DOI: https://doi.org/10.52756/ijerr.2024.v46.011 *Int. J. Exp. Res. Rev.*, Vol. 46: 139-149 (2024)

Original Article | Peer Reviewed | Open Access

International Journal of Experimental Research and Review (IJERR)
© Copyright by International Academic Publishing House (IAPH)
ISSN: 2455-4855 (Online)
www.iaph.in

Machine Learning-Based Prediction System for Risk Assessment of Hypertension Using Symptoms Investigations

Simranjit Kaur[1*], Khushboo Bansal[1] and Yogesh Kumar[2]

[1]Department of CSE, Desh Bhagat University, Mandi Gobindgarh, Punjab, India; [2]Department of CSE, School of Technology, Pandit Deendayal Energy University, Gandhinagar, Gujarat, India

E-mail/Orcid Id:

SK, dhillon.simranjit@gmail.com; *KB*, erkhushboo1985@gmail.com;
YK, Yogesh.arora10744@gmail.com, https://orcid.org/0000-0002-2879-0441

Article History:

Received: 26th Mar., 2024
Accepted: 15th Dec., 2024
Published: 30th Dec., 2024

Keywords:
Hypertension, Cardiovascular, Machine Learning, Random Forest, Disease Symptoms, Clinical Applications

How to cite this Article:
Simranjit Kaur, Khushboo Bansal and Yogesh Kumar (2024). Machine Learning-Based Prediction System for Risk Assessment of Hypertension Using Symptoms Investigations. *International Journal of Experimental Research and Review*, *46*, 139-149.
DOI: https://doi.org/10.52756/ijerr.2024.v46.011

Abstract: Hypertension is a common condition of cardiovascular disease that poses significant health challenges among the public on a larger scale globally. It is important to accurately predict the risk of hypertension to save people and improve overall quality of life. Traditionally, the detection of hypertension relies on clinical criteria such as blood pressure measurement and examination of medical history. However, these methods have drawbacks involving potential human error, time consumption, and the possibility of missed diagnoses. The paper aims to identify the features or symptoms of hypertension disease and predict its risk factors using machine learning algorithms. Apart from this, it is of utmost importance to identify the symptoms as they play a pivotal role in recognizing the type of risk for hypertension. To successfully conduct the work, a dataset of 13 attributes, including gender, age, smoking habits, etc, has been used, which is further visualized graphically to understand the pattern among them. Later, multiple machine learning-based learning techniques have been applied and examined on the basis of standard metrics. Results indicate that random forest models outperform existing approaches, achieving an accuracy of 87.26% in predicting low and high-risk hypertension. Furthermore, classification reports reveal superior precision, recall, and F1-score for random forests compared to alternative models. Insights from learning curves and confusion matrices provide a valuable understanding of model performance and data sufficiency. Overall, this research highlights the impact of machine learning in accurately predicting the risk of hypertension and underscores the importance of ongoing research efforts to translate these findings into practical clinical applications.

Introduction

Hypertension, a prevalent and serious global health issue, stands as a pivotal risk factor for cardiovascular ailments like coronary heart disease and stroke. Characterized by elevated blood pressure (>140/90 mmHg), hypertension (HTN) places considerable strain on patients' hearts, compelling them to function significantly harder than usual to maintain adequate blood circulation throughout the body (Gupta, 2023). The condition, which affects the body's arteries, is prevalent among young adults, impacting approximately 1 in 8 individuals aged between 20 and 40 years. This incidence is anticipated to rise due to lifestyle factors and the lowering of diagnostic thresholds for hypertension. While the mechanisms remain unclear, early-life factors profoundly influence blood pressure (BP), with BP exhibiting strong tracking within individuals from adolescence into adulthood (Geevar, 2022). Elevated BP in youth correlates with abnormalities detected through brain and heart imaging, heightening the risk of cardiovascular events in later life (Hinton et al., 2020). Despite ongoing advancements in cardiovascular disease treatment in Canada, coronary artery disease and stroke persist as leading causes of premature mortality and disability among adults (Wyss et al., 2020). Hence, addressing modifiable risk factors such as dyslipidemia

*Corresponding Author: dhillon.simranjit@gmail.com

and hypertension constitutes a cornerstone of both medical intervention and public health policy. Notably, hypertension often manifests without noticeable symptoms, with severe cases giving rise to symptoms like headaches, blurred vision, and chest pain (Figure 1) (Thongtang et al., 2022). Regular blood pressure monitoring remains the primary method for detecting hypertension, as untreated hypertension can lead to various health complications, including kidney disease, heart disease, and stroke (Schmidt et al., 2020). Early prediction and effective management significantly reduce hypertension-related incidence and mortality rates, underscoring the importance of establishing a highly accurate risk prediction system for hypertension management (Carey et al., 2018). Artificial intelligence has emerged as an important tool in the healthcare sector in recent years, demonstrating effectiveness across diverse clinical contexts. Machine learning (ML) advancements have notably impacted hypertension research, particularly by integrating disparate data sources such as electronic health records and environmental data (Himabindu et al., 2024). ML models leverage these integrated datasets to identify population- and individual-specific patterns in hypertension pathology. ML algorithms, tailored for precise outcome predictions without explicit programming, exhibit varying degrees of success in hypertension research, reflecting the field's complexity and diversity (Montagna et al., 2022).

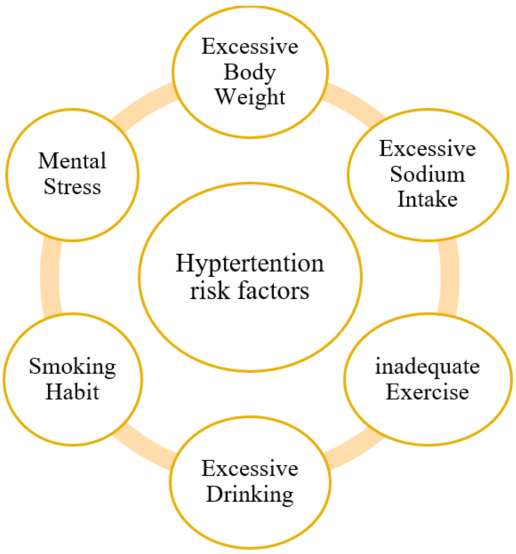


Figure 1. Risk factors of hypertension.

Despite the significant advancements in applying machine learning techniques for hypertension prediction and management, several research gaps remain. First, while numerous models have achieved high accuracy, the generalizability of these models across diverse populations, varying demographic factors and different healthcare settings remains underexplored. Additionally, most studies focus on individual datasets with limited representation of global diversity, indicating the need for more comprehensive, multi-national datasets. Thus, the goal of the paper is to work on these limitations and develop a system which is used to identify the symptoms of hypertension and explore the risk assessment concerning the given symptoms using multiple machine learning models. This framed objective will address the research question of how machine learning techniques are used to identify the risk of hypertension and which model performed well for this dataset.

Related Work

The section discusses how machine learning approaches help to identify the early prediction of hypertension disease. Additionally, it also defines how risk can be assessed and symptoms can be identified to cure the disease by taking some necessary action to overcome and improve the overall quality life of patients. The related work carried out studies on hypertension detection by different researchers. Shrivastava et al. (2023) aimed to develop of machine-learning system for predicting blood pressure. Their suggested model incorporated variations in diastolic and systolic blood pressure together with 11 other lifestyle characteristics. It aimed to provide a full understanding of the elements that shoot up the blood pressure and generated individualized predictions for people by including these aspects. The evaluation findings indicated that the random forests algorithm outperformed all other classifiers in accurately predicting both systolic and diastolic blood pressure. Silva et al. (2022) analysed previous work on the prediction of hypertension. The authors used ASRview algorithm for the screening and compared 21 articles from Jan 2018 to May 2021 based on data balancing, selection of variable, performance measures, etc. They found that random forest, support vector machine, and extreme gradient boosting were the most efficient algorithms for predicting the risk of hypertension. Bae et al. (2022) introduced a system to predict masked hypertension using clinical blood pressure along with other clinical characteristics. They analyzed the system using 809 ambulatory blood pressure monitoring (ABPM) studies conducted on chronic kidney disease participants among Children with non-hypertensive clinic BP. Islam et al. (2022) did a study on hypertension using surveys that were based on the general population. The researchers created a consolidated dataset by standardizing individual-level data from the most recent nationally representative Demographic and Health

DOI: https://doi.org/10.52756/ijerr.2024.v46.011

Surveys conducted in India, Nepal, as well as Bangladesh. The dataset included variables such as socio-demographic, blood pressure (BP), economic parameters, haemoglobin, weight, height, and random blood glucose. The definition of hypertension was based on the criteria established by JNC-7. In order to carry out this task, researchers put up six distinct machine learning models to forecast hypertension and its related risk variables. Kaur et al. (2024) developed an automated method for identifying hypertension by exploring various deep-learning techniques. They used a dataset that included blood pressure measures, medical history, demographic information, and lifestyle factors. Initially, they applied pre-processing techniques to the dataset to address multiple characteristics for hypertension prediction, resolve missing values, normalize features, and address class imbalance. On training the models, it was found that GRU exhibited the highest accuracy of 99.68% with 0.99 as precision, F1 score, and recall. Additionally, the Embedded GRU model yielded an accuracy of 98.10%, while the Bidirectional LSTM model produced superior outcomes predicting stroke with 98.85% accuracy. Chobufo et al. (2020) used machine learning algorithms to identify hypertension in people based on the selected features. A set of questionnaires taken from World Hypertension Day from 2015 to 2019 were compiled and analysed. A total of 20206 individuals participated in the survey and five machine learning techniques along with the balancing techniques were used for testing. Researchers stated that detecting hypertension at a population level posed certain challenges but using ML approach could be useful to enhance the effectiveness and efficiency of screening programs. Du et al. (2023) developed a risk prediction system based on visualization for personalized hypertension management, employing SHAP and machine learning as auxiliary tools to construct ten hypertension risk prediction models using 1617 anonymized health check datasets. Model performance was assessed using accuracy, ROC curve analysis, and F1-score metrics. Fang et al. (2023) stated that the percentage of patients in China with hypertension who have improved their knowledge, treatment, and control rates is 51.6%, 45.8%, and 16.8% respectively. However, these rates are still considered poor. In addition, they referenced the viewpoints of clinical investigations that indicated that implementing appropriate medications or making lifestyle modifications could considerably diminish the advancement of the condition for individuals who are at risk. They created a system that used LightGBM and KNN with and accuracy of 86% and recall of 92% to forecast hypertension for the next five years. Anand et al. (2024) evaluated a machine learning model for predicting pulmonary hypertension (PH) using echocardiography data from 7853 patients, split into training, validation, and testing sets. A gradient boosting machine handled missing data without imputation. PH was defined as a mean pulmonary artery pressure above 20 mm Hg. The cohort's average age was 64, with 44% females and 81% diagnosed with PH. The model, incorporating 19 echocardiographic features, achieved an AUC of 0.83 in the testing set, with 82% accuracy, 88% sensitivity, 89% positive predictive value, and 54% negative predictive value. Bardhan and Roga (2024) introduced an innovative method to automatically

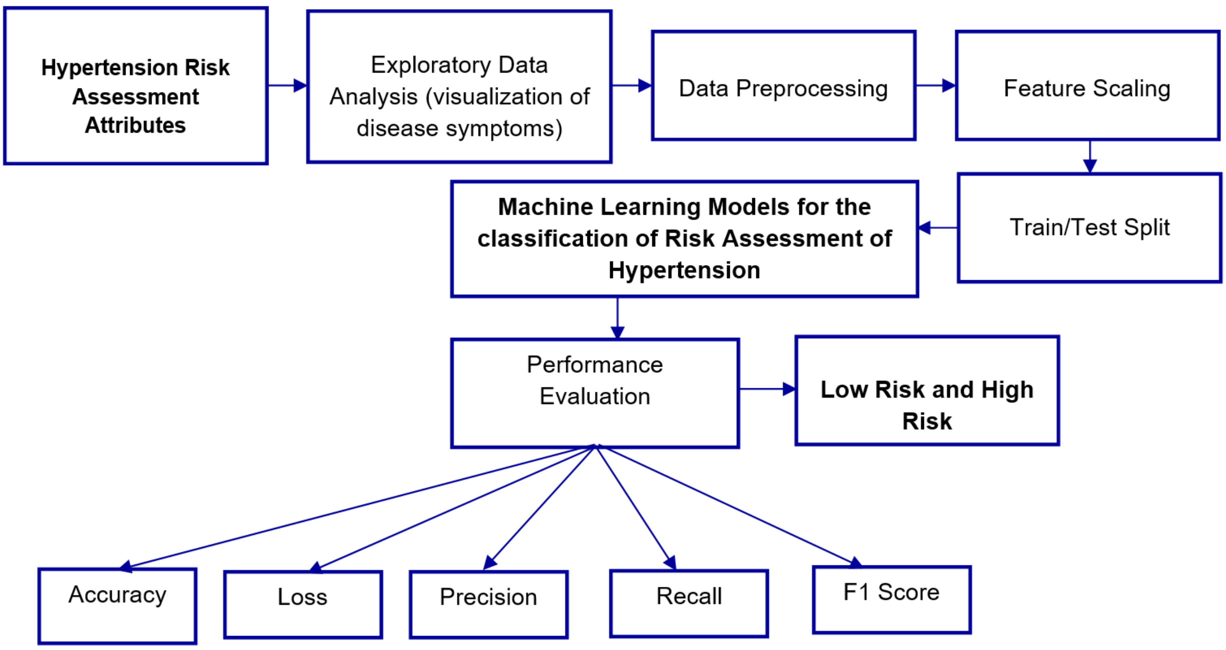


Figure 2. System design for prediction of low and high-risk of hypertension.

DOI: https://doi.org/10.52756/ijerr.2024.v46.011

classify the risk indexing of hypertension using cardiac ECM images via machine learning. The process included pre-processing the image, Bi-dimensional Empirical Mode Decomposition, extracting and selecting features, and classification using a statistical T-test for hypertension indexing. On applying 150 normal and hypertensive cardiac ECM images, their method achieved a classification accuracy of 98.9%, sensitivity of 97%, and specificity of 100%.

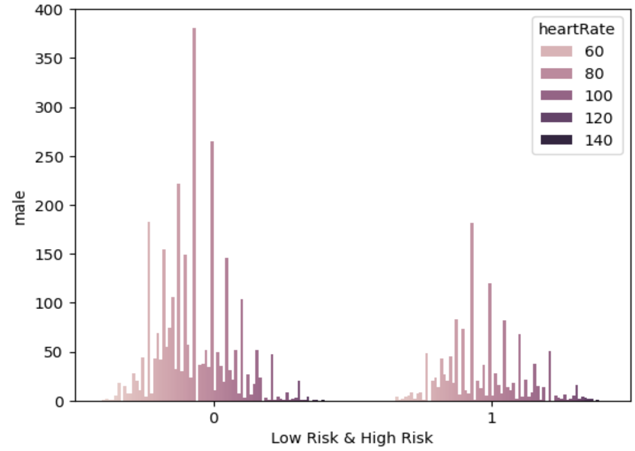
Working Methodology

The methodology section of the paper comprises a detailed analysis and step-by-step approach to handling hypertension risk based on the feature analysis of disease symptoms. The section also discusses how risk factors are pre-processed and machine learning models are applied to predict hypertension risk (as shown in Figure 2).

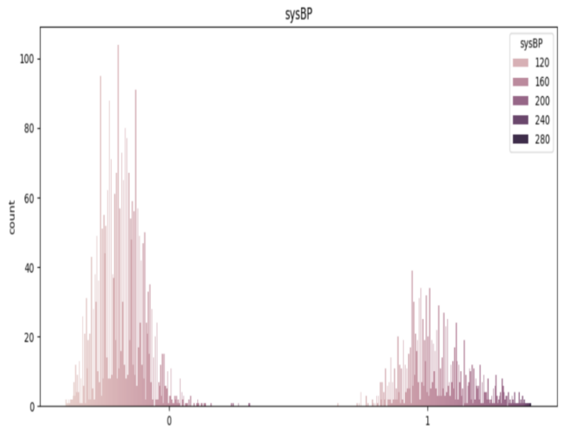
Data Source and disease symptoms

The dataset consists of various demographic and health-related factors intended for forecasting hypertension risk. Each record contains details such as gender, age, smoking habits, body mass index (BMI), high blood pressure medication (BPMeds), presence of diabetes, total cholesterol levels, systolic as well as diastolic blood pressure, heart rate, glucose levels, and the corresponding hypertension risk classification (0 for low risk, 1 for high risk). With a total of 13 attributes, this dataset offers a comprehensive insight into the factors associated with hypertension, facilitating the development of predictive models for risk assessment and prevention strategies (/https://www.kaggle.com/datasets/khan1803115/hypertension-risk-model-main).

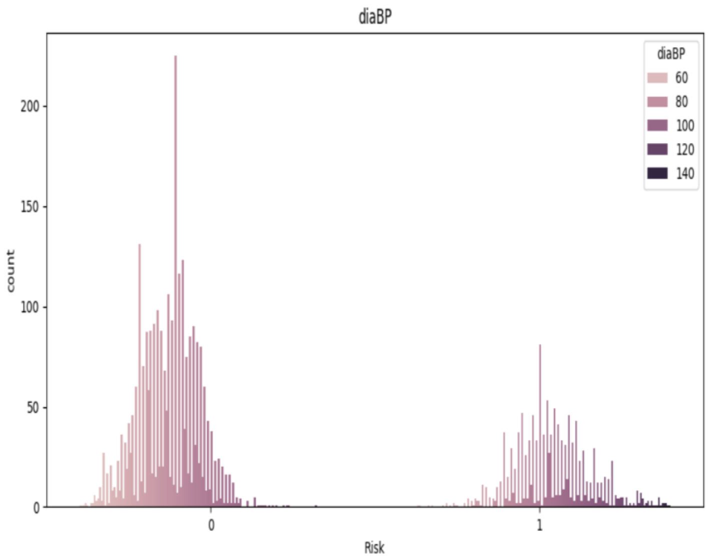
The visualization helps identify the various risk



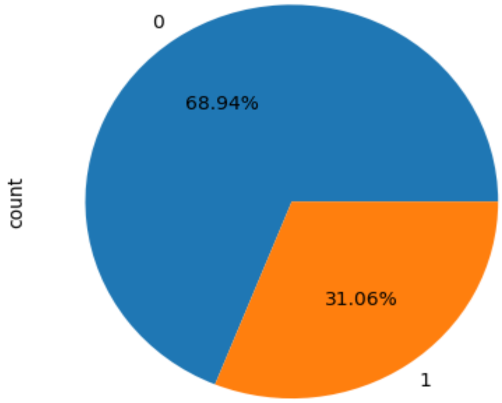
(a) Low & High Risk with heart rate



(b) Systolic Blood Pressure Level



c) Diastolic Blood Pressure Level



d) Total Count of High and Low Risk Hypertension

Figure 3. Visualization based on risk assessment parameters.

DOI: https://doi.org/10.52756/ijerr.2024.v46.011

parameters of hypertension risk assessment, as shown in Figure 3. Here, we observe that variation in heart rate may also be a cause of hypertension, and the type of blood pressure is one of the parameters that reflect the symptoms of hypertension. Additionally, we can check the correlation between the risk parameters based on the disease symptoms. Figure 4 shows the heatmap depicting the correlation between the attributes based on the risk symptom attributes.

of them are in the range, thus enhancing the performance of the machine learning models. After the data is scaled, it is split into two sets: one to use for training the models (70%) while the other to test (30%) the ability of the models to generalize. These scaled features are analyzed using different machine learning classifiers to assess the models' performance for hypertension risk prediction, which will serve as a proper guideline for doctors to make clinical decisions.

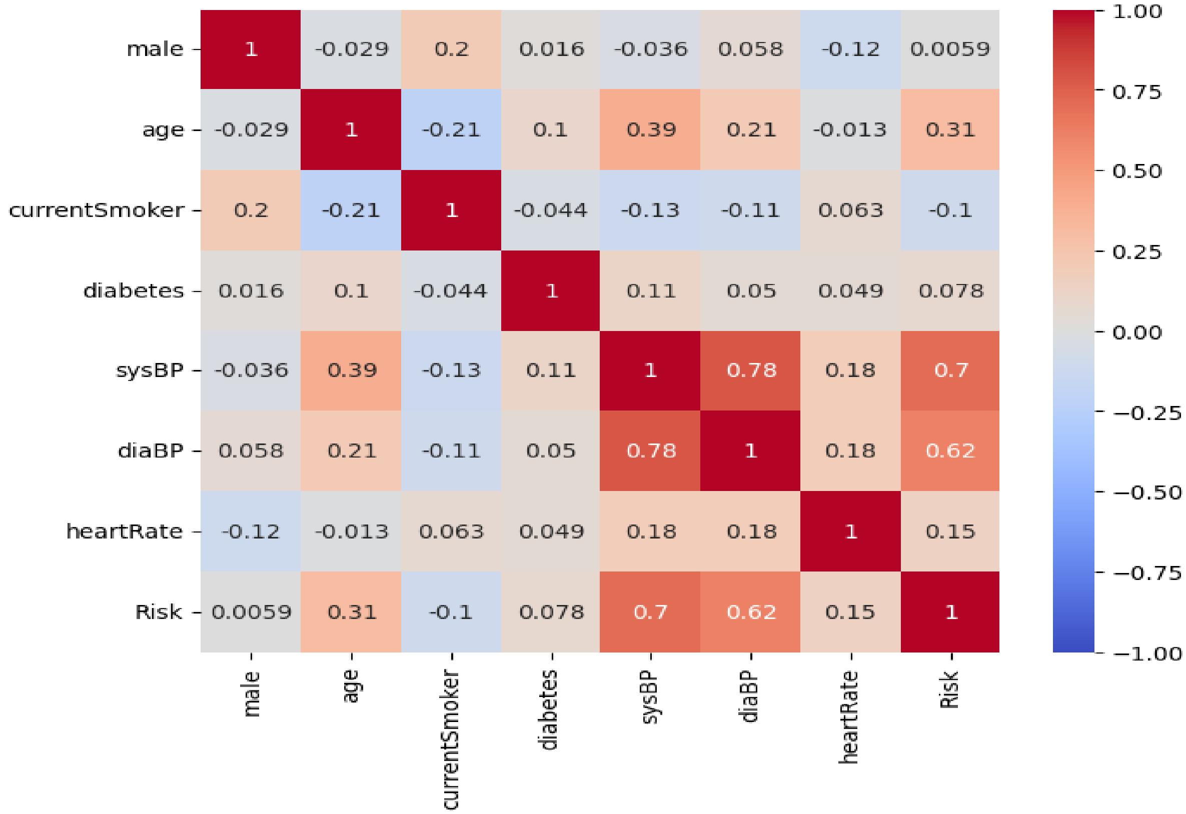


Figure 4. Correlation between the hypertension risk attributes.

After properly visualizing the hypertension disease features, pre-processing was conducted to address missing values in the database and encode string values into numeric ones during the data encoding phase. Feature scaling is a key operation during data pre-processing and helps make all features consistent. In this paper, feature scaling is conducted using the *min-max scaler*, whereby data is normalized to fall within a certain range, often 0 and 1. This method normalises the values of features by making proportional transformations that are applied to the minimum and maximum of the datasets. Doing so reduces the influence of the outliers in the learning process and avoids the effects of negative values to be imparted into the computations.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Here, X refers to the original value, X_{min} and X_{max} refers to the minimum and maximum value of the dataset, respectively and X' is the scaled value. It guarantees all

Applied learning models for hypertension risk prediction

Machine learning models have been applied to the preprocessed and scaled data to classify hypertension risk prediction. Logistic regression is a statistical technique that predicts the result of a dependent variable of type categorical by considering one or more predictor factors. The logistic regression model employs the logistic function to model the probability where a particular given input is related to a certain class (Lanza et al. 2024). On the other hand, random forest is also used to classify hypertension risk assessment. Unlike logistic regression, which is based on a single equation, Random Forest operates by constructing many decision trees for disease parameters during training and outputting the prediction of every tree. In the random forest, bootstrap sampling is performed on dataset on repeated samples with a replacement where each tree is trained on the subset of the original disease data (Singh et al. 2024). Meanwhile,

DOI: https://doi.org/10.52756/ijerr.2024.v46.011

Table 1. Evaluative parameters of the model.

Parameter	Definition	Equation
Accuracy	The ability of the model to accurately predict the class	$\frac{\text{True Negative} + \text{True Positive}}{\text{False Positive} + \text{True Positive} + \text{False Negative} + \text{True Negative}}$
Loss	The rate to which model predictions are wrong.	$\frac{(\text{Actual Value} - \text{Predicted Value})^2}{\text{Total observations}}$
Recall	The percentage of accurate predictions made for true positives.	$\frac{\text{True Positive}}{\text{False Negative} + \text{True Positive}}$
Precision	The percentage of positive predictions that appear to be true.	$\frac{\text{True Positive}}{\text{False Positive} + \text{True Positive}}$
F1 Score	A composite metric that includes precision as well as recall into a single value.	$2\,\frac{\text{Recall} * \text{Precision}}{\text{Precision} + \text{Recall}}$

K-Nearest Neighbors is an effective and simple algorithm for the classification of any disease prediction. It works based on the principle of finding the k nearest data points in the feature space and making predictions based on the disease features and attributes (Kumar et al. 2024).

$$d(Xi, Xj) = \sum n = 1N(Xi, n - Xj, n)\text{^}2 \quad (2)$$

Where N = number of features, and Xi,n and Xj,n are the values of the n-th feature of data points Xi and Xj, respectively. Naive Bayes is a classifier with probabilities that relies on Bayes' theorem where the features are not dependent on each other. It is normally used for large datasets and when features are independent of each other's (Desiani et al. 2019).

$$P(\, Ck \mid X\,) = P(X)P(\, X \mid Ck\,) \cdot P(Ck) \quad (3)$$

Support Vector Machine is the machine learning classifier used to classify healthcare data, which completely depends upon the classification of multiple disease features. The Support Vector Machine's primary objective is identifying a hyperplane that achieves the highest possible margin across classes. This hyperplane is then used to make accurate predictions regarding various risk types of hypertension (Singh et al. 2019). The equation can represent the hyperplane:

$$w \cdot x + b = 0 \quad (4)$$

W = normal vector to the hyperplane and x = feature vector with b as bias term. AdaBoost, also known as Adaptive Boosting, is a technique in ensemble learning that forms a powerful classifier by bringing together several weak classifiers. It focuses on the examples that are difficult to classify by giving them higher weights, and subsequent weak learners are forced to concentrate on the misclassified examples of the previous ones (Karima and Anggraeni 2024). It can be computed as:

$$F(x) = argmax_k(\textstyle\sum_{m=1}^{M} \alpha_m . h_m^k(x)) \quad (5)$$

Here F(x) = binary class classification, α= weight, and $h_m^k(x)$ = prediction of weak learners.

Performance Metrics

The efficacy of the machine learning models employed in this paper during training with the hypertension dataset has been evaluated using several metrics, as presented in Table 1 (Kaur et al., 2023; Kaur et al., 2022; and Koul et al., 2024).

Results and Discussion

This part of the paper discusses detailed results for different machine learning classification models used in predicting hypertension and their low and high-risk analysis. The applied models predict performance using various parameters, including accuracy, recall, loss, and precision, along with the F1-score. Additionally, the confusion matrix helps identify the actual and true labels for low and high-risk analysis. Learning curves are also generated to visualize cross-validation accuracy and training accuracy for each applied machine-learning model. The main purpose of applying these models is to depict their performance accurately in predicting hypertension risk. Table 2 and 3 show the prediction results for classifying hypertension risk analysis. From Table 2, the models demonstrate promising accuracy in predicting hypertension, with values ranging from 84.90% (KNN) to 87.26% (Random Forest). These results indicate that the predictive features utilized are well-suited for this clinical task. The relatively high accuracy across different algorithms suggests that machine learning could be effectively integrated into clinical decision-making tools for hypertension risk assessment. Among all the applied models, the Random Forest model obtained the highest accuracy of 87.26%, which could be particularly valuable in clinical settings. The reason is to balance the model's prediction performance with its interpretability and allow clinicians

DOI: https://doi.org/10.52756/ijerr.2024.v46.011

to identify the key risk factors of hypertension in patients. Likewise, models like AdaBoost and SVM also show strong performance with accuracies of 86.60% and 86.79%, respectively and can also be useful to provide reliable support for clinicians by enhancing risk prediction and enabling targeted interventions.

Table 2. Prediction accuracy and loss for risk analysis of hypertension.

Models	Accuracy	Loss
Logistic Regression	86.32	0.13
Random Forest	87.26	0.12
KNN	84.90	0.15
Naïve Bayes	85.94	0.14
SVM	86.79	0.13
Ada Boost	86.60	0.13

From Table 3, Logistic Regression and Random Forest achieve the highest precision in case of low-risk hypertension, with scores of 0.88 and 0.92, respectively. This high precision is clinically relevant, as it suggests these models are effective at accurately identifying individuals who are truly at low risk, potentially reducing unnecessary anxiety and interventions for patients misclassified as high-risk. Both models also exhibit high recall, indicating they are effective at identifying a significant proportion of actual low-risk cases, which is crucial for ensuring that no low-risk patient is wrongly categorized.

Table 3. Prediction of low and high-risk hypertension.

Model	Low-risk Hypertension			High-Risk Hypertension		
	A	**B**	**C**	**A**	**B**	**C**
Logistic Regression	0.88	0.94	0.91	0.82	0.68	0.74
Random Forest	0.92	0.89	0.91	0.76	0.82	0.79
KNN	0.86	0.94	0.90	0.81	0.63	0.71
Naïve Bayes	0.88	0.93	0.90	0.79	0.70	0.74
SVM	0.89	0.93	0.91	0.81	0.71	0.76
Ada Boost	0.91	0.91	0.91	0.77	0.77	0.77

A: Precision; B: Recall; **C:** F1-score

For high-risk hypertension, Random Forest stands out with the highest recall as 0.82 and a competitive F1 Score as 0.79. This performance is clinically significant because it means the model is effective at detecting high-risk individuals who need immediate attention, potentially improving early intervention and treatment outcomes. However, its slightly lower precision (0.76) suggests a higher rate of false positives, which could lead to over-treatment or increased healthcare costs. Logistic Regression and SVM also perform well with F1 Scores of 0.74 and 0.76, respectively, providing a good balance between precision and recall. While offering a consistent F1 Score of 0.77, Ada Boost shows average performance across precision and recall. This means that while Ada Boost may not be the top performer, it still offers reliable predictions, potentially suitable for scenarios where a consistent performance is valued over the highest precision or recall. Overall, these models' varying strengths highlight the importance of selecting the right model based on clinical needs, whether it is prioritizing precision for low-risk predictions or maximizing recall for high-risk detection.

Further confusion matrix and learning curves depicted the results for all the parameters. A confusion matrix (as shown in Figure 5) is typically used in classifying hypertension risk assessment using machine learning models to analyze performance. The instances in a predicted class as well as the actual class of hypertension for low and high risk, are represented by each row and column of the matrix, respectively.

As shown in Figure 6, learning curves are important measures for monitoring and optimizing the performance of applied classifiers of machine learning, including those designed for predicting the risk of hypertension risk assessment. Learning curves are used to diagnose the model behavior, guide data collection efforts, select the best model, and understand the tradeoffs between bias and variance.

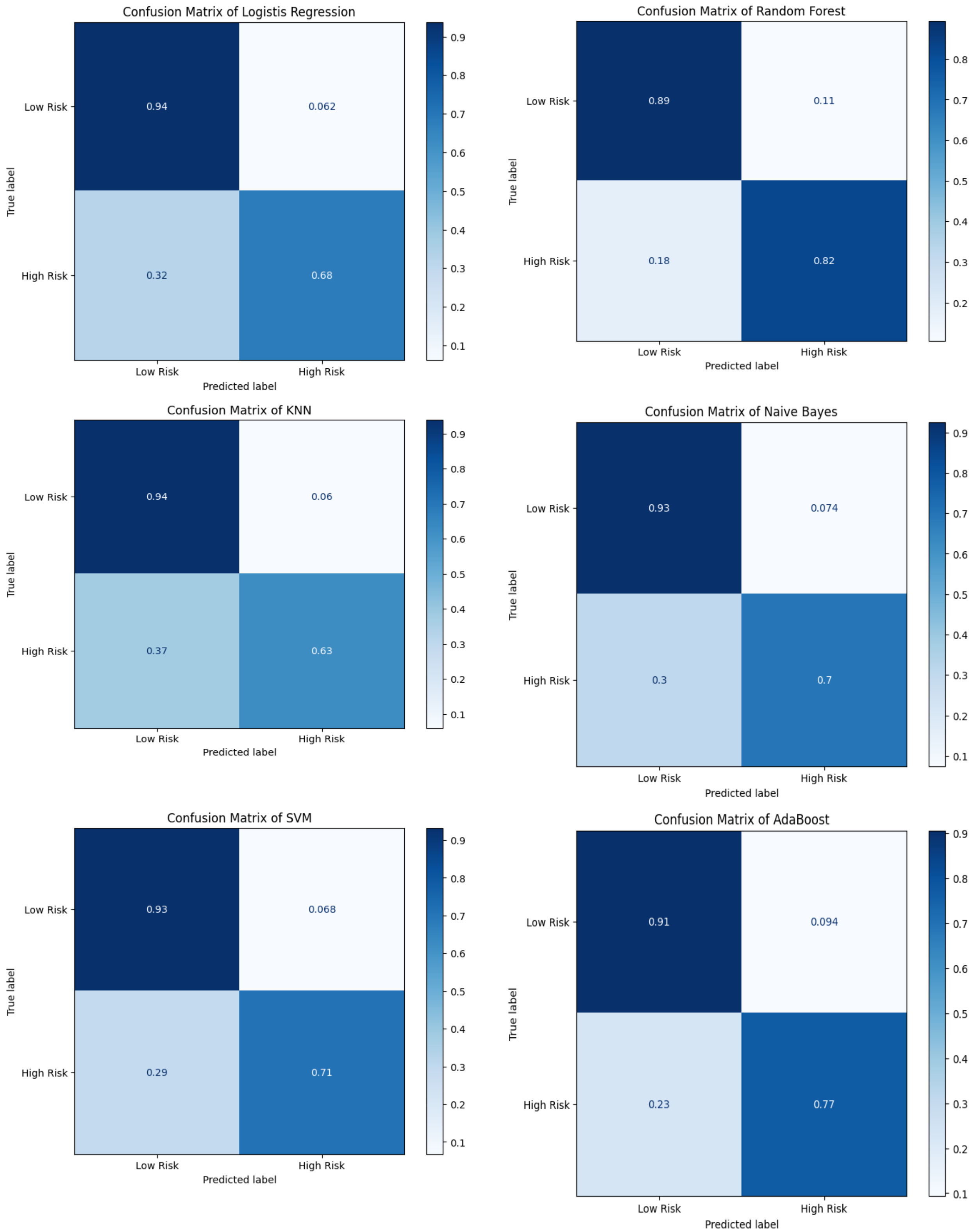


Figure 5. Confusion matrix to predict true labels for low and high-risk hypertension.

Conclusion and Future Direction

The results of this paper support the ability of machine learning models, especially the Random Forest algorithm to correctly and accurately predict the risk of hypertension. The proposed model achieved approximately 87.26% of prediction accuracy with the balancing precision, recall, and F1 score. This means that the model is reliable for identifying persons with the possibility of developing hypertension, and the number of false positives and negatives is kept to the barest minimum, which is very useful in clinical practice. Looking at the learning curves and getting deeper into the confusion matrices helped to make conclusions regarding the performance of models as well as the sufficiency of data used. This showed that while the current models are good, they could very likely be improved. For example,

DOI: https://doi.org/10.52756/ijerr.2024.v46.011

exploring additional features like diet and activity level, genetic factors, age, sex, and social position, pollution and stress levels could help to increase the potential of such models. This would help develop a more comprehensive and personalized approach to hypertension risk assessment. avenue for the possibility of the new innovative machine learning hypertension risk prediction tools to be used in the real world. If integrated into daily practice, this results in a possible reversal of healthcare models shifting from reactive to more preventive, particularly as far as hypertension is concerned.

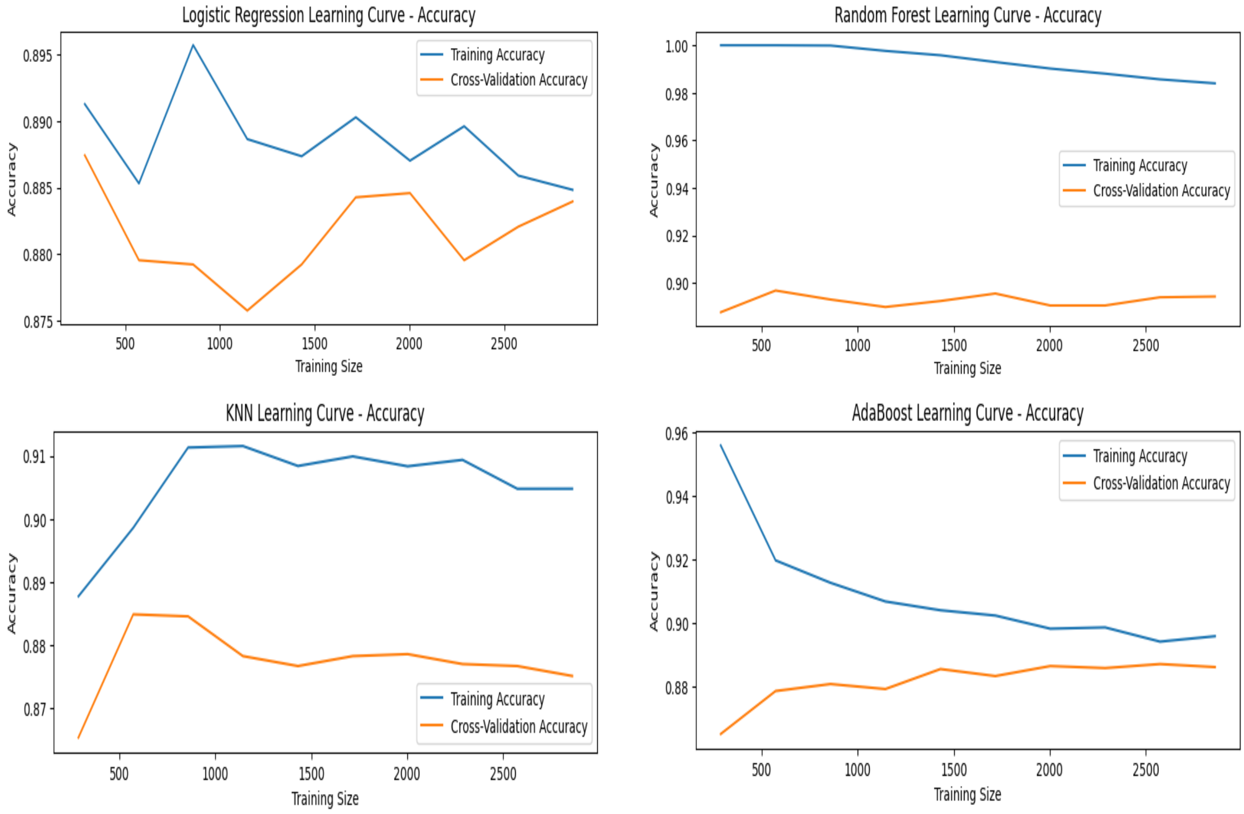


Figure 6. ML learning curves for risk analysis of hypertension.

In practical terms, implementing these machine learning algorithms in electronic health record (EHR) systems is one of the areas where promising results in the delivery of healthcare could possibly be achieved. If this type of risk assessment for hypertension could be incorporated into annual physicals or other clinical examinations, then clinicians would have the opportunity to screen for these individuals. Early detection could lead to early treatment through suitable changes in lifestyle, use of medicines or frequent monitoring, which may drastically reduce cases of hypertension-causing diseases like heart disease and stroke. Nevertheless, for such models to gain acceptance for application in clinical and health-care settings, they need further validation studies involving various populations and clinical solutions. This would ensure that the models are accurate and fair, hence pertinent across the different demographics of people, region, and health care institutions. However, issues of ethical nature like patients' consent, data protection, and clarity of proposed algorithms will be key factors determining trust of both the healthcare professionals and the patients. Hence, in a real sense, this paper creates an

Conflict of Interest

No conflict of interest

References

Anand, V., Weston, A. D., Scott, C. G., Kane, G. C., Pellikka, P. A., & Carter, R. E. (2024). Machine Learning for Diagnosis of Pulmonary Hypertension by Echocardiography. In *Mayo Clinic Proceedings, 99(2)*, 260-270. https://doi.org/10.1016/j.mayocp.2023.05.006

Bae, S., Samuels, J. A., Flynn, J. T., Mitsnefes, M. M., Furth, S. L., Warady, B. A., ... & CKiD Study. (2022). Machine Learning–Based Prediction of Masked Hypertension Among Children with Chronic Kidney Disease. *Hypertension*, *79*(9), 2105-2113. https://doi.org/10.1161/HYPERTENSIONAHA.121.18794

Bardhan, S., & Roga, S. (2024). Hypertension detection and indexing from cardiac ECM image analysis. *Multimedia Tools and Applications*, *83*(10), 30541-30561.

https://link.springer.com/article/10.1007/s11042-023-16746-3

Carey, R. M., Muntner, P., Bosworth, H. B., & Whelton, P. K. (2018). Prevention and control of hypertension: JACC health promotion series. *Journal of the American College of Cardiology*, *72*(11), 1278-1293. https://www.jacc.org/doi/abs/10.1016/j.jacc.2018.07.008

Chobufo, M. D., Gayam, V., Soluny, J., Rahman, E. U., Enoru, S., Foryoung, J. B., ... & Nfor, T. (2020). Prevalence and control rates of hypertension in the USA: 2017–2018. *International Journal of Cardiology Hypertension*, *6*, 100044. https://doi.org/10.1016/j.ijchy.2020.100044

Desiani, A., Primartha, R., Arhami, M., & Orsalan, O. (2019). Naive Bayes classifier for infant weight prediction of hypertension mother. In *Journal of Physics: Conference Series, 1282(1),*012005. IOP Publishing.10.1088/1742-6596/1282/1/012005

Du, J., Chang, X., Ye, C., Zeng, Y., Yang, S., Wu, S., & Li, L. (2023). Developing a hypertension visualization risk prediction system utilizing machine learning and health check-up data. *Scientific Reports*, *13*(1), 18953. https://www.nature.com/articles/s41598-023-46281-y

Fang, M., Chen, Y., Xue, R., Wang, H., Chakraborty, N., Su, T., & Dai, Y. (2023). A hybrid machine learning approach for hypertension risk prediction. *Neural Computing and Applications*, *35*(20), 14487-14497. https://link.springer.com/article/10.1007/s00521-021-06060-0

Geevar, Z., Krishnan, M. N., Venugopal, K., Sanjay, G., Harikrishnan, S., Mohanan, P. P., ... & Thankappan, K. R. (2022). Prevalence, awareness, treatment, and control of hypertension in young adults (20–39 Years) in Kerala, South India. *Frontiers in Cardiovascular Medicine*, *9*, 765442. https://doi.org/10.3389/fcvm.2022.765442

Gupta, A. (2023). The Impact of Gender and Age on the Relationship between Body Mass Index and Hypertension among Middle-Aged Inhabitants of Kolkata, India. *Int. J. Exp. Res. Rev.*, *30,* 228-235. https://doi.org/10.52756/ijerr.2023.v30.020

Himabindu, D. D., Pranalini, B., Kumar, M., Neethika, A., Sree N, B., C, M., B, H., & S, K. (2024). Deep CNN-based Classification of Brain MRI Images for Alzheimer's Disease Diagnosis. International Journal of Experimental Research and Review, *41*(Spl Vol), 43-54. https://doi.org/10.52756/ijerr.2024.v41spl.004

Hinton, T. C., Adams, Z. H., Baker, R. P., Hope, K. A., Paton, J. F., Hart, E. C., & Nightingale, A. K. (2020). Investigation and treatment of high blood pressure in young people: too much medicine or appropriate risk reduction? *Hypertension*, *75*(1), 16-22. https://doi.org/10.1161/HYPERTENSIONAHA.119.13820

Islam, S. M. S., Talukder, A., Awal, M. A., Siddiqui, M. M. U., Ahamad, M. M., Ahammed, B., ... & Maddison, R. (2022). Machine learning approaches for predicting hypertension and its associated factors using population-level data from three South Asian countries. *Frontiers in Cardiovascular Medicine*, *9*, 839379. https://www.frontiersin.org/articles/10.3389/fcvm.2022.839379/full

Karima, C., & Anggraeni, W. (2024). Performance Analysis of the Ada-Boost Algorithm For Classification of Hypertension Risk With Clinical Imbalanced Dataset. *Procedia Computer Science*, *234*, 645-653. https://doi.org/10.1016/j.procs.2024.03.050

Kaur, S., Bansal, K., & Kumar, Y. (2022). Artificial Intelligence approaches for Predicting Hypertension Diseases: Open Challenges and Research Issues. In *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*, 338-343. IEEE. https://ieeexplore.ieee.org/abstract/document/10072978

Kaur, S., Bansal, K., & Kumar, Y. (2023). Machine Learning based Approaches for Accurately Diagnosis and Detection of Hypertension Disease. In *2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 10,* 169-173. IEEE. https://ieeexplore.ieee.org/abstract/document/10434428

Kaur, S., Bansal, K., Kumar, Y., & Changela, A. (2024). A Comprehensive Analysis of Hypertension Disease Risk-Factors, Diagnostics, and Detections Using Deep Learning-Based Approaches. *Archives of Computational Methods in Engineering*, *31*(4), 1939-1958. https://link.springer.com/article/10.1007/s11831-023-10035-w

Koul, A., Bawa, R. K., & Kumar, Y. (2024). Enhancing the detection of airway disease by applying deep learning and explainable artificial intelligence. *Multimedia Tools and Applications*, 1-

33. https://link.springer.com/article/10.1007/s11042-024-18381-y

Kumar, Y., Kaur, I., & Mishra, S. (2024). Foodborne disease symptoms, diagnostics, and predictions using artificial intelligence-based learning approaches: a systematic review. *Archives of Computational Methods in Engineering*, *31*(2), 553-578. https://link.springer.com/article/10.1007/s11831-023-09991-0

Lanza, V. E., Silva, G. O., Quiroga, C. C. C., Cavalcante, M. A., Barroso, W. K. S., Brandão, A. A., ... & Lopes, R. D. (2024). Therapeutic Adherence According to the Morisky Scale in Patients with Hypertension. *International Journal of Cardiovascular Sciences*, *37*, e20220138. https://www.scielo.br/j/ijcs/a/kyQtk698hGJTmmLckbFjyQS/

Montagna, S., Pengo, M. F., Ferretti, S., Borghi, C., Ferri, C., Grassi, G., ... & Parati, G. (2022). Machine learning in hypertension detection: a study on world hypertension day data. *Journal of Medical Systems*, *47*(1), 1. https://link.springer.com/article/10.1007/s10916-022-01900-5

Schmidt, B. M., Durao, S., Toews, I., Bavuma, C. M., Hohlfeld, A., Nury, E., ... & Kredo, T. (2020). Screening strategies for hypertension. *Cochrane Database of Systematic Reviews*, (5). https://doi.org/10.1002/14651858.CD013212.pub2

Shrivastava, A., Chakkaravarthy, M., & Shah, M. A. (2023). A new machine learning method for predicting systolic and diastolic blood pressure using clinical characteristics. *Healthcare Analytics*, *4*, 100219. https://doi.org/10.1016/j.health.2023.100219

Silva, G. F., Fagundes, T. P., Teixeira, B. C., & Chiavegatto Filho, A. D. (2022). Machine learning for hypertension prediction: a systematic review. *Current Hypertension Reports*, *24*(11), 523-533. https://link.springer.com/article/10.1007/s11906-022-01212-6

Singh, J., Sandhu, J. K., & Kumar, Y. (2024). Metaheuristic-based hyperparameter optimization for multi-disease detection and diagnosis in machine learning. *Service Oriented Computing and Applications*, 1-20. https://link.springer.com/article/10.1007/s11761-023-00382-8

Singh, N., Singh, P., & Bhagat, D. (2019). A rule extraction approach from support vector machines for diagnosing hypertension among diabetics. *Expert Systems with Applications*, *130*, 188-205. https://doi.org/10.1016/j.eswa.2019.04.029

Thongtang, N., Sukmawan, R., Llanes, E. J. B., & Lee, Z. V. (2022). Dyslipidemia management for primary prevention of cardiovascular events: Best in-clinic practices. *Preventive Medicine Reports*, *27*, 101819. https://doi.org/10.1016/j.pmedr.2022.101819

Wyss, F., Coca, A., Lopez-Jaramillo, P., Ponte-Negretti, C., Wyss, F. S., Restrepo, G., ... & Iñiguez, A. (2020). Position statement of the Interamerican Society of Cardiology (IASC) on the current guidelines for the prevention, diagnosis and treatment of arterial hypertension 2017–2020. *International Journal of Cardiology Hypertension*, *6*, 100041. https://doi.org/10.1016/j.ijchy.2020.100041

How to cite this Article:

Simranjit Kaur, Khushboo Bansal and Yogesh Kumar (2024). Machine Learning-Based Prediction System for Risk Assessment of Hypertension Using Symptoms Investigations. *International Journal of Experimental Research and Review*, *46*, 139-149.

DOI : https://doi.org/10.52756/ijerr.2024.v46.011