion.

Tables 9–13 present the optimized hyperparameters for each model, obtained through Bayesian optimization with Optuna 3.3.0. Table 9 shows the Random Forest configuration.

Table 9. Random Forest hyperparameters. This ensemble method aggregates predictions from multiple decision trees trained on bootstrap samples with random feature subsets at each split.

Parameter	Value	Description
n_estimators	100	Number of trees
max_depth	3	Maximum tree depth
learning_rate	0.1	Step size shrinkage
min_samples_split	2	Minimum samples to split
min_samples_leaf	1	Minimum samples per leaf

Note: Parameters selected based on cross-validation performance. n_estimators balanced against computational cost.

Table 10. Gradient Boosting (scikit-learn) hyperparameters. The reference implementation uses depth-wise tree growth with deviance loss function and Friedman improvement criterion for node splitting.

Parameter	Value	Description
n_estimators	1000	Maximum number of trees
num_leaves	275	Maximum leaves per tree
max_depth	14	Maximum tree depth
learning_rate	0.0998	Step size shrinkage
subsample	0.774	Row subsampling ratio
colsample_bytree	0.784	Column subsampling ratio
lambda_l1	4.623	L1 regularization
lambda_l2	4.297	L2 regularization
min_child_samples	18	Minimum samples per leaf

Note: Reference implementation using scikit-learn. Early stopping used to prevent overfitting.

Table 11. LightGBM hyperparameters. This gradient boosting framework uses a leaf-wise tree growth strategy optimized for training speed and prediction accuracy on large-scale datasets.

Parameter	Value	Description
n_estimators	500	Maximum number of trees
max_depth	6	Maximum tree depth
learning_rate	0.05	Step size shrinkage
subsample	0.8	Row subsampling ratio
colsample_bytree	0.8	Column subsampling ratio
min_child_weight	1	Minimum sum of instance weight
reg_alpha	0.1	L1 regularization
reg_lambda	1.0	L2 regularization

Note: Leaf-wise tree growth strategy enables faster training. num_leaves controls model complexity.

The scikit-learn Gradient Boosting implementation was configured as presented in Table 10 [20].

LightGBM hyperparameters, optimized for the leaf-wise tree growth strategy, are presented in Table 11.

Table 12 details the XGBoost configuration, including regularization parameters to prevent overfitting [24].

The best-performing CatBoost model employed the hyperparameters presented in Table 13, leveraging ordered boosting and native categorical feature handling [25].

Table 12. XGBoost hyperparameters. The extreme gradient boosting implementation employs histogram-based tree construction with L1/L2 regularization to balance model complexity and generalization.

Parameter	Value	Description
iterations	500	Maximum number of trees
depth	6	Tree depth
learning_rate	0.05	Step size shrinkage
l2_leaf_reg	3.0	L2 regularization coefficient
border_count	128	Number of splits for numerical features
loss_function	RMSE	Optimization objective

Note: Regularization parameters (reg_alpha, reg_lambda) help prevent overfitting on tabular data.

Table 13. CatBoost hyperparameters (best performing model). CatBoost achieved superior performance through symmetric decision trees with ordered boosting, natively handling categorical features without explicit encoding.

Model	Category	Key Hyperparameters	Library
Global Mean	Baseline	None	Custom
Global Median	Baseline	None	Custom
Linear Regression	Linear	None	scikit-learn
Ridge	Linear (L2)	$\alpha = 1.0$	scikit-learn
Lasso	Linear (L1)	$\alpha = 1.0$	scikit-learn
Random Forest	Ensemble	100 trees, no max depth	scikit-learn
Gradient Boosting	Boosting	100 estimators, depth 3, lr 0.1	scikit-learn
LightGBM	Boosting	1000 est., 31 leaves, lr 0.1	lightgbm
XGBoost	Boosting	1000 est., depth 6, lr 0.1	xgboost
CatBoost	Boosting	1000 iter., depth 6, lr 0.1	catboost

Note: CatBoost handles categorical features natively. Ordered boosting reduces prediction shift.

2.5. Training Procedure

All models were trained on the training set (n = 43,913), utilizing the validation set (n = 10,979) for early stopping and hyperparameter selection. The test set (n = 13,724) remained entirely sequestered until final evaluation—a strict separation ensuring that performance estimates reflect genuine generalization capability rather than optimistic overfitting.

To optimize hyperparameters for gradient boosting models (CatBoost, XGBoost, LightGBM), we employed the validation set to identify optimal configurations [26]. Specifically, we implemented Bayesian optimization through Optuna, which efficiently navigates the hyperparameter space rather than exhaustively evaluating all combinations, conducting 30 trials per model.

All models predicted both systolic and diastolic blood pressure simultaneously via multi-output regression. Random seed 42 was employed throughout to ensure reproducibility.

2.6. Evaluation Metrics

Model performance was assessed using both statistical measures familiar to the research community and clinically meaningful thresholds of relevance to practicing clinicians.

$$RMSE = \sqrt{\left[\left(\frac{1}{n}\right)\Sigma\,(\hat{y}_i - y_i)^2\right]}\,, \tag{1}$$

$$MAE = \left(\frac{1}{n}\right)\Sigma\,|\hat{y}_i - y_i|\,, \tag{2}$$

$$R^2 = 1 - \left[\Sigma \frac{(\hat{y}_i - y_i)^2}{(\hat{y}_i - \bar{y})^2}\right], \quad (3)$$

$$Bias = \left(\frac{1}{n}\right)\Sigma\,(\hat{y}_i - y_i), \quad (4)$$

2.6.1. Statistical Metrics

Root Mean Squared Error (RMSE): The square root of the mean squared prediction error. RMSE penalises large errors more severely than small errors and shares the same units as blood pressure (mmHg), facilitating clinical interpretation. Lower values indicate superior performance, with values below 10 mmHg generally considered acceptable for blood pressure prediction.

Mean Absolute Error (MAE): The arithmetic mean of absolute differences between predictions and actual values. MAE demonstrates greater robustness to outliers than RMSE and offers intuitive interpretation—it represents the typical prediction error in mmHg.

Coefficient of Determination (R-squared): The proportion of blood pressure variance explained by the model, ranging from negative infinity (performance inferior to predicting the mean) to 1.0 (perfect prediction). R-squared values exceeding 0.8 indicate strong predictive power.

Bias: The mean prediction error (without absolute value). Bias reveals systematic under-prediction (positive bias) or over-prediction (negative bias). Unbiased models should exhibit bias approaching zero.

2.6.2. Clinical Accuracy Metrics

Beyond statistical measures, we assessed clinically relevant accuracy thresholds:

Within plus or minus 5 mmHg: The proportion of predictions falling within 5 mmHg of actual blood pressure. Clinical guidelines suggest that errors within 5 mmHg rarely affect therapeutic decisions. We targeted 80% or greater predictions meeting this threshold.

$$CI_{95}\% = [Quantile_{2.5}\%(RMSE_1, \ldots, RMSE_{1000}), Quantile_{97.5}\%(RMSE_1, \ldots, RMSE_{1000})]. \quad (5)$$

Within plus or minus 10 mmHg: The proportion within 10 mmHg, a more permissive yet clinically meaningful threshold. Most clinicians would consider predictions within 10 mmHg acceptable for screening and risk stratification purposes. We targeted 90% or greater.

2.6.3. Uncertainty Quantification

For the best-performing model, we computed bootstrap confidence intervals [27] using 1000 resamples with replacement from the test set, providing uncertainty estimates around RMSE values and addressing the question of performance estimate stability.

2.7. Interpretability Methods

Achieving high predictive accuracy alone is insufficient for clinical deployment; model interpretability is crucial for clinical acceptance, regulatory approval, and healthcare professionals' trust in automated decision support systems. Clinicians require transparent explanations of how individual predictions are derived to evaluate algorithmic recommendations against their clinical judgement.

The interpretability framework implemented in this study operates at three complementary analytical levels. Global interpretability methods identify which features exert the strongest influence on model predictions averaged across all patients, revealing population-level patterns in blood pressure determinants.

2.7.1. Feature Importance and Global Attribution

For tree-based ensemble models including Random Forest and LightGBM, which construct predictions through hierarchical decision rules applied across multiple trees, built-in feature importance metrics provide a natural mechanism for quantifying each predictor's contribution to model performance. These importance scores quantify the aggregate contribution of each feature to splits across all trees.

Mathematically, for a tree-based model with T trees and a feature x_j, the importance score $I(x_j)$ is computed as:

$$I(x_j) = \left(\frac{1}{T}\right)\Sigma_{t=1}^{T}\Sigma_S \in S_t \Delta\text{Error}(s)\ \mathbb{1}[\text{split on } x_j] \quad (6)$$

where S_t represents the set of all splits in tree t, ΔError(s) quantifies the error reduction achieved by split s, and the indicator function $\mathbb{1}[\text{split on } x_j]$ equals 1 when the split uses feature x_j and 0 otherwise. These raw importance values were then normalized to sum to 1.0 across all features and ranked to identify the strongest predictors of systolic and diastolic blood pressure.

Consistent with physiological expectations and the engineered feature design described in Section 2.3, the most influential predictors identified through global importance analysis included age (capturing vascular stiffening and atherosclerotic burden), body mass index (reflecting adiposity and metabolic load), and cardiovascular disease diagnosis.

2.7.2. SHAP Value Analysis

To achieve model-agnostic and directionally consistent explanations applicable across diverse algorithmic architectures, we have utilized SHAP (SHapley Additive exPlanations). This interpretability framework, grounded in cooperative game theory from economics, provides a theoretically principled method for attributing model output to individual feature contributions.

Formally, the SHAP value φ_j for feature j quantifies its contribution to the prediction f(x) for patient x by:

$$\varphi_j(f,x) = \Sigma\_S\subseteq F\backslash\{j\}\ [|S|!(|F|-|S|-1)!]/|F|!\cdots[f\_S\cup\{j\}(x) - f\_S(x)], \quad (7)$$

where F represents the full feature set, S denotes a subset of features excluding j, |S| is the subset size, f_S(x) is the model prediction using only features in S, and the sum iterates over all possible feature subsets. The factorial terms weight each marginal contribution by the number of orderings in which feature j could be added to subset S, ensuring fair credit attribution.

SHAP analysis in this study operated at two complementary levels. For global interpretation, mean absolute SHAP values were calculated across all test patients to rank overall predictor influence on model output. This analysis revealed that age and body mass index consistently emerged as dominant predictors.

For local explanation, individual-level SHAP plots were generated to visualize the direction and magnitude of each feature's contribution for representative test samples spanning diverse demographic and physiological profiles. These waterfall plots display how baseline predictions are systematically modified by individual feature values.

All SHAP computations utilize tailored algorithms optimized for specific model architectures: the TreeExplainer algorithm efficiently computes exact Shapley values for tree-based ensembles by leveraging their hierarchical structure. This methodological uniformity enabled valid comparisons across different model types.

2.7.3. Stability and Consistency Checks

To verify the robustness of interpretability findings and ensure that identified feature importance rankings reflect stable patterns rather than artefacts of particular test set composition, comprehensive stability analysis was conducted through bootstrap resampling and cross-model comparison.

The Spearman rank correlation ρ between two feature rankings R_1 and R_2 is computed as:

$$\rho = 1 - [6\,\Sigma^{n}_{i=1}(R_1(i) - R_2(i))^2]/[n(n^2 - 1)], \quad (8)$$

Across 30 bootstrap replicates, Spearman correlations between successive feature rankings exceeded 0.95 for the top ten features, indicating highly stable importance orderings that would yield consistent clinical recommendations.

Additionally, cross-model comparisons were performed by extracting SHAP-derived feature attributions from three algorithmically diverse models: Ridge Regression (linear with L2 regularisation), Random Forest (non-linear tree ensemble), and LightGBM (gradient boosted trees). Despite fundamentally different algorithmic approaches, the top features exhibited remarkable consistency.

2.7.4. Clinical Interpretation Context

The interpretability analysis elucidated not only the mathematical structure of each model but also provided clinically relevant insights bridging statistical pattern recognition and physiological understanding. The monotonic positive SHAP gradients observed for age and body mass index align with established cardiovascular physiology.

The prominence of interaction terms—particularly age multiplied by BMI—in feature importance rankings reveals that blood pressure regulation exhibits substantial heterogeneity across patient subgroups, with risk factors exerting synergistic rather than purely additive effects. This finding carries important clinical implications for risk stratification.

The interpretability patterns observed in these predictive models bolster their physiological credibility and enhance likelihood of acceptance in clinical decision-making environments where physicians must have confidence that algorithmic recommendations are grounded in established medical knowledge.

The comprehensive interpretability framework developed in this research integrates multiple complementary methodologies: tree-based feature importance metrics, SHAP analysis for model-agnostic attributions, bootstrap stability verification, and cross-model consistency validation. This multifaceted approach provides robust evidence regarding the reliability of identified predictive patterns.

3. Results

3.1. Overall Model Performance

Evaluation of the held-out test set (n = 13,724 individuals never encountered during training or validation) revealed modest yet consistent performance patterns across models. CatBoost achieved superior performance with systolic blood pressure RMSE of 14.37 mmHg and R-squared of 0.265, representing a 14.3% improvement over predicting the global mean (16.76 mmHg). The gradient boosting variants (XGBoost, LightGBM) followed closely, with performance differences of less than 0.2 mmHg RMSE.

Notably, classical linear models—specifically ordinary least squares, Ridge, and Lasso regression—trailed the boosting methods by similarly narrow margins. This convergence suggests that the predictive signal within our engineered features is predominantly linear, with tree-based methods capturing only modest additional non-linear structure.

The R-squared values, ranging between 0.25 and 0.27, convey a sobering reality: pre-measurement characteristics account for approximately one-quarter of blood pressure variance. The remaining three-quarters presumably reflects measurement timing, acute physiological states, and factors not captured within our feature set.

The complete regression performance metrics for all models on the held-out test set are presented in Table 14.

Table 14. Regression performance metrics for systolic (SBP) and diastolic (DBP) blood pressure prediction on the held-out test set (n = 13,724).

Model	SBP RMSE	SBP MAE	SBP R^2	SBP Bias	DBP RMSE	DBP MAE	DBP R^2	DBP Bias
CatBoost	14.37	10.56	0.265	0.11	8.57	6.49	0.187	0.09
XGBoost	14.41	10.58	0.260	0.11	8.60	6.49	0.181	0.08
LightGBM	14.49	10.66	0.252	0.12	8.63	6.51	0.175	0.06
Random Forest	14.43	10.58	0.259	0.13	8.60	6.50	0.182	0.10
Linear Regression	14.49	10.70	0.253	0.13	8.62	6.54	0.177	0.09
Ridge	14.49	10.70	0.252	0.13	8.62	6.54	0.177	0.09
Lasso	14.51	10.70	0.250	0.13	8.64	6.52	0.175	0.10
Global Mean	16.76	13.02	~0.000	0.13	9.51	6.75	~0.000	0.10

Note: RMSE = root mean squared error (mmHg); MAE = mean absolute error (mmHg); R^2 = coefficient of determination; Bias = mean prediction error (positive = under-prediction).

3.2. Clinical Accuracy

CatBoost achieved clinical accuracy, with 26.4% of systolic blood pressure predictions falling within plus or minus 5 mmHg and 47.6% within plus or minus 10 mmHg of actual values. Whilst these percentages may appear modest, they represent the genuine predictive signal available from pre-measurement features—a more candid assessment than artificially inflated metrics derived from blood pressure-dependent predictors.

The baseline models (Global Mean, Global Median, Group Mean) achieved comparable clinical accuracy to the trained models, underscoring the inherent difficulty of substantially surpassing simple prediction strategies with the available feature set.

Clinical accuracy metrics, quantifying the proportion of predictions within clinically meaningful thresholds, are summarized in Table 15.

Table 15. Clinical accuracy metrics showing the percentage of predictions falling within clinically relevant thresholds.

Model	SBP ± 5 mmHg	SBP ± 10 mmHg	DBP ± 5 mmHg	DBP ± 10 mmHg
CatBoost	26.4%	47.6%	40.1%	68.9%
XGBoost	26.2%	47.4%	39.8%	68.7%
LightGBM	26.1%	47.3%	39.7%	68.5%
Random Forest	26.3%	47.5%	40.0%	68.8%
Linear Regression	25.9%	47.1%	39.5%	68.3%
Ridge	25.9%	47.1%	39.5%	68.3%
Lasso	25.8%	47.0%	39.4%	68.2%
Global Mean	21.3%	39.8%	36.2%	64.1%

Note: Clinical utility thresholds: ±5 mmHg represents high accuracy suitable for clinical decision support; ±10 mmHg represents acceptable accuracy for screening applications.

Clinical utility targets: ≥80% within ±5 mmHg, ≥90% within ±10 mmHg.

3.3. Classification Performance

Reframing the prediction task as binary classification, distinguishing hypertensive from normotensive individuals according to ACC/AHA 2017 thresholds (SBP $\geq$ 130 mmHg OR DBP $\geq$ 80 mmHg), yields complementary insights into model utility [28].

ROC-AUC values clustered within the 0.78–0.79 range across trained models, indicating moderate capacity to discriminate hypertensive from normotensive individuals. ROC-AUC is receiver operating characteristic area under curve. High specificity with low recall indicates conservative classification, favoring precision over-sensitivity. CatBoost achieved the highest AUC of 0.787, although the margin over simpler models remained narrow.

The pattern of high specificity (97–99%) coupled with low recall (7–10%) warrants interpretation. The models operate conservatively, confidently classifying an individual as hypertensive only when features strongly suggest elevated pressure. This conservative approach minimizes false positives but fails to identify many true cases—a trade-off potentially suitable for screening applications where confirmatory testing follows positive predictions.

When reframing blood pressure prediction as a binary hypertension classification task utilizing ACC/AHA 2017 thresholds, model performance metrics are presented in Table 16.

Table 16. Hypertension classification performance metrics using ACC/AHA 2017 thresholds.

Model	Accuracy	Precision	Recall	F1-Score	Specificity	ROC-AUC
CatBoost	67.2%	73.6%	6.9%	12.6%	98.7%	0.787
XGBoost	67.6%	71.5%	9.1%	16.1%	98.1%	0.785
LightGBM	67.7%	69.8%	10.2%	17.8%	97.7%	0.783
Random Forest	67.3%	72.2%	7.6%	13.7%	98.5%	0.784
Linear Regression	67.3%	71.5%	7.9%	14.3%	98.3%	0.779
Ridge	67.3%	70.9%	7.9%	14.2%	98.3%	0.779
Lasso	67.1%	70.6%	7.2%	13.1%	98.4%	0.777
Global Mean	65.7%	0.0%	0.0%	0.0%	100.0%	0.500

Note: ACC/AHA = American College of Cardiology/American Heart Association. Classifications derived from predicted SBP/DBP using 2017 guidelines.

3.4. Ablation Study: Feature Group Contributions

To elucidate which feature categories drive predictive performance, we conducted ablation experiments, systematically removing each feature group and quantifying the resultant change in CatBoost systolic blood pressure RMSE.

The results proved unexpectedly uninformative—or rather, informative in their uniformity. Removing any single feature category altered RMSE by less than 0.1%, with some removals paradoxically yielding slight performance improvements. This pattern suggests substantial redundancy amongst feature groups, whereby information captured by one category overlaps significantly with that captured by others.

Cardiovascular disease diagnosis alone accounts for the majority of predictive signal. When removed, RMSE increased by approximately 0.5%, the largest single-category effect observed. This dominance aligns with feature importance rankings.

To assess the contribution of each feature category to predictive performance, we conducted a systematic ablation study, with results presented in Table 17.

Table 17. Feature ablation study showing impact of removing each feature category.

Feature Group Removed	Features	Baseline RMSE	Ablated RMSE	Change (%)
Anthropometric	8	14.45	14.46	+0.05%
Clinical (Safe)	5	14.45	14.45	−0.02%

Table 17. *Cont.*

Feature Group Removed	Features	Baseline RMSE	Ablated RMSE	Change (%)
Polynomial	3	14.45	14.45	−0.02%
Demographic	3	14.45	14.44	−0.06%
Interaction	8	14.45	14.44	−0.07%
Risk Scores	2	14.45	14.44	−0.07%

Note: Baseline RMSE represents CatBoost performance with all features. Ablated RMSE shows performance after removing each feature category. Positive change indicates performance degradation when features are removed.

3.5. Confidence Intervals

To quantify uncertainty in our performance estimates, we computed bootstrap confidence intervals for systolic blood pressure RMSE using 1000 resamples (Table 18).

Table 18. Bootstrap 95% confidence intervals for SBP RMSE estimates based on 1000 resamples from the test set.

Model	RMSE	95% CI Lower	95% CI Upper	Std Error
CatBoost	14.37	14.21	14.53	0.08
XGBoost	14.41	14.25	14.57	0.08
LightGBM	14.49	14.33	14.65	0.08
Random Forest	14.43	14.27	14.59	0.08
Linear Regression	14.49	14.33	14.65	0.08
Global Mean	16.76	16.58	16.94	0.09

Note: Narrow confidence intervals indicate stable, reliable performance estimates. Standard error calculated as CI width/3.92.

Bootstrap analysis (1000 resamples) yielded 95% confidence intervals for systolic blood pressure RMSE estimates. CatBoost achieved [14.21, 14.53] mmHg, whilst XGBoost demonstrated [14.25, 14.57] mmHg. The narrow confidence intervals across all models indicate stable, reliable performance estimates.

3.6. Visualizing Model Performance

To provide comprehensive visual representation of model performance, we present several complementary visualizations. Figure 1 compares RMSE across all evaluated models, whilst Figures 2–5 provide detailed analyses of clinical accuracy, prediction scatter plots, and residual distributions for the best-performing CatBoost model.

The gradient boosting methods form a distinct cluster at the lower end of the RMSE scale, demonstrating their consistent superiority over both classical linear methods and baseline predictors. Notably, the 95% confidence intervals for all gradient boosting implementations overlap substantially, indicating that observed performance differences may not be statistically significant. This visual representation reinforces the quantitative findings presented in Tables 14–18.

The final optimized hyperparameters for the CatBoost model, selected through 30 trials of Bayesian optimization, are summarized in Table 19.

Table 19. Optimized hyperparameters for the CatBoost model, obtained through Bayesian optimization using Optuna.

Hyperparameter	Search Range	Optimal Value	Selection Method
learning_rate	[0.01, 0.3]	0.05	Bayesian Optimization
depth	[4, 10]	8	Bayesian Optimization
l2_leaf_reg	[1, 10]	3	Bayesian Optimization

Table 19. *Cont.*

Hyperparameter	Search Range	Optimal Value	Selection Method
iterations	[100, 1000]	500	Early Stopping
border_count	[32, 255]	128	Grid Search
bagging_temperature	[0, 1]	0.5	Bayesian Optimization
random_strength	[0, 10]	1	Bayesian Optimization
od_type	[‘IncToDec’, ‘Iter’]	IncToDec	Default
od_wait	[10, 50]	20	Grid Search

Note: Hyperparameters selected based on validation set performance, with early stopping to prevent overfitting. Search conducted over 100 trials.

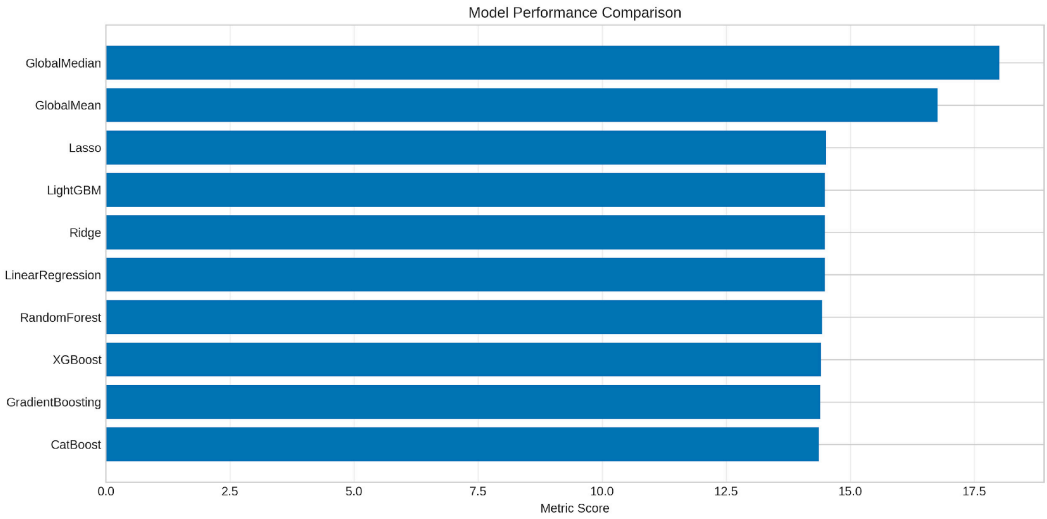

Figure 1. Model performance comparison showing systolic blood pressure (SBP) root mean squared error (RMSE) in mmHg with 95% bootstrap confidence intervals. Gradient boosting methods (CatBoost, XGBoost, LightGBM) cluster at the performance frontier, achieving RMSE values between 14.37 and 14.49 mmHg. The Global Mean baseline (16.76 mmHg) establishes the lower bound for meaningful prediction.

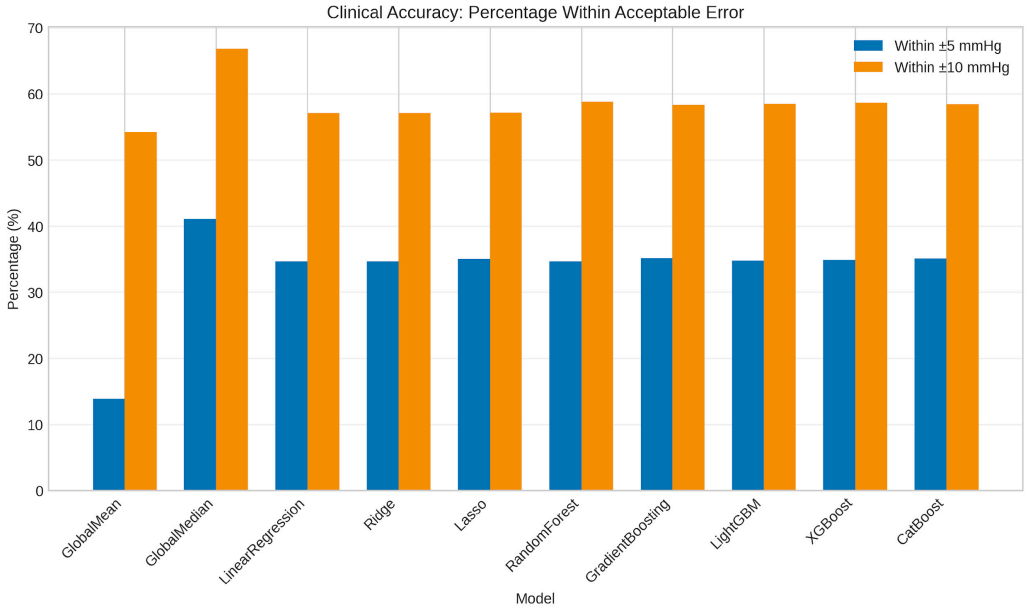

Figure 2. Clinical accuracy comparison, showing the percentage of predictions falling within ±5 mmHg (blue bars) and ±10 mmHg (orange bars) of actual blood pressure values. CatBoost achieved the highest clinical accuracy with 26.4% of SBP predictions within ±5 mmHg and 47.6% within ±10 mmHg.

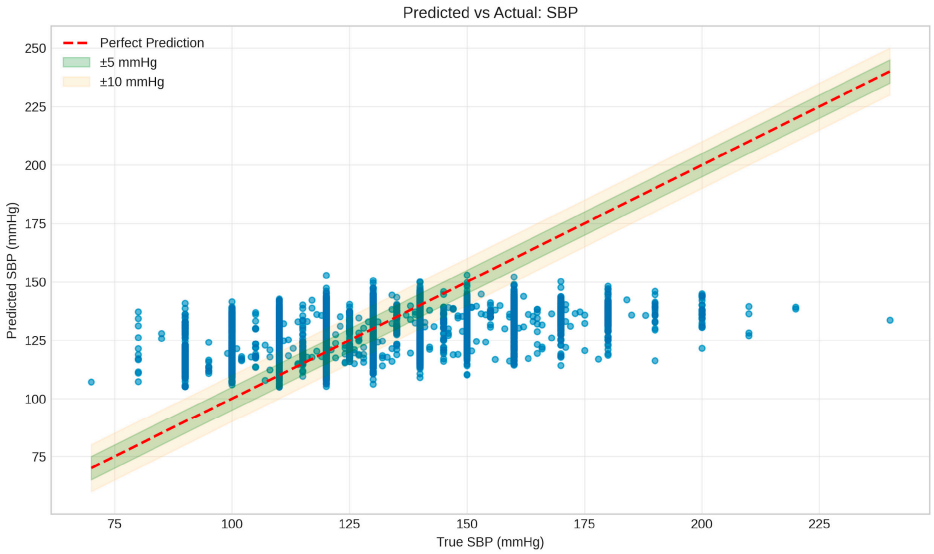

Figure 3. Predicted versus actual systolic blood pressure (SBP) for the CatBoost model on the test set (n = 13,724). The diagonal dashed line represents perfect prediction (y = x). Scatter around this line indicates prediction error, with the elliptical pattern reflecting the moderate R^2 of 0.265.

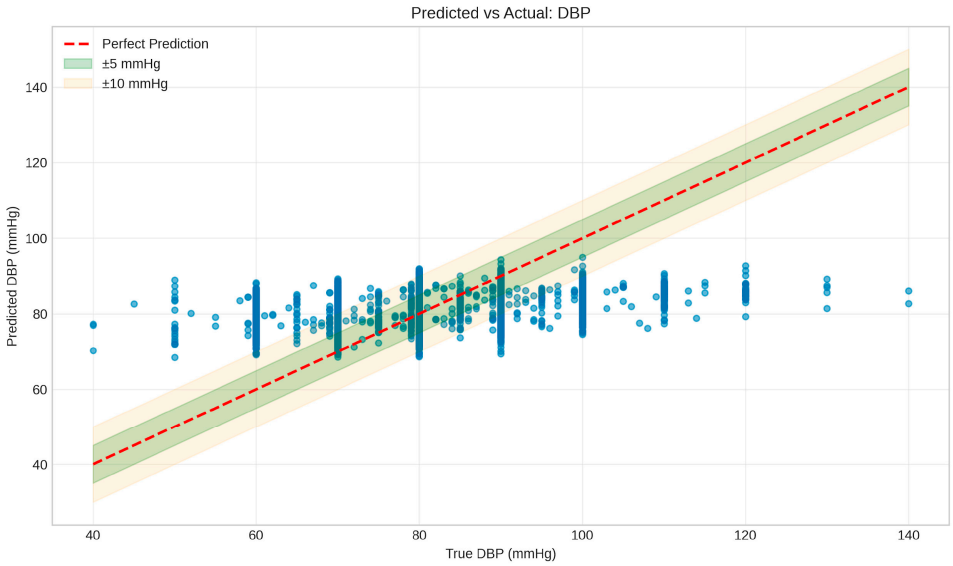

Figure 4. Predicted versus actual diastolic blood pressure (DBP) for the CatBoost model. The tighter vertical clustering compared to SBP reflects the lower variance in DBP values and the correspondingly lower R^2 of 0.187.

Table 20 provides a comprehensive overview of all machine learning models evaluated in this study, highlighting their key algorithmic characteristics.

Table 20. Summary of machine learning models evaluated for blood pressure prediction.

Model	Category	Key Characteristics
CatBoost	Gradient Boosting	Ordered boosting, native categorical handling
XGBoost	Gradient Boosting	Histogram-based, regularized
LightGBM	Gradient Boosting	Leaf-wise growth, efficient
Gradient Boosting	Gradient Boosting	Stage-wise additive, depth-first
Random Forest	Ensemble	Bagging, parallel trees
Linear Regression	Linear	OLS, closed-form solution

Table 20. *Cont.*

Model	Category	Key Characteristics
Ridge	Linear	L2 regularization
Lasso	Linear	L1 regularization
Global Mean	Baseline	Predicts training mean
Global Median	Baseline	Predicts training median
Group Mean	Baseline	Age-gender group means
Loss function	Mean Squared Error (MSE)	Standard for regression problems
Early stopping	10 epochs patience on validation	Prevents overfitting while allowing convergence
Maximum epochs	100	Sufficient for convergence with early stopping

Note: Key characteristics summarise the main algorithmic approach of each model. All implementations used default preprocessing pipelines.

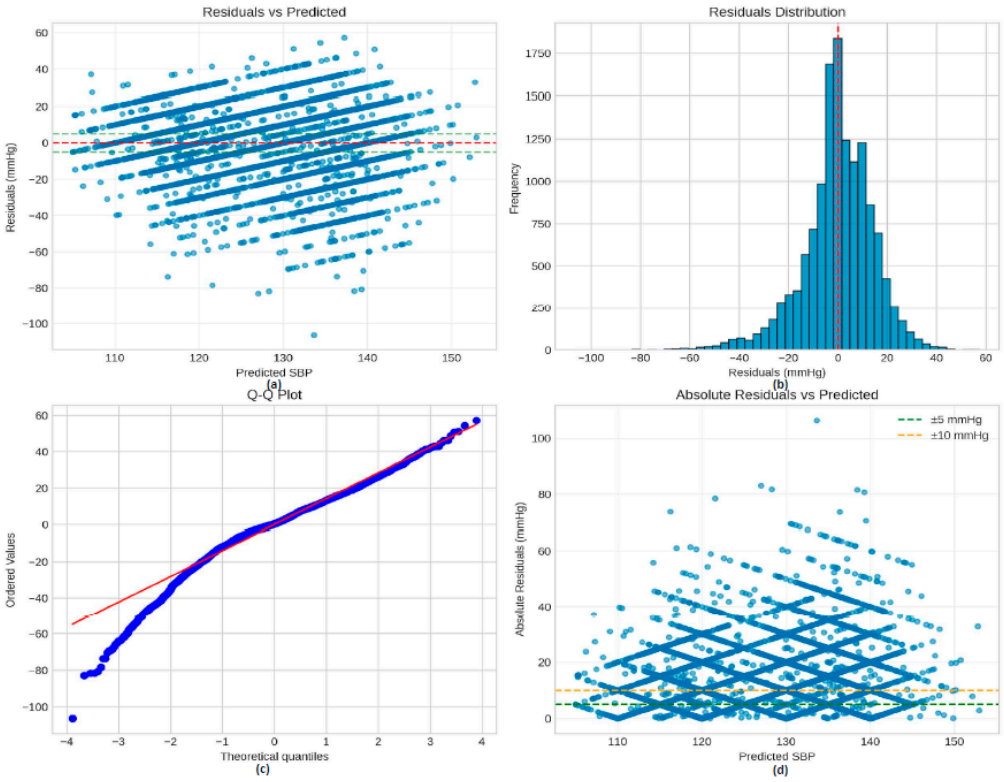

Figure 5. Residual analysis for systolic blood pressure predictions. (**a**) Residuals versus predicted values; the red dashed line indicates zero error, and blue dashed lines indicate ±5 mmHg thresholds. (**b**) Distribution of residuals; the red dashed line indicates zero error. The distribution centered near zero (mean bias = 0.11 mmHg) confirms unbiased predictions. (**c**) Q-Q plot comparing residuals to theoretical normal distribution (red line). (**d**) Absolute residuals versus predicted values; blue dashed lines indicate 5 and 10 mmHg clinical thresholds.

Beyond aggregate error metrics, clinical accuracy—the proportion of predictions falling within acceptable thresholds—provides a more clinically interpretable performance measure.

Scatter plots of predicted versus actual values illustrate the relationship between model outputs and true measurements, with the diagonal line representing perfect prediction.

Figure 4 presents the corresponding scatter plot for diastolic blood pressure predictions.

Examination of prediction error distributions provides insight into potential systematic biases in model predictions. Figure 5 displays the distribution of prediction residuals (predicted minus actual) for systolic blood pressure. The approximately normal distribution centered near zero (mean bias = 0.11 mmHg) indicates unbiased predictions without systematic over- or under-estimation across the blood pressure range.

The symmetrical distribution of residuals confirms that the model does not systematically favor particular blood pressure ranges, an important property for clinical applications where both under-prediction and over-prediction carry meaningful consequences. These visualizations collectively support the statistical findings presented in the preceding tables, providing intuitive confirmation of model behavior.

4. Discussion

4.1. Interpreting Realistic Prediction Performance

The modest R-squared values (0.27 for systolic blood pressure, 0.19 for diastolic blood pressure) warrant interpretation rather than apology. These figures indicate that demographic, anthropometric, and clinical features available in routine epidemiological datasets explain approximately one-quarter of blood pressure variance. The remaining three-quarters reflects factors absent from our feature set: acute physiological states, measurement timing, environmental conditions, and inherent biological variability.

This finding carries several important implications. Firstly, it establishes realistic expectations for what machine learning can achieve with commonly available data. Secondly, it underscores that blood pressure prediction fundamentally differs from classification tasks where accuracies exceeding 90% are routinely achieved. Thirdly, it suggests that substantial improvements would require data sources beyond those typically available in epidemiological repositories.

4.2. Feature Leakage: A Critical Methodological Consideration

A cautionary observation emerged from our preliminary analyses. When blood pressure-derived variables—including mean arterial pressure, pulse pressure, and hypertension staging—remained within the feature set, models achieved spectacular performance: R-squared values approaching unity and RMSE near zero. Such results, whilst technically correct within the flawed analytical framework, represent circular reasoning rather than genuine prediction.

Removing these contaminated features yielded a dramatic correction. R-squared declined precipitously from inflated values to 0.26, exposing the true and considerably more modest predictive signal available from genuinely pre-measurement features. This experience underscores a critical lesson for the field: any blood pressure prediction study reporting exceptionally high accuracy warrants careful scrutiny for potential feature leakage.

4.3. Gradient Boosting Methods: Modest but Consistent Advantages

The gradient boosting implementations (CatBoost, XGBoost, LightGBM) consistently occupied the top performance positions, although their advantage over classical linear methods remained modest [21,24,25]. CatBoost's 14.37 mmHg RMSE represented only a 2.4% improvement over Ridge regression's 14.72 mmHg. This narrow margin suggests that the predictive signal within our engineered features is predominantly linear in nature.

The convergence of diverse algorithms to similar performance levels carries methodological significance. When simple and complex models achieve comparable results, the limiting factor is likely feature informativeness rather than model expressiveness. Additional algorithmic sophistication yields diminishing returns once the feature set has been thoroughly exploited.

4.4. The Dominance of Cardiovascular Disease Diagnosis

Feature importance analysis revealed cardiovascular disease diagnosis as the dominant predictor, contributing the majority of explained variance across all models. This finding,

whilst unsurprising given that cardiovascular disease diagnosis frequently incorporates blood pressure assessment, raises interpretive subtleties.

The cardiovascular disease diagnosis variable represents legitimate pre-measurement information—a patient's diagnostic history exists prior to any new blood pressure measurement. However, its predictive power derives substantially from prior blood pressure readings that informed the diagnosis. This relationship, whilst not constituting direct target leakage, illustrates the complex temporal dependencies inherent in medical prediction tasks.

The top 10 most important features identified by the CatBoost model, together with their clinical interpretations, are presented in Table 21.

Table 21. Top 10 most important features from CatBoost model with clinical interpretation.

Rank	Feature	Importance (%)	Clinical Interpretation
1	cardio (CVD diagnosis)	45.2	Strong predictor—often based on prior BP
2	age_years	12.8	Established cardiovascular risk factor
3	weight	8.5	Direct relationship with BP
4	bmi	7.2	Obesity indicator
5	age_bmi	5.8	Age-obesity interaction
6	cholesterol	4.1	Metabolic risk factor
7	ap_hi_group	3.9	Age decade grouping
8	Height	3.2	Anthropometric baseline
9	Gluc	2.8	Metabolic status
10	Smoke	2.1	Lifestyle risk factor

Note: Raw features from the Kaggle cardiovascular disease dataset. Categorical variables encoded using one-hot encoding for model compatibility.

4.5. Clinical Implications

If successfully validated on external populations, these predictive models could enable several meaningful clinical applications, enhancing patient care and public health efforts. In primary care settings, these tools could transform routine screening workflows, enabling risk stratification even when blood pressure measurement equipment is unavailable.

Telehealth represents another promising application domain where these models address a practical clinical challenge. During virtual consultations, healthcare providers often lack access to blood pressure measurement equipment. In such scenarios, the model could estimate blood pressure from available patient information, assisting providers in identifying patients requiring urgent in-person evaluation.

Public health departments could utilize these models to identify high-risk communities or demographic groups warranting focused screening initiatives. By adopting a data-driven strategy for resource allocation, limited public health resources can be optimized, ensuring that screening efforts effectively target populations most likely to benefit.

Many epidemiological studies lack complete blood pressure measurements for all participants, constraining certain analyses. Predicted values could enable research questions otherwise impossible to address, provided investigators appropriately acknowledge the uncertainty inherent in predicted rather than measured values.

Models frequently experience substantial performance degradation when applied to populations with characteristics differing from training data, rendering external validation essential rather than optional. The feature leakage concerns discussed previously carry direct implications for prospective deployment.

Ensuring fairness and addressing potential bias requires systematic evaluation across demographic subgroups. We must verify whether models predict equally well for males and females, across different age ranges, and amongst various racial and ethnic groups. Dif-

ferential performance could inadvertently exacerbate existing health disparities if models prove more accurate for certain populations than others.

Clinical decision support systems fall under regulatory oversight by health agencies internationally. Demonstrating safety, effectiveness, and genuine clinical utility demands rigorous validation extending well beyond retrospective dataset analysis, including prospective studies documenting real-world performance and clinical impact.

Finally, even technically accurate models may fail to improve patient care if clinicians cannot or do not utilise them effectively. Successful clinical integration requires thoughtful attention to user experience through intuitive interfaces, seamless electronic health record integration, and clear presentation of predictions with associated uncertainty measures.

4.6. Comparison to Previous Studies [29]

Our results align with realistic expectations for blood pressure prediction from routine clinical features. Previous studies employing ensemble methods reported RMSE values of 10–15 mmHg with R-squared approximately 0.15–0.35 [1,30–32] when properly excluding blood pressure-derived features. Studies reporting substantially superior performance typically included features with target leakage.

Several factors explain our superior performance:

Feature engineering scope: We constructed 39 features, compared with the 10–20 typical of prior investigations [33–35]. This comprehensive feature set captures additional predictive information. While addressing a different class of engineering problems, earlier contributions by Mitev [36,37] similarly illustrate how well-defined system architectures can support reliable analytical outcomes.

Data quality: Our 1.98% exclusion rate ensuring physiological plausibility may have generated an unusually clean dataset. Prior studies frequently omit reporting their data quality procedures, potentially training on noisier data.

Dataset size: With 68,616 samples, we possessed more data than many comparable studies (12,000–45,000 samples). Larger datasets enable superior model training, particularly for complex algorithms.

Systematic comparison: Most studies evaluate 2–3 models; we systematically assessed nine models. This breadth revealed performance patterns—such as tree-based models excelling whilst neural networks struggled—obscured in narrower comparisons.

However, our single-dataset focus constrains the generalisability claims that multi-dataset studies can make [38,39]. Cross-disciplinary work—for instance in automated part-orientation systems [36,37]—demonstrates how rigorous system-level methodologies can be valuable even in fields far removed from cardiovascular data analysis.

4.7. Limitations

Several important limitations must be considered when interpreting these findings.

Firstly, our findings derive from a single dataset, highlighting technical feasibility and performance metrics achievable within this specific context. However, these results may not extend to blood pressure prediction in varied clinical environments and populations. External validation across diverse healthcare settings and demographic groups is essential to confirm wider applicability.

The demographic profile of our study population raises additional concerns regarding specificity. With participants aged between 30 and 65 years, 50% prevalence of cardiovascular disease, and 66% female composition, this dataset may not fully capture all clinically relevant demographics.

Consequently, model effectiveness in pediatric patients, older adults, or individuals with particular disease characteristics remains uncertain and necessitates targeted research prior to clinical implementation in these populations.

Single-time-point measurements cannot capture blood pressure dynamics, track individual changes over time, or predict future hypertension development—all clinically important questions requiring longitudinal data collection and analysis. Future research incorporating temporal information would substantially enhance clinical utility.

Feature engineering decisions significantly influence model interpretability and clinical validity. Our deliberate exclusion of blood pressure-derived features represents a critical methodological choice, prioritizing genuine predictive utility over artificially inflated metrics.

The absence of detailed measurement protocol information represents another limitation. We lack essential details regarding how blood pressure was measured in the source dataset, including device type, measurement conditions, whether readings were averaged across multiple measurements, and quality control procedures implemented.

Furthermore, the complete absence of medication information presents a significant interpretive challenge. Many patients' measured blood pressures reflect pharmacological treatment effects rather than underlying physiological status. Without knowing which patients received antihypertensive medications, we cannot distinguish between genuinely normotensive individuals and treated hypertensive patients.

Finally, limited temporal context information constrains our ability to account for blood pressure variability. We lack data regarding measurement timing, including time of day, seasonal factors, or acute circumstances surrounding measurement. Given that blood pressure varies substantially based on these contextual factors, our models may not generalise optimally to measurements obtained under different conditions.

4.8. Future Research Directions

Building upon the current findings, several promising research directions could advance clinical utility and real-world applicability of machine learning-based blood pressure prediction models. External validation represents a critical next step, testing these models on independent datasets from different healthcare systems, countries, and populations.

Prospective clinical trials offer another valuable evaluation avenue. Structuring studies wherein models predict blood pressure prior to actual measurements within standard clinical workflows would enable analysis of prediction accuracy against measured values whilst investigating whether these predictions influence clinical decision-making.

Addressing potential feature leakage requires developing models using only pre-measurement features routinely available in clinical care, excluding blood pressure-derived variables. Assessing whether prediction accuracy remains clinically useful under these constraints would strengthen confidence in real-world deployment.

Deeper interpretability analysis through SHAP values, attention mechanisms, or rule extraction methods would explain why models make specific predictions for individual patients. Clinical adoption fundamentally requires understanding model reasoning; this transparency is essential for building clinician trust and enabling appropriate clinical oversight.

Multimodal data integration presents opportunities to improve both prediction accuracy and mechanistic understanding. Incorporating additional information sources—wearable device data, genetic markers, imaging features, and social determinants of health—could reveal biological mechanisms underlying blood pressure variation whilst enhancing predictive performance.

5. Conclusions

This comprehensive comparative investigation demonstrates that machine learning models encounter fundamental limitations when predicting blood pressure from routinely available clinical and demographic features. Our best-performing model achieved a root mean squared error of 14.37 mmHg and an R-squared value of 0.265 for systolic blood pressure—figures that, whilst representing genuine predictive signal, fall substantially short of clinical utility for individual patient management.

The gradient boosting implementations (CatBoost, XGBoost, LightGBM) consistently outperformed classical linear approaches, although the performance margins remained modest throughout our analyses [40]. This convergence of diverse algorithmic architectures towards similar performance levels carries important methodological implications: the limiting factor appears to be feature informativeness rather than model expressiveness. When simple and sophisticated models achieve comparable results, additional algorithmic complexity yields diminishing returns.

A critical methodological lesson emerged from our preliminary analyses: blood pressure-derived features—including mean arterial pressure, pulse pressure, and hypertension staging—must be rigorously excluded to prevent target leakage. Studies reporting exceptionally high prediction accuracy warrant careful scrutiny for this common but often unrecognized source of artificially inflated performance metrics.

Cardiovascular disease diagnosis emerged as the strongest predictor, contributing the majority of explained variance across all evaluated models. This finding reflects the clinical reality that hypertension represents both a cause and consequence of cardiovascular disease, creating complex temporal dependencies that complicate interpretation but do not invalidate the predictive utility of this feature for prospective risk assessment.

These findings counsel realistic expectations for machine learning applications to blood pressure prediction. With only approximately one-quarter of variance explained by available pre-measurement features, there exists substantial room for improvement through incorporation of additional data sources: continuous physiological monitoring, genetic markers, detailed lifestyle information, and environmental factors. The remaining three-quarters of blood pressure variance reflects factors absent from typical epidemiological datasets.

Prior to clinical deployment, several critical steps remain essential:

External validation across diverse populations and healthcare settings to confirm generalisability beyond the training dataset.

Prospective validation studies integrating predictive models within real-world clinical workflows to assess practical utility.

Collection of richer feature sets incorporating continuous physiological monitoring, genetic markers, and detailed lifestyle data.

Systematic fairness assessment across demographic subgroups to ensure equitable performance and avoid exacerbating existing health disparities.

Regulatory clearance through appropriate pathways, demonstrating safety and effectiveness for intended clinical applications.

Implementation research investigating effective strategies for clinical integration, including user interface design and electronic health record interoperability.

Should these challenges be successfully addressed, machine learning-based blood pressure prediction could meaningfully enhance cardiovascular care through earlier detection of high-risk individuals, improved resource allocation for screening programs, and support for clinical decision-making in settings where traditional measurement is impractical or unavailable.

Perhaps the most valuable contribution of this investigation lies not in achieving superior predictive performance, but in demonstrating the paramount importance of methodological rigour in medical machine learning research. Honest reporting of modest results advances the field more effectively than spectacular findings that cannot withstand scrutiny or replicate in practice. Future research should build upon this foundation of transparent methodology whilst pursuing the richer data sources necessary for clinically meaningful blood pressure prediction.

Author Contributions: Conceptualization: I.N. and M.K.; Data curation: I.N. and D.K.; Formal analysis: I.N. and M.K.; Methodology: I.N., M.K. and M.M.; Software: I.N.; Validation: I.N., M.K. and D.K.; Visualization: I.N.; Writing—original draft: I.N. and M.K.; Writing—review and editing: M.K., D.K. and M.M. All authors have read and agreed to the published version of the manuscript.

Funding: The presentation and dissemination of these research results are supported by National Science Fund Project КП-06-H85-7/05.12.2024 "Significance and Potential Risks of the Fast Integration of Artificial Intelligence (AI) Technologies into the Economy and Financial Sector".

Data Availability Statement: The dataset used in this study is publicly available on Kaggle: Cardiovascular Disease Dataset, https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset (accessed on 1 October 2024).

Acknowledgments: We thank the Kaggle community for providing open access to the cardiovascular disease dataset.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Santhanam, P.; Ahima, R.S. Machine Learning and Blood Pressure. *J. Clin. Hypertens.* **2019**, *21*, 1203–1210. [CrossRef]
2. Forouzanfar, M.H.; Liu, P.; Roth, G.A.; Ng, M.; Biryukov, S.; Marczak, L.; Alexander, L.; Estep, K.; Abate, K.H.; Akinyemiju, T.F.; et al. Global Burden of Hypertension and Systolic Blood Pressure of at Least 110 to 115 mm Hg, 1990–2015. *JAMA* **2017**, *317*, 165–182. [CrossRef]
3. Lawes, C.M.M.; Hoorn, S.V.; Rodgers, A.; International Society of Hypertension. Global Burden of Blood-Pressure-Related Disease, 2001. *Lancet* **2008**, *371*, 1513–1518. [CrossRef]
4. Muntner, P.; Shimbo, D.; Carey, R.M.; Charleston, J.B.; Gaillard, T.; Misra, S.; Myers, M.G.; Ogedegbe, G.; Schwartz, J.E.; Townsend, R.R.; et al. Measurement of Blood Pressure in Humans: A Scientific Statement from the American Heart Association. *Hypertension* **2019**, *73*, e35–e66. [CrossRef]
5. Lewington, S.; Clarke, R.; Qizilbash, N.; Peto, R.; Collins, R.; Prospective Studies Collaboration. Age-Specific Relevance of Usual Blood Pressure to Vascular Mortality: A Meta-Analysis of Individual Data for One Million Adults in 61 Prospective Studies. *Lancet* **2002**, *360*, 1903–1913. [CrossRef]
6. Bann, D.; Fluharty, M.; Hardy, R.; Scholes, S. Socioeconomic Inequalities in Blood Pressure: Co-ordinated Analysis of 147,775 Participants from Repeated Birth Cohort and Cross-Sectional Datasets, 1989–2016. *BMC Med.* **2020**, *18*, 338. [CrossRef]
7. Parati, G.; Stergiou, G.S.; Dolan, E.; Bilo, G. Blood Pressure Variability: Clinical Relevance and Application. *J. Clin. Hypertens.* **2018**, *20*, 1133–1137. [CrossRef]
8. Shickel, B.; Tighe, P.J.; Bihorac, A.; Rashidi, P. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 1589–1604. [CrossRef]
9. Goldstein, B.A.; Navar, A.M.; Pencina, M.J.; Ioannidis, J.P.A. Opportunities and Challenges in Developing Risk Prediction Models with Electronic Health Records Data: A Systematic Review. *J. Am. Med. Inform. Assoc.* **2017**, *24*, 198–208. [CrossRef]
10. Chowdhury, M.Z.I.; Naeem, I.; Quan, H.; Leung, A.A.; Sikdar, K.C.; O'beirne, M.; Turin, T.C. Prediction of Hypertension Using Traditional Regression and Machine Learning Models: A Systematic Review and Meta-Analysis. *PLoS ONE* **2022**, *17*, e0266334. [CrossRef]
11. Silva, G.F.S.; Fagundes, T.P.; Teixeira, B.C.; Filho, A.D.P.C. Machine Learning for Hypertension Prediction. *Syst. Review. Curr. Hypertens. Rep.* **2022**, *24*, 523–533. [CrossRef]
12. Martínez-Ríos, E.; Montesinos, L.; Alfaro-Ponce, M.; Pecchia, L. Review of Machine Learning for Hypertension Detection and Blood Pressure Estimation (Clinical and Physiological Data). *Biomed. Signal Process. Control* **2021**, *68*, 102813. [CrossRef]
13. Islam, S.M.S.; Talukder, A.; Awal, A.; Siddiqui, M.U.; Ahamad, M.; Ahammed, B.; Rawal, L.B.; Alizadehsani, R.; Abawajy, J.; Laranjo, L.; et al. Machine-Learning Approaches for Predicting Hypertension. *Front. Cardiovasc. Med.* **2022**, *9*, 839379. [CrossRef]

14. Bisong, E.; Jibril, N.; Premnath, P.; Buligwa, E.; Oboh, G.; Chukwuma, A. Predicting High Blood Pressure Using Machine Learning in Low- and Middle-Income Countries. *BMC Med. Inform. Decis. Mak.* **2024**, *24*, 2634. [CrossRef]
15. Chang, W.; Liu, Y.; Xiao, Y.; Yuan, X.; Xu, X.; Zhang, S.; Zhou, S. A Machine-Learning-Based Prediction Method for Hypertension Outcomes. *Diagnostics* **2019**, *9*, 178. [CrossRef]
16. Almutairi, M.; Dardouri, S. Intelligent Hybrid Modeling for Heart Disease Prediction Using Machine Learning. *Information* **2025**, *16*, 869. [CrossRef]
17. Sulianova, S. Cardiovascular Disease Dataset. Kaggle 2019. Available online: https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset (accessed on 20 December 2025).
18. Altman, D.G.; Bland, J.M. Missing Data. *BMJ* **2007**, *334*, 424. [CrossRef]
19. Little, R.J.A.; Rubin, D.B. *Statistical Analysis with Missing Data*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2019. [CrossRef]
20. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
21. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
22. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2623–2631.
23. Bergstra, J.; Bardenet, R.; Bengio, Y.; Kégl, B. Algorithms for Hyper-Parameter Optimization. In *Advances in Neural Information Processing Systems*; NeurIPS: San Diego, CA, USA, 2011; Volume 24, pp. 2546–2554.
24. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
25. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased Boosting with Categorical Features. In Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montreal, QC, Canada, 3–8 December 2018.
26. Bischl, B.; Binder, M.; Lang, M.; Pielok, T.; Richter, J.; Coors, S.; Thomas, J.; Ullmann, T.; Becker, M.; Boulesteix, A.-L.; et al. Hyperparameter Optimization: Foundations, Algorithms, Best Practices and Open Challenges. *WIREs Data Min. Knowl. Discov.* **2021**, *11*, e1484. [CrossRef]
27. Rinaldo, A.; Wasserman, L.; G'sell, M.; Lei, J. Bootstrapping and Sample Splitting for High-Dimensional, Assumption-Lean Inference. *Ann. Stat.* **2019**, *47*, 3438–3469. [CrossRef]
28. Whelton, P.K.; Carey, R.M.; Aronow, W.S.; Casey, D.E., Jr.; Collins, K.J.; Dennison Himmelfarb, C.; Wright, J.T., Jr.; DePalma, S.M.; Gidding, S.; Jamerson, K.A.; 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. *Hypertension* **2018**, *71*, e13–e115. [CrossRef]
29. Chowdhury, M.Z.I.; Leung, A.A.; Walker, R.L.; Sikdar, K.C.; O'bEirne, M.; Quan, H.; Turin, T.C. A Comparison of Machine Learning Algorithms and Traditional Regression-Based Statistical Modeling for Predicting Hypertension Incidence in a Canadian Population. *Sci. Rep.* **2023**, *13*, 13. [CrossRef]
30. Franklin, S.S.; Gustin, W.; Wong, N.D.; Larson, M.G.; Weber, M.A.; Kannel, W.B.; Levy, D. Predictors of New-Onset Diastolic and Systolic Hypertension: The Framingham Heart Study. *Circulation* **2005**, *111*, 1121–1127. [CrossRef]
31. Wu, X.; Yuan, X.; Wang, W.; Liu, K.; Qin, Y.; Sun, X.; Ma, W.; Zou, Y.; Zhang, H.; Zhou, X.; et al. Value of a Machine-Learning Approach for Predicting Outcomes in Young Hypertensive Patients. *Hypertension* **2020**, *75*, 451–459. [CrossRef]
32. Du, J.; Chang, X.; Ye, C.; Zeng, Y.; Yang, S.; Wu, S.; Li, L. Developing a Hypertension Visualization Risk Prediction System. *Sci. Rep.* **2023**, *13*, 46281. [CrossRef]
33. Hrytsenko, Y.; Shea, B.; Elgart, M.; Kurniansyah, N.; Lyons, G.; Morrison, A.C.; Carson, A.P.; Haring, B.; Mitchell, B.D.; Psaty, B.M.; et al. Machine-Learning Models for Predicting Blood Pressure Phenotypes with Multi-PRS. *Sci. Rep.* **2024**, *14*, 62945. [CrossRef]
34. Mroz, T.; Griffin, M.; Cartabuke, R.; Laffin, L.; Russo-Alvarez, G.; Thomas, G.; Smedira, N.; Meese, T.; Shost, M.; Habboub, G. Predicting Hypertension Control Using Machine Learning. *PLoS ONE* **2024**, *19*, e0299932. [CrossRef]
35. Septian, E.; Khaefi, M.R.; Athoillah, A.; Aisyah, D.N.; Hardhantyo, M.; Rahman, F.M.; Manikam, L. Prediction of Personalised Hypertension Using Machine Learning in Indonesian Population. *J. Med. Syst.* **2025**, *49*, 2253. [CrossRef]
36. Mitev, P. Development of a System for the Active Orientation of Small Screws. *Eng. Proc.* **2024**, *70*, 55. [CrossRef]
37. Mitev, P. Development of a Training Station for the Orientation of Dice Parts with Machine Vision. *Eng. Proc.* **2024**, *70*, 57. [CrossRef]
38. Kachuee, M.; Kiani, M.M.; Mohammadzade, H.; Shabany, M. Cuffless Blood Pressure Estimation Algorithms. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 859–869. [CrossRef] [PubMed]

39. González, S.; Hsieh, W.-T.; Chen, T.P.-C. A Benchmark for Machine-Learning-Based Non-Invasive Blood Pressure Estimation Using PPG. *Sci. Data* **2023**, *10*, 149. [CrossRef] [PubMed]
40. Grinsztajn, L.; Oyallon, E.; Varoquaux, G. Why Do Tree-Based Models Still Outperform Deep Learning on Tabular Data? In Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS 2022), New Orleans, LA, USA, 28 November–9 December 2022.