Indonesian Journal of Electrical Engineering and Computer Science
Vol. 37, No. 1, January 2025, pp. 347~355
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v37.i1.pp347-355

Enhancing hypertension prediction: a hybrid machine learning optimization approach

Abd Allah Aouragh[1], Mohamed Bahaj[1], Fouad Toufik[2]
[1]MIET Laboratory, Faculty of Sciences and Techniques, Hassan 1st University, Settat, Morocco
[2]Computer Sciences Laboratory, Higher School of Technology, Mohammed V University, Sale, Morocco

Article Info

Article history:

Received May 7, 2024
Revised Sep 8, 2024
Accepted Sep 29, 2024

Keywords:

Ant colony optimization
Bayesian optimization
Feature selection
Hyperparameter optimization
Hypertension
Machine learning
Particle swarm optimization
SMOTE

ABSTRACT

Early identification of hypertension is crucial to prevent its serious complications, which can lead to devastating health effects by threatening lifestyle quality and significantly increasing premature mortality. This study aims to evaluate the effectiveness of machine learning techniques in predicting the presence of hypertension from an unbalanced dataset consisting of 4,363 records and 35 features. To balance the dataset, we employed the synthetic minority over-sampling technique (SMOTE) algorithm. In addition, to select the most relevant features, we used ant colony optimization. Next, we applied various algorithms, including logistic regression (LR), K-nearest neighbors (KNNs), support vector machine (SVM), extra trees (ETs), and AdaBoost (AB). We also evaluated the optimization of hyperparameters using two methods: Bayesian optimization (BO) and particle swarm optimization (PSO). The results reveal that the combination of AB with BO demonstrated superior performance, with an accuracy of 97.60%, a recall of 98.93%, and a precision of 98.59%. This research emphasizes the potential of machine learning techniques for anticipating hypertension and highlights the importance of optimization techniques in improving predictive models' performance.

Corresponding Author:

Abd Allah Aouragh
MIET Laboratory, Faculty of Sciences and Techniques, Hassan 1st University
Settat, Morocco
Email: abdallahaouragh@gmail.com

1. INTRODUCTION

Hypertension, also known as high blood pressure, is a medical condition characterized by persistently high blood pressure that affects more than a billion people worldwide [1]. It represents a serious global health problem, which notably increases the risk of cardiovascular disease, stroke, and kidney disorders [1], [2]. Worldwide, around 10 million deaths are associated with hypertension every year, making it one of the foremost factors leading to death [1], [2]. The risks and consequences of hypertension include damage to vital organs, reduced quality of life, and an increased risk of premature death [2], [3]. In addition, hypertension is a prominent risk element for other serious health issues, such as kidney disease and diabetes [2], [3]. Available treatments cover antihypertensive drugs, lifestyle modifications, and regular monitoring [3]. However, challenges remain, including treatment non-adherence and the costs associated with long-term disease management [4]. Consequently, early recognition and intervention are crucial to mitigating the risk of serious complications and reducing the morbidity and mortality associated with hypertension, thus helping to alleviate the burden of the disease on healthcare systems [5]. Machine learning techniques provide a valuable tool in the healthcare field, offering the ability to analyze large medical datasets to predict diseases [6], [7]. Their use can help to identify risk factors, detect early warning signs of conditions such as hypertension, and tailor interventions in a personalized way [8]. By integrating data balancing [9] and feature selection [10] methods, as well as hyperparameter optimization [11], we can improve the accuracy of predictive models, thus contributing to more efficient patient management and reducing the burden on healthcare systems [12]. These techniques' effectiveness in anticipating healthcare needs paves the way for a proactive approach to medical care, promoting optimized clinical outcomes and better allocation of medical resources.

In this context, Padmanabhan *et al.* [13] assert that while artificial intelligence (AI) and machine learning (ML) hold the potential for reducing the global burden of hypertension, their full potential is yet to be realized due to several challenges. They advocate for a closer collaboration between clinicians and ML experts to successfully integrate AI into hypertension treatment, underscoring the importance of continued research and development to overcome existing hurdles. Despite acknowledging these challenges, they remain optimistic about the future role of AI and ML in hypertension care. Martinez-Ríos *et al.* [14] carried out a literature review on the use of machine learning techniques in healthcare, with a particular focus on studies of hypertension. Their analysis revealed that machine learning models have been widely used to detect hypertensive subjects from physiological signals, notably electrocardiograph (ECG) and photoplethysmography (PPG) waves. However, they noted insufficient exploration of the correlation between socio-demographic or clinical data and physiological signals, despite their potential impact on blood pressure. Finally, they raised concerns about the interpretability of models and their training time, which can be mitigated by employing feature selection methods. Lafreniere *et al.* [15] have suggested employing artificial neural networks (ANNs) as a robust machine learning method to anticipate the presence of hypertension in at-risk individuals while limiting human error. Their research identifies key risk factors based on medical history, current health status, and patient demographics, and then uses them to predict hypertension in a given individual. Their neural network model achieves an accuracy of around 82%. Fang *et al.* [16] suggested lifestyle changes, drug treatment, or machine learning techniques to curb hypertension progression in preclinical individuals. They developed a predictive model using a combination of KNN and LightGBM to estimate hypertension risk over five years based on age and blood indicators. Their model exhibited reliability, with precision and recall exceeding 86% and 92%, respectively.

Hung *et al.* [17] developed predictive models based on machine learning to detect masked and uncontrolled hypertension from clinical features collected during a single consultation. Thirty-three clinical characteristics were used for constructing models utilizing ANN, logistic regression (LR), extreme gradient boosting (XGBoost), and random forest (RF). The results showed that the RF, XGboost, and ANN models outperformed the LR model, with better efficiency in terms of area under the ROC curve (AUC) (83.7%). Zhao *et al.* [18] assessed and contrasted the efficacy of four machine learning algorithms: LR, CatBoost, ANNs, and RF. Using a dataset comprising 29,700 instances gathered during physical examinations, the findings indicated that the RF model outperformed the others, with an AUC-ROC of 92%, accuracy of 82%, sensitivity of 83%, and specificity of 81%. Yan *et al.* [19] used a particle swarm optimization (PSO) algorithm to improve the structure of a backpropagation neural network, resulting in an effective model for predicting hypertension risk. Compared with other methods such as LR, the PSO model stood out for its high performance, with an accuracy of 85.38%, a sensitivity of 43.90%, a specificity of 96.66%, and an AUC of 86%. Chang *et al.* [20] used medical data from a Beijing hospital to predict hypertension outcomes with a machine learning model. They employed a sequential gain progression tabu search feature selection method to identify key medical variables influencing hypertension outcomes and then utilized the XGBoost algorithm for prediction due to its stability. Their model achieved high performance on test datasets, with AUC, precision, F1-score, and recall values of 92%, 94%, 87%, and 80%, respectively.

Our research significantly advances the field of machine learning in hypertension prediction by addressing key gaps in existing studies. We observe several unresolved issues: the comprehensive management of unbalanced data, the effective selection of relevant features from a large set, and the optimization of hyperparameters, which are often neglected. To address data imbalance, we utilize synthetic minority over-sampling technique (SMOTE), which boosts the reliability of predictions. Furthermore, we introduce an innovative application of ant colony optimization (ACO) for feature selection, and compare Bayesian optimization (BO) and PSO for hyperparameter tuning. Our approach, tested across five widely used machine learning algorithms-LR, support vector machine (SVM), k-nearest neighbors, extra trees (ET), and AB-outperforms existing methods, setting a new standard in the field. This study not only highlights the importance of addressing these unresolved issues but also offers a practical framework for improving predictive models in clinical settings, paving the way for future research to build on these findings.

In the following sections of this document, the structure is as follows: section 2 offers a comprehensive explanation of the materials and methods used. In section 3 outlines and discusses the results, including an analysis of the effects of the techniques used. Finally, section 4 summarizes the main findings and suggests research directions for future investigations.

2. MATERIALS AND METHODS

2.1. Proposed method

As part of our study to improve hypertension prediction using a hybrid machine learning optimization approach, we developed a rigorous methodology to ensure a systematic workflow. Our study began with the acquisition of the Hypertension Arterial Mexico dataset, comprising 35 hypertension-relevant attributes. Next, we conducted exploratory data analysis and pre-processing tasks, including normalization, to prepare the dataset for in-depth analysis. To overcome the imbalance inherent in the dataset, we used the SMOTE to effectively balance the classes. Next, we undertook feature selection, opting for the ant colony optimization algorithm to identify the most discriminating features for hypertension prediction. Next, we applied five machine learning algorithms: LR, KNNs, SVM, ET, and AB. Each algorithm underwent a rigorous evaluation to measure its performance in predicting hypertension outcomes. To achieve optimal model performance, we performed hyperparameter optimization using two distinct techniques: BO and PSO. By comparing the results of these optimization methods, we sought to identify the most effective approach for refining our machine learning models. Our methodology encompasses a structured and comprehensive approach aimed at harnessing the power of hybrid machine learning methods to enhance the precision of hypertension prognosis. The methodology employed in this study is depicted in Figure 1.

Figure 1. Proposed method

2.2. Dataset

The Hypertension Arterial Mexico dataset used in our study is the result of the consolidation of three datasets from the governmental National Health and Nutrition Survey (ENSANUT) [21], [22]. It offers a wide range of information concerning 4363 patients in Mexico, covering a wide range of 35 biological and anthropometric parameters, such as ambient temperature, uric acid, albumin, cholesterol (HDL, LDL, total), creatinine, glucose result, insulin, C-reactive protein, triglycerides, mean glucose, glycosylated hemoglobin, ferritin, folic acid, homocysteine, transferrin, vitamin B12, and vitamin D, as well as demographic details such as gender and age. In addition, the final column, “risk_hypertension” indicates whether the patient is at risk of developing hypertension, with the value “1” for those at risk and “0” for those who are not. The dataset shows a notable imbalance between classes, with 2816 (64.54%) cases with hypertension and 1,547 (35.46%) cases without. This disparity can lead to bias in machine learning models, compromising the reliability of predictions. To remedy this situation, it is imperative to address this imbalance before training the models. Thus, oversampling techniques will be employed to increase minority class representation.

2.3. Balancing dataset: SMOTE

The fundamental principle of dataset balancing lies in the correction of distribution imbalances between the different classes of a target variable, with the aim of avoiding bias in machine learning models. This correction is essential to guaranteeing accurate and reliable predictions, particularly in cases where the minority group is notably underrepresented. In our study, we examined several oversampling techniques, including ADASYN, SVMSMOTE, K-Means SMOTE, and SMOTE, of which SMOTE proved to be the most successful [10], [23]. SMOTE, is a commonly used method that generates new synthetic examples for the minority class. Specifically, SMOTE randomly chooses an instance from the minority class and determines its k nearest neighbors also belonging to the same class. Then, new synthetic examples are created by interpolating the features of the selected example and its neighbors in the attribute space. This interpolation generates examples along the line connecting the initial example to its neighbors, thus preserving the existing data distribution [10], [24], [25]. The use of SMOTE effectively balanced our dataset, improved the ability of our models to correctly generalize the data, and produced more accurate predictions while preserving the representativeness of the original data. The choice of the SMOTE technique was based on its reputation for effectiveness in oversampling minority classes in unbalanced datasets, particularly in the healthcare field, where it has proven its ability to improve the effectiveness of predictive models by ensuring a balanced representation of minority classes. Furthermore, SMOTE has been widely validated in the literature for its ability to generate synthetic samples that closely resemble the original data distribution, facilitating robust model training and evaluation [10], [24], [25].

2.4. Feature selection: ant colony optimization

In machine learning, feature selection plays a crucial role, consisting of identifying the most relevant features within a dataset to build predictive models. This step aims to diminish the dimensionality of the data by eliminating duplicate features while preserving essential information for prediction tasks [10], [26]. ACO is inspired by the optimization capabilities observed in ant colonies, helping them find the most efficient paths to food sources. In feature selection, ACO is used to discover an optimal subset of characteristics that maximizes the effectiveness of predictive models. It uses a stochastic search process where potential solutions (feature sets) are explored and evaluated against a specified performance metric, guided by artificial pheromones, to identify promising solutions [27]. ACO offers several significant advantages over traditional techniques like grid search. Firstly, it efficiently explores the search space of feature subsets, even in complex situations, leading to the discovery of promising solutions. Secondly, ACO is resilient to noisy data and high-dimensional problems, rendering it suitable for real-world applications. In addition, it identifies non-redundant and informative feature subsets, improving the performance of predictive models by attenuating overfitting and facilitating generalization. Finally, the computational efficiency of ACO exceeds that of grid search thanks to its intelligent and efficient exploration of the search space [27].

2.5. Machine learning algorithms

Machine learning algorithms are extensively utilized in the medical field for classifying diseases and predicting diagnoses. In our study on hypertension prediction, we meticulously chose algorithms based on their proven performance in medical diagnostics, capacity to manage complex classifications, and their established credibility in the research community [28], [29]. This careful selection ensures that the algorithms used are particularly suited for delivering accurate and reliable hypertension predictions

2.5.1. Logistic regression

LR is a binary classification algorithm that calculates the probability of an observation belonging to a given class. By fitting a logistic curve, it models the correlation between the independent variables and the likelihood of the dependent variable occurring. Its major strengths lie in its simplicity, its effectiveness even in the presence of limited data, and its ability to provide probability estimates for each class [9], [29].

2.5.2. K-nearest neighbors

The KNN algorithm is a classification method that assigns a label to an observation according to the classifications of its closest neighbors in the feature space. It operates according to the principle that analogous points tend to aggregate within the same space. Its advantages include its simplicity of implementation, its adaptability to diverse data types, and its ability to capture complex structures in the data without making assumptions about their underlying distribution [28], [29].

2.5.3. Support vector machine

SVM is a powerful classification algorithm that seeks to locate the best-fitting hyperplane for separating data into different classes. It operates by identifying the optimal decision boundary between different groups by widening the distance between the nearest instances in every group. Its advantages lie in its proficiency in handling datasets with many dimensions, its flexibility to handle both linear and non-linear classification problems, and its robustness to overfitting problems [28], [29].

2.5.4. Extra trees

ET are a variation of decision trees that integrate more randomness into the node-splitting process. Unlike decision trees, which examine all possibilities, ETs randomly select splitting thresholds from a predefined set of values. This approach aims to reduce model variance, prevent overfitting, and improve generalization. In addition, these trees are more robust to outliers and noise, while generally offering shorter training [10], [29].

2.5.5. AdaBoost

AB is a boosting strategy that iteratively combines several weak classifiers to form a more resilient overall model. At each iteration, AB adjusts the weights of the instances according to their performance, placing greater emphasis on poorly classified instances. This approach promotes the construction of a model less prone to overfitting and capable of generalizing efficiently on new data. By using weak classifiers, AB offers great flexibility, can be adapted to a variety of classification problems, and is robust to noise, making it suitable for complex medical datasets [28], [29].

2.6. Hyperparameter optimization with metaheuristics

Hyperparameter optimization is an essential phase in the process of building machine learning models, aimed at finding optimal values for the parameters that govern algorithm behavior. In our study, we have favored the use of advanced hyperparameter optimization techniques such as BO and PSO, approaches widely recognized in the field of machine learning research. Unlike traditional methods such as grid search, these techniques offer several significant advantages. They enable a more efficient exploration of the hyperparameter space, considerably reducing the computation time needed to find optimal configurations. In addition, they are better adapted to complex search spaces and are often associated with heightened accuracy, performance, and generalization [30], [31].

2.6.1. Bayesian optimization

BO is a hyperparameter optimization method based on the concept of Bayesian probability. Its principle is based on constructing a probabilistic model to estimate the proficiency of a machine learning model as a function of hyperparameters. Unlike grid search, which exhaustively explores all possible hyperparameter values, BO uses an iterative approach. At each iteration, it selects the most promising hyperparameter values based on previous observations of the objective function [31]. This approach enables a more efficient exploration of the hyperparameter space, focusing on the most promising regions. The advantages of BO lie in its ability to converge rapidly on optimal solutions, even in complex, high-dimensional search spaces. In addition, it makes efficient use of computational resources by dynamically adapting the distribution of hyperparameter evaluations according to the information accumulated during the optimization process [9], [11], [30].

2.6.2. Particle swarm optimization

PSO is an optimization technique inspired by the collective behavior of bird or insect swarms. Its principle is based on the movement of several “particles” in a multi-dimensional search space, each representing a potential solution. These particles adjust their movement according to their own best solution found so far, as well as the best solution found by the swarm as a whole [31]. This collaboration between particles promotes exploration of the solution space and helps to converge towards global optimums. In the context of hyperparameter optimization, PSO offers several advantages. Firstly, it offers efficient exploration of the hyperparameter space, enabling promising configurations to be found quickly. Another advantage of PSO is its ease of parallelization, which speeds up the optimization process by efficiently leveraging available computing resources [31].

3. RESULTS AND DISCUSSION

The experiments in our study were performed on Google Colab, a cloud-hosted machine learning platform. To implement the experiments, we used the Python programming language and exploited essential libraries such as Pandas, Scikit-learn, Scikit-optimize, and HypONIC for data manipulation, machine learning model building, and hyperparameter optimization.

Our study’s evaluation approach relies on a series of rigorous evaluation measures. Some of these are accuracy, recall, precision, F1-score, and AUC-ROC. They are all necessary for figuring out how well the models can predict the future and how well they can be used with new data [32]. In Tables 1–5, the results of these tests are shown. They show how well the algorithms worked when compared to the data processing methods of feature selection and hyperparameter optimization.

For the original dataset, as summarized in Table 1, the AB algorithm stood out with the best performance in several measures: an accuracy of 92.81%, a recall of 98.57%, an F1-score of 93.15%, and an AUC-ROC of 92.86%. In terms of precision, ETs achieved the best result, with a score of 90.93%. After applying the SMOTE technique to balance the dataset, Table 2 revealed a significant improvement in all the evaluation metrics, with an average increase of 1.64%. This time, AB outperformed the other algorithms, recording the highest scores for accuracy (94.59%), precision (95.60%), F1-score (94.48%), and AUC-ROC (94.58%). In terms of recall, ETs achieved the best result, with a value of 94.83%.

After applying the ACO algorithm for feature selection on the smote balanced dataset, the results presented in Table 3 reveal the predominance of AB, which improved by 1.60%. The AB algorithm recorded a remarkable accuracy of 96.18%, a recall of 97.32%, an F1-score of 96.20%, and an AUC-ROC of 96.19%. In terms of precision, the highest score was recorded by ETs, reaching 96.72%. After applying the BO, Table 4 reveals a slight performance improvement of 1.42%. Once again, AB stands out, achieving the highest values for accuracy (97.60%), recall (98.93%), F1-score (97.62%), and AUC-ROC (97.61%). Likewise, SVM achieved the highest precision of 98.59%. After the application of PSO, Table 5 also reveals a slight performance improvement of 1.05%, but not as significant as that of BO. Once again, AB stands out by achieving the highest values for accuracy (97.07%), recall (98.39%), F1-score (97.09%), and AUC-ROC (97.08%). In addition, SVM achieved the highest precision of 96.84%. Measurements for the different models are shown in Figure 2, providing a graphical representation to enhance clarity and comprehensibility.

The results above demonstrate the advantages of data balancing, feature selection, and hyperparameter optimization in the prediction of hypertension. The combination of SMOTE, ACO, BO, and the AB model produced the best results, with an accuracy rate of 97.60% and an F1-score of 97.62%. The AB and ET ensemble models were the most accurate and had the highest recall, F1-score, and AUC-ROC. This is because they are naturally good at figuring out complicated relationships between variables and making correct predictions. Their ability to train weak models and aggregate them significantly improved overall model performance. On the other hand, the SVM stood out for its precision because it uses decision hyperplanes that maximize the margin between different classes, enabling it to identify positive examples more easily. In our study, BO proved to be a more suitable and efficient approach to enhance the performance of our machine learning models than PSO. O relies on a probabilistic approach to determine optimal hyperparameters based on the results of previous iterations, which enables it to converge more quickly on high-quality solutions. In addition, this method is better suited to dealing with highly complex, non-linear search spaces. However, while PSO is an effective global optimization method, it can struggle with high-dimensional spaces and discontinuous solution regions, potentially leading to suboptimal configurations.

When comparing our results to those of previous studies, such as those by [13], [14], we find similarities in the application of machine learning techniques for hypertension prediction. For instance, both our study and theirs highlight the effectiveness of ensemble methods like AB and ET. However, our study uniquely contributes by focusing on the critical role of data balancing, feature selection, and hyperparameter optimization-areas less emphasized in earlier research. For example, while [15] achieved an accuracy of 82% using neural networks, and [16] attained an accuracy of 86% with a combination of KNN and LightGBM, our approach utilizing AB and ET demonstrated superior performance, achieving an accuracy of 97.60% and an F1-score of 97.62%. A significant strength of our study is the comprehensive methodological approach, which includes advanced techniques such as SMOTE for data balancing, ACO for feature selection, and BO for hyperparameter tuning. This approach not only enhances the accuracy and reliability of the predictive models but also provides a practical framework that can be readily applied in clinical settings. This is particularly relevant given the limitations noted in previous studies, such as those highlighted by [17], [18], who identified challenges in model interpretability and training times. Our work mitigates these issues by employing a robust feature selection process and optimized hyperparameter settings.

Despite these advances, our study has limitations, such as the potential lack of representativeness of our dataset, which may affect the generalizability of the findings. Furthermore, the absence of certain critical variables, as pointed out in previous studies, might influence the accuracy of the models. Future research should explore the application of our methodologies to more diverse datasets and different medical conditions, as well as their integration into real-world clinical workflows. Additionally, examining the impact of these techniques on patient outcomes will be crucial for validating their clinical utility. Overall, our study establishes a new benchmark in hypertension prediction and suggests future research directions, including improving data representativeness and exploring additional variables.

Table 1. Original dataset

	Accuracy	Precision	Recall	F1-score	AUC-ROC
LR	84.38%	85.93%	81.93%	83.88%	84.36%
KNN	85.57%	85.48%	93.40%	89.27%	82.44%
SVM	82.13%	81.20%	93.94%	87.11%	77.42%
ET	91.22%	**90.93%**	91.41%	91.17%	91.22%
AB	**92.81%**	88.30%	**98.57%**	**93.15%**	**92.86%**

Table 2. Balanced SMOTE dataset

	Accuracy	Precision	Recall	F1-score	AUC-ROC
LR	86.25%	87.31%	91.98%	89.58%	83.97%
KNN	87.05%	89.48%	83.72%	86.51%	87.02%
SVM	84.31%	84.42%	92.69%	88.36%	80.96%
ET	92.10%	93.01%	**94.83%**	93.91%	91.01%
AB	**94.59%**	**95.60%**	93.38%	**94.48%**	**94.58%**

Table 3. Balanced SMOTE dataset + ACO

	Accuracy	Precision	Recall	F1-score	AUC-ROC
LR	88.82%	87.39%	90.52%	88.93%	88.83%
KNN	90.59%	88.46%	93.20%	90.77%	90.62%
SVM	89.09%	87.07%	91.59%	89.28%	89.11%
ET	93.35%	**96.72%**	89.62%	93.04%	93.32%
AB	**96.18%**	95.10%	**97.32%**	**96.20%**	**96.19%**

Table 4. Balanced SMOTE dataset + ACO + BO

	Accuracy	Precision	Recall	F1-score	AUC-ROC
LR	89.17%	87.22%	91.59%	89.35%	89.19%
KNN	91.04%	88.42%	94.28%	91.26%	91.06%
SVM	93.26%	**98.59%**	87.66%	92.80%	93.21%
ET	94.06%	98.43%	89.45%	93.72%	94.02%
AB	**97.60%**	96.34%	**98.93%**	**97.62%**	**97.61%**

Table 5. Balanced SMOTE dataset + ACO + PSO

	Accuracy	Precision	Recall	F1-score	AUC-ROC
LR	88.91%	87.16%	91.06%	89.06%	88.93%
KNN	90.95%	88.27%	94.28%	91.18%	90.98%
SVM	92.46%	**96.84%**	87.66%	92.02%	92.42%
ET	93.88%	94.71%	92.84%	93.77%	93.87%
AB	**97.07%**	95.82%	**98.39%**	**97.09%**	**97.08%**

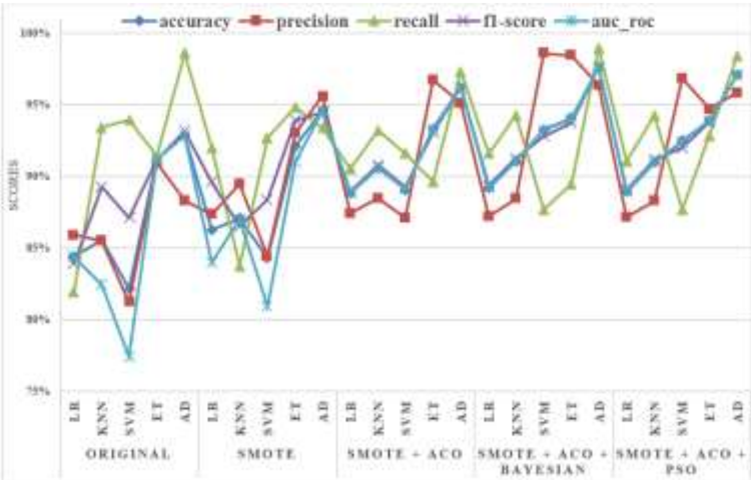

Figure 2. Metrics of different algorithms

4. CONCLUSION

This study highlights the effectiveness of machine learning techniques such as AB and ET in predicting hypertension, achieving an impressive accuracy of 97.60%. By integrating these algorithms with advanced feature selection methods like ant colony optimization, hyperparameter optimization through BO, and data balancing via SMOTE, we have demonstrated a significant enhancement in predictive performance. These findings suggest that such a comprehensive approach can substantially improve diagnostic accuracy and patient management in hypertension. The implications of our research extend beyond the immediate results. The combination of these techniques offers a robust framework for developing more precise and reliable predictive tools, which could transform how hypertension and potentially other cardiovascular conditions are diagnosed and managed. Future research should focus on validating these findings across diverse datasets and clinical settings to ensure broader applicability and robustness. Additionally, exploring the integration of these methods with electronic health records and real-time patient monitoring systems could further enhance their utility. Ultimately, our study contributes to the advancement of predictive medicine by providing a practical and effective approach for hypertension prediction. It opens avenues for further exploration into personalized treatment strategies and the optimization of healthcare delivery, with the potential to significantly impact both research and clinical practice.

REFERENCES

[1] S. Brouwers, I. Sudano, Y. Kokubo, and E. M. Sulaica, “Arterial hypertension,” *The Lancet*, vol. 398, no. 10296, pp. 249–261, Jul. 2021, doi: 10.1016/S0140-6736(21)00221-X.

[2] “Hypertension,” 2024. https://www.who.int/news-room/fact-sheets/detail/hypertension (accessed Jul. 26, 2024).

[3] J. A. Whitworth and J. Chalmers, “World Health Organisation-International Society of Hypertension (WHO/ISH) hypertension guidelines,” *Clinical and Experimental Hypertension*, vol. 26, no. 7–8, pp. 747–752, 2004, doi: 10.1081/CEH-200032152.

[4] B. V. Silva, C. Sousa, D. Caldeira, A. Abreu, and F. J. Pinto, “Management of arterial hypertension: challenges and opportunities,” *Clinical Cardiology*, vol. 45, no. 11, pp. 1094–1099, Nov. 2022, doi: 10.1002/clc.23938.

[5] P. Rodríguez-Lesmes, “Estimating the gains of early detection of hypertension over the marginal patient,” *PLoS ONE*, vol. 16, no. 7 July, p. e0254260, Jul. 2021, doi: 10.1371/journal.pone.0254260.

[6] A. A. Aouragh and M. Bahaj, “Advanced cardiovascular disease diagnosis with machine learning: exploring KBest, t-SNE, grid search, and ensemble methods,” in *Proceedings - 2023 IEEE International Conference on Advances in Data-Driven Analytics and Intelligent Systems, ADACIS 2023*, Nov. 2023, pp. 1–6, doi: 10.1109/ADACIS59737.2023.10424089.

[7] I. Jovovic, D. Babic, T. Popovic, S. Cakic, and I. Katnic, “Disease prediction using machine learning algorithms,” in *2023 27th International Conference on Information Technology, IT 2023*, Feb. 2023, pp. 1–4, doi: 10.1109/IT57431.2023.10078464.

[8] S. A. Alowais *et al.*, “Revolutionizing healthcare: the role of artificial intelligence in clinical practice,” *BMC Medical Education*, vol. 23, no. 1, p. 689, Sep. 2023, doi: 10.1186/s12909-023-04698-z.

[9] A. A. Aouragh and M. Bahaj, “Advancing breast cancer diagnosis with machine learning: exploring data balancing, feature selection, and bayesian optimization,” in *2023 International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications, CloudTech 2023*, Nov. 2023, pp. 01–06, doi: 10.1109/CloudTech58737.2023.10366058.

[10] A. A. Aouragh and M. Bahaj, “Feature selection and dimensionality reduction for unbalanced liver disease classification with optimized machine learning algorithms,” in *Colloquium in Information Science and Technology, CIST*, Dec. 2023, pp. 547–552, doi: 10.1109/CiSt56084.2023.10409967.

[11] J. Wu, X. Y. Chen, H. Zhang, L. D. Xiong, H. Lei, and S. H. Deng, “Hyperparameter optimization for machine learning models based on Bayesian optimization,” *Journal of Electronic Science and Technology*, vol. 17, no. 1, pp. 26–40, 2019, doi: 10.11989/JEST.1674-862X.80904120.

[12] M. Bekbolatova, J. Mayer, C. W. Ong, and M. Toma, “Transformative potential of ai in healthcare: definitions, applications, and navigating the ethical landscape and public perspectives,” *Healthcare (Switzerland)*, vol. 12, no. 2, p. 125, 2024.

[13] S. Padmanabhan, T. Q. B. Tran, and A. F. Dominiczak, “Artificial intelligence in hypertension: seeing through a glass darkly,” *Circulation Research*, vol. 128, no. 7, pp. 1100–1118, Apr. 2021, doi: 10.1161/CIRCRESAHA.121.318106.

[14] E. Martinez-Ríos, L. Montesinos, M. Alfaro-Ponce, and L. Pecchia, “A review of machine learning in hypertension detection and blood pressure estimation based on clinical and physiological data,” *Biomedical Signal Processing and Control*, vol. 68, p. 102813, Jul. 2021, doi: 10.1016/j.bspc.2021.102813.

[15] D. Lafreniere, F. Zulkernine, D. Barber, and K. Martin, “Using machine learning to predict hypertension from a clinical dataset,” in *2016 IEEE Symposium Series on Computational Intelligence, SSCI 2016*, Dec. 2017, pp. 1–7.

[16] M. Fang *et al.*, “A hybrid machine learning approach for hypertension risk prediction,” *Neural Computing and Applications*, vol. 35, no. 20, pp. 14487–14497, Jul. 2023, doi: 10.1007/s00521-021-06060-0.

[17] M. H. Hung *et al.*, “Prediction of masked hypertension and masked uncontrolled hypertension using machine learning,” *Frontiers in Cardiovascular Medicine*, vol. 8, Nov. 2021, doi: 10.3389/fcvm.2021.778306.

[18] H. Zhao *et al.*, “Predicting the risk of hypertension based on several easy-to-collect risk factors: a machine learning method,” *Frontiers in Public Health*, vol. 9, Sep. 2021, doi: 10.3389/fpubh.2021.619429.

[19] Y. Yan *et al.*, “Application of back propagation neural network model optimized by particle swarm algorithm in predicting the risk of hypertension,” *Journal of Clinical Hypertension*, vol. 24, no. 12, pp. 1606–1617, Dec. 2022, doi: 10.1111/jch.14597.

[20] W. Chang *et al.*, “Prediction of hypertension outcomes based on gain sequence forward tabu search feature selection and xgboost,” *Diagnostics*, vol. 11, no. 5, p. 792, Apr. 2021, doi: 10.3390/diagnostics11050792.

[21] A. F. F. Jiménez and V. S. S. Lee, “Hipertension arterial Mexico data set,” *Kaggle*, 2024.

[22] Secretaría de salud de México ENSANUT, “Encuesta Nacional de Salud y Nutrición,” *Ensanut*, vol. 1, p. 47, 2018, [Online]. Available: https://ensanut.insp.mx/encuestas/ensanut2018/doctos/informes/ensanut_2018_presentacion_resultados.pdf.

[23] S. Gavrylenko, Z. Vladislav, and N. Khatsko, “Methods for improving the quality of classification on imbalanced data,” in *2023 IEEE 4th KhPI Week on Advanced Technology, KhPI Week 2023 - Conference Proceedings*, Oct. 2023, pp. 1–5, doi: 10.1109/KhPIWeek61412.2023.10312879.

[24] D. Elreedy, A. F. Atiya, and F. Kamalov, “A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning,” *Machine Learning*, vol. 113, no. 7, pp. 4903–4923, 2024, doi: 10.1007/s10994-022-06296-4.

[25] R. Blagus and L. Lusa, “SMOTE for high-dimensional class-imbalanced data,” *BMC Bioinformatics*, vol. 14, no. 1, p. 106, Dec. 2013, doi: 10.1186/1471-2105-14-106.

[26] D. Theng and K. K. Bhoyar, “Feature selection techniques for machine learning: a survey of more than two decades of research,” *Knowledge and Information Systems*, vol. 66, no. 3, pp. 1575–1637, Mar. 2024, doi: 10.1007/s10115-023-02010-5.

[27] A. Hashemi and M. B. Dowlatshahi, “Exploring ant colony optimization for feature selection: a comprehensive review,” in *Applications of Ant Colony Optimization and its Variants: Case Studies and New Developments*, 2024, pp. 45–60.

[28] A. A. Aouragh, M. Bahaj, and N. Gherabi, “Comparative study of dimensionality reduction techniques and machine learning algorithms for alzheimer’s disease classification and prediction,” in *2022 IEEE 3rd International Conference on Electronics, Control, Optimization and Computer Science, ICECOCS 2022*, Dec. 2022, pp. 1–6, doi: 10.1109/ICECOCS55148.2022.9983211.

[29] I. H. Sarker, “Machine learning: algorithms, real-world applications and research directions,” *SN Computer Science*, vol. 2, no. 3, p. 160, May 2021, doi: 10.1007/s42979-021-00592-x.

[30] T. Ban, M. Ohue, and Y. Akiyama, “Efficient hyperparameter optimization by using Bayesian optimization for drug-target interaction prediction,” in *IEEE International Conference on Computational Advances in Bio and Medical Sciences, ICCABS*, Oct. 2017, vol. 2017-October, pp. 1–6, doi: 10.1109/ICCABS.2017.8114299.

[31] A. Indrawati and I. N. Wahyuni, “Enhancing machine learning models through hyperparameter optimization with particle swarm optimization,” in *Proceedings - 2023 10th International Conference on Computer, Control, Informatics and its Applications: Exploring the Power of Data: Leveraging Information to Drive Digital Innovation, IC3INA 2023*, Oct. 2023, pp. 244–249, doi: 10.1109/IC3INA60834.2023.10285736.

[32] G. Naidu, T. Zuva, and E. M. Sibanda, “A review of evaluation metrics in machine learning algorithms,” in *Lecture Notes in Networks and Systems*, vol. 724 LNNS, 2023, pp. 15–25.

BIOGRAPHIES OF AUTHORS

Abd Allah Aouragh is currently pursuing his Ph.D. at the MIET Laboratory, Faculty of Sciences and Techniques, Hassan First University, Settat, Morocco. His research is focused on advancing medical diagnosis support systems, with a particular emphasis on leveraging machine learning, deep learning, and computer vision techniques to develop innovative solutions. For inquiries or collaboration opportunities, he can be contacted at email: abdallahaouragh@gmail.com.

Mohamed Bahaj received his Ph.D. in Mathematics and Computer Science from the University Hassan 1st, Morocco, and currently serves as a full professor in the Department of Mathematics and Computer Sciences at the University Hassan 1st, Faculty of Sciences and Technology, Settat, Morocco. With a robust academic background, he has contributed over 60 peer-reviewed papers, spanning areas such as Intelligent Systems, Ontologies Engineering, Partial and Differential Equations, Numerical Analysis, and Scientific Computing. He has provided valuable peer review services for various journals and mentored several Ph.D. students in computer sciences and mathematics. He actively engages in workshops, seminars, and academic forums to enhance teaching methodologies and research practices. For inquiries, he can be contacted at email: mohamedbahaj@gmail.com.

Fouad Toufik received his Ph.D. in Computer Science from the University Hassan 1st, Settat, Morocco. He currently holds the position of Professor of Computer Sciences at the Higher School of Technology SALE, Mohammed V University, Morocco. With expertise in artificial intelligence, big data, and database architectures, his research interests lie at the intersection of these fields. His academic pursuits aim to advance knowledge and innovation in these areas, contributing to the development of cutting-edge technologies. For inquiries or collaboration opportunities, he can be contacted at email: toufik.fouad@gmail.com.