Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics e-ISSN: [2656-8624](https://portal.issn.org/resource/ISSN/2656-8624)
Homepage: https://ijeeemi.org/; Vol. 7, No. 2, pp. 387-400, May 2025

RESEARCH PAPER

OPEN ACCES

Combinations of Optimization Method and Balancing Technique in Hypertension Classification with Machine Learning

Natalia Intan Suryani Lu'o[1], Daniel Febrian Sengkey[1,2,3], and Victor Florencia Ferdinand Joseph[4,5,6]

[1] Department of Electrical Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Manado, Indonesia
[2] Results Analysis, Bioinformatics, and BioBank Data Team, BioMolecular Laboratory, Universitas Sam Ratulangi, Manado, Indonesia
[3] Department of Research, Development, and Innovation, Indonesian Artificial Intelligence Society, Jakarta, Indonesia
[4] Department of Cardiology and Vascular Medicine, Faculty of Medicine, Universitas Sam Ratulangi, Manado, Indonesia
[5] Department of Cardiology and Vascular Medicine, R.D Kandou Central General Hospital, Manado, Indonesia
[6] Hypertension Working Group, Indonesian Heart Association, Jakarta, Indonesia

ABSTRACT

Hypertension is a condition in which blood vessels experience continuous pressure higher than normal limits which can cause pain and even death. Hypertension is classified into several classes based on the measured blood pressure. To correctly diagnose hypertension is a critical task that requires medical specialists who are unfortunately not evenly distributed in every region. This research aims to implement Particle Swarm Optimization for hyperparameter tuning in machine learning algorithms in hypertension disease classification. This approach is developed by comparing the performance of Random Forest (RF), Light Gradient Boosting Machine (LGBM), and Extra Trees (ET). Each algorithm was trained using hyperparameters tuned based on previous research literature, tuned with Grid Search and Cross-validation (GSCV), and Particle Swarm Optimization with Cross-validation (PSO-CV). Several evaluation metrics were used in this study, such as precision, recall, F1-score, ROC-AUC, PR-AUC, but due to the case of data imbalance, recall became the main metric in this study. The experimental results show that the combination of LGBM and PSO-CV is the best combination of algorithms and hyperparameter optimization methods with precision, recall, F1-score, ROC-AUC, and PR-AUC values of 0.22, 0.63, 0.33, 0.79, and 0.24, respectively. Not only that, every model that uses PSO-CV is proven to be consistent and has a significant improvement, so the results of this study prove that PSO-CV can have a positive influence on model performance in this study, especially in the case of unbalanced data. In this study, it can also be seen which features affect hypertension based on the best model, namely LGBM + PSO-CV, so that it can be a literacy for readers and a reference for feature work.

PAPER HISTORY
Received Feb. 10, 2025
Accepted April 30, 2025
Published May 18, 2025

KEYWORDS
Classification;
Hyperparameter Tuning;
Hypertension;
Machine Learning;
Particle Swarm Optimization

CONTACT:
natalialuo026@student.unsrat.ac.id
danielsengkey@unsrat.ac.id
victorjoseph@unsrat.ac.id

1. INTRODUCTION

Hypertension is a condition in which blood vessels experience increased pressure above normal limits that occurs continuously, hence it is commonly referred to as high blood pressure [1]. It is categorized as a serious medical condition because it can cause complications and trigger other deadly diseases, such as coronary heart disease, heart failure, aortic dissection, stroke, and chronic kidney disease [2], [3]. The Silent Killer is a nickname that can describe hypertension because sufferers often do not realize that they have the disease before blood pressure measurements [4]. Moreover, some cases state that it is possible that a chronic hypertension patient who momentarily has normal blood pressure, is misdiagnosed if based only on the readings of the sphygmomanometer [5]. This phenomenon is called Masked Hypertension, where patients with suspected hypertension have normal blood pressure when examined [6]. There are two risk factors for hypertension, namely unchangeable factors such as age, gender, and genetics. The second is factors that arise from human lifestyle. However, it can still be changed by taking appropriate treatment, such as smoking, alcohol consumption, excessive salt consumption, obesity, stress, and lack of physical activity [7]. Knowing the factors that

Corresponding author: Daniel Febrian Sengkey, danielsengkey@unsrat.ac.id, Department of Electrical Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Jl. Kampus Unsrat, Bahu, Manado, Indonesia, 95115.
DOI: https://doi.org/10.35882/ijeeemi.v7i2.86

cause hypertension can undoubtedly make it easier for someone to take early prevention of the deadly disease. Therefore, regular blood pressure monitoring is needed to prevent the disease. In some places with limited medical facilities, further diagnosis is challenging due to the availability of the corresponding experts, who make it possible to request the appropriate advanced treatments [8], [9], [10], [11]. In the era of technological development in the Industrial Revolution 4.0, existing innovations are growing rapidly, one of which is Artificial Intelligence technology commonly known as Artificial Intelligence (AI), especially Machine Learning (ML) which can perform classification tasks with its algorithms which are known to have high accuracy so that it brings major changes in various industrial sectors, especially in the health industry [12]. Machine Learning (ML) is part of artificial intelligence commonly known as Artificial Intelligence (AI). ML is a computer science discipline that focuses on building computers or machines that can learn and have some kind of intelligence [13]. In short, machine learning is used to teach machines how to handle data efficiently [14]. In a study, it is explained that there are several tasks in machine learning which are categorized into 3 learning types, supervised, unsupervised, and reinforcement [15]. Supervised learning focuses on developing a function that associates inputs with corresponding outputs using given input-output examples. In contrast, unsupervised learning operates without supervision, identifying patterns in data independently. Meanwhile, reinforcement learning involves learning a set of decision-making rules based on feedback gathered from interactions with the environment.

Classification is one of the important tasks in supervised learning, which is the process of categorizing a set of data into a particular class from several available classes. In other words, classification is the process of building a model based on training data (*training set*), which will then be used to perform classification on test data (*test set*) or on new data that the model has never seen before. The classification task has so many algorithms, and some of the selected algorithms used for the classification task in this research include Random Forest (RF), Light Gradient Boosting (LGBM), and Extra Trees (ET). RF is an algorithm that is often used in classification problems, where it works by approaching various decision trees through majority voting to reach a final decision. This approach is also referred to as the ensemble technique [16]. LGBM itself is often used in classification tasks, especially for gradient boosting, which has several advantages such as being efficient and flexible when used on large datasets, and is known for its high accuracy compared to other algorithms [17]. Having similar characteristics to random forest, which is included in the ensemble learning method, ET was chosen in this study. The thing that distinguishes ET from RF lies in the level of randomization applied when building a decision tree. Not only that, by randomly selecting features and split points without the most optimal split based on certain metrics, the ET approach can produce more diverse models that can improve performance on some datasets [18].

Several studies have proven that the Particle Swarm Optimization (PSO) method can improve the performance of machine learning algorithms in various fields. For example, PSO has been applied for hyperparameter tuning in Arabic sentiment analysis, COVID-19 big data, and heart disease prediction and classification using a combination of PSO and Ant Colony Optimization (ACO) [19], [20], [21]. PSO also played a key role in improving the Extreme Learning Machine Method (ELM) for classifying CKD cases [22] These results show that PSO is able to optimize hyperparameters effectively and improve model accuracy.

A study in CoVID-19 analysis [20] shows that the application of PSO can improve model performance, with the best accuracy of 92.3% on MLP models optimized with PSO. However, this research only focuses on PSO without comparing it with other optimization methods. On the other hand, another study, [23], showed that Grid Search Cross-Validation (GSCV) was able to improve the accuracy of hypertension classification models by 13.7%, but this study only used limited machine learning algorithms. Meanwhile, the study "Photoplethysmography-based non-invasive blood pressure monitoring via ensemble model and imbalanced dataset processing" [24] obtained an F1-score value of 81.6% with an AUC of 0.895, but did not discuss the handling of unbalanced data that could affect model performance.

Based on previous studies, there is a gap in the application of PSO for hyperparameter tuning in hypertension classification. To the best of our knowledge, there has been no research that specifically compares the PSO method with GSCV in the context of hypertension classification by considering the handling of data imbalance and using various machine learning algorithms. Therefore, this study aims to implement PSO for hyperparameter tuning in machine learning algorithms in hypertension classification and compare the performance of models optimized using PSO-CV and GS-CV as a baseline. The contributions of this research are as follows:

1. Implement PSO for hyperparameter tuning in hypertension classification using ensemble learning-based algorithms such as RF, LGBM, and ET.
2. Comparing the performance of PSO-CV and GS-CV and proving the superiority of PSO-CV.

Corresponding author: Daniel Febrian Sengkey, danielsengkey@unsrat.ac.id, Department of Electrical Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Jl. Kampus Unsrat, Bahu, Manado, Indonesia, 95115.
DOI: https://doi.org/10.35882/ijeeemi.v7i2.86

3. Using SMOTE balancing technique and weighting techniques which are the default parameters of the algorithms used in this research simultaneously and see the best performance to be used.
4. Using stroke dataset in hypertension classification by considering relevant features as consideration.

The remainder of this paper is organized as follows: In Section 2, we present the materials as well as the methods used in this research, such as the data sources, and the steps/stages in the research. Then we present the research results in Section 3, and in Section 4 we continue with a discussion of what has been obtained in Section 3. Finally, in Section 5 we conclude this research and potential issues to be explored further.

2. MATERIALS & METHOD

The course of this research loosely adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) [25], [26]. In the initial stage, business understanding, we focus on conducting literature studies and understanding the basic concepts of theory to identify a problem to be studied. The data understanding stage involves collecting and understanding data by exploring the data to be used. The data preprocessing stage involves data cleaning, and data transformation consisting of data normalization and categorical data encoding, to ensure the data is ready to be used in machine learning model training. This stage also involves splitting the data using the train_test_split technique for the training set used to train and build the model and the test set for the final evaluation of the model. Fig. 1 shows the course of research, as adapted from the CRISP-DM.

A. Dataset

An openly available stroke prediction dataset from Kaggle is used for this research. The dataset consists of a total of 5110 rows and 12 features [27]. We focus on the classification of hypertension so that the features used as labels are hypertension features which are categorized into 2 classes, namely 0 for non-hypertensive patients and 1 for hypertensive patients. The dataset has an imbalance in the target variable, where class 0 has a data population of 90.3%, while class 1 has only 9.7% of the data. The distribution of data for training can be seen in Table 1 and the feature columns of the dataset are described in Table 2. Feature data types include numeric and categorical.

Table 1. Label distribution in the training and testing dataset. The numbers indicate an imbalance in frequency between the hypertension status.

Label	Dataset		
	Training	Testing	Total
0	3689	922	4610
1	398	100	499
	4087	1022	5109

No 9.1 In addition to unbalanced data, there are other limitations in this study, such as a relatively small dataset that increases the risk of overfitting the developed model due to the limitations of statistical analysis [28]. Another limitation is the features of the dataset that mostly come from certain demographic groups, thus underrepresenting the diversity of the population as a whole.

Based on the discussion, we found that the limitations in this study may affect the extent to which the research findings can be generalized to the wider population. For example, a model that shows high performance on this dataset may not necessarily show similar performance when applied to a significantly different population, especially since the model was designed to be more sensitive to the dataset used in this study. Therefore, cautious interpretation is required in this study, and further studies with larger and more diverse data are recommended to validate and strengthen the findings.

B. Data Preprocessing

Several tasks are performed in the data preprocessing stage, including handling missing values, where the missing values are only found in the BMI column, which is 201 rows. Since the missing values amounted to <5 of the total data used, the decision taken was to fill the empty values using median value imputation, after that handling outliers, where further handling of outlier data was only done on the BMI feature because outlier cases on the Glucose Level feature usually reflect actual diabetes cases, both type 1 and type 2. Then, encoding categorical variables was applied to all features with categorical data types in the dataset. In this study, the decision to eliminate some features did not use the feature selection method but was only done manually on less relevant features generated from feature importance, namely id, type_gender, and type_domicile. Data type handling is only done on the age column which is in the form of a float and will be converted into an integer to simplify numeric data by rounding the number to the nearest integer. The last stage before separating the data is scaling and normalization, where scaling is done on features that have numeric (integer) data types, including Age, Average Glucose Level, and BMI by using the MinMaxScaler function. When the data has been scaled, the data type of the features that have gone through the scaling stage will change from integer to float. The dataset is divided into 2 types, namely training data and testing data using a division pattern of 80:20 for the train set and test set. The split was done using the train_test_split technique and stratification to ensure that the class distribution of the dataset remains the same for each subset of data generated. This technique is particularly useful when working with unbalanced datasets.

Handling imbalanced data is a process that is not mandatory for balanced data. However, in the case of the dataset that will be used for modeling has a very extreme class imbalance in the target variable (label), where the imbalance ratio is 1:10 where class 0 (no-hypertension) as the majority class has 4611 rows of data and class 1

Corresponding author: Daniel Febrian Sengkey, danielsengkey@unsrat.ac.id, Department of Electrical Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Jl. Kampus Unsrat, Bahu, Manado, Indonesia, 95115.
DOI: https://doi.org/10.35882/ijeeemi.v7i2.86

(hypertension) as the minority class only has 498 rows of data. Handling data imbalance is necessary in cases of extreme data imbalance to prevent bias in the model to be built, where the model learns more from the majority class so that it loses information for the minority class [29]. This process is done using the Synthetic Minority Over-sampling Technique (SMOTE) with the best sampling strategy of 75% and using the default parameters of the algorithm such as class_weight and scale_pos_weight. SMOTE is only applied to the training data to avoid the occurrence of leakage data that will affect model evaluation on data that has never been seen. SMOTE works by generating synthesized data for minority classes through interpolation of existing data, thus supporting improved underrepresentation of classes without simply repeating the same data or duplicating it. On the other hand, the workings of class weighting using class_weight and scale_pos_weight are to provide a greater penalty for misclassification of minority classes. This approach allows the model to focus more on the minority classes without changing the original data distribution. SMOTE was only applied to the training data to avoid data leakage that would affect model evaluation on unseen data, and SMOTE's own sampling_strategy setting was limited to 75% to avoid potential overfitting due to extreme data imbalance that could potentially generate a lot of synthesized data.

As explained in the previous paragraph, data imbalance is a condition where one class is much more dominant than the other, which can cause machine learning models to be biased towards the majority class. In the real world, especially in the context of the medical world such as hypertension classification, the majority class usually contains “non-hypertensive” cases and the minority class contains “hypertensive” cases, because a person who does not have a certain disease is usually more than a person who has a certain disease [30]. If not handled properly, models trained on such datasets may overlook patients who really need medical attention. Therefore, it is important to apply balancing techniques as well as evaluation strategies that emphasize the importance of detection in minority classes [31]. In this study, data balancing techniques such as SMOTE and algorithm default parameters such as class_weight and scale_pos_weight were used to address the imbalance between hypertensive and non-hypertensive classes. These techniques were shown to effectively improve the representation of minority classes by synthesizing new examples based on existing samples. As a result, the model becomes better able to recognize important patterns from hypertensive patient data, thus improving performance on metrics such as recall and F1-score, which are very important in a medical context.

C. Modeling and Hyperparameter Tuning

In the modeling stage, several experiments were conducted to prove the effectiveness of the PSO method with the application of machine learning algorithms used in this study, namely Random Forest, LGBM, and Extra Trees. The initial stage begins with the creation of a basic model for each algorithm that will be used as a reference for the next step. After obtaining initial insight from the application of several balancing techniques such as SMOTE, class_weight, and scale_pos_weight parameters in each algorithm, the next stage is the last part of model building in this study, namely the application of GSCV and PSO-CV methods for hyperparameter tuning of machine learning algorithms to be used.

In the experimental configuration of this study, a laptop with Windows 11 Home Insider Preview Single Language 64-bit operating system (10.0, Build 26244), Intel(R) Core(TM) i3-1005G1 processor, CPU running at 1.20GHz, and 4 GB RAM was used to run the Python 3 classifier code on both Jupyter Notebook and Google Colab. Several packages and libraries, such as Scikit Learn, sklearn, Pandas, Numpy, Pyswarm, and Matplotlib, were used to run the code during training. In this implementation, the dataset is divided into 80% for the training set and 20% for the test set.

Furthermore, the model design will be carried out using 3 machine learning algorithms, namely Random Forest (RF), Light Gradient Boosting (LGBM), and Extra Trees (ET). The reason for choosing these algorithms is because their characteristics are suitable for classification tasks, including hypertension classification. The three algorithms are ensemble learning algorithms that are able to handle some of the constraints that exist in this study, especially in the dataset used. Some of the reasons are that the algorithms are able to capture complex or non-linear patterns, do not require excessive preprocessing, resistance to overfitting, high performance on variable data, and are able to overcome data imbalance problems such as in this study.

In this research, hyperparameter tuning for machine learning algorithms using PSO is carried out and then compared with GS as a baseline model. In the model training stage, the training data is implemented in the 5-fold cross-validation method for all machine learning algorithms. This is done so that there is no need to explicitly separate the dataset into validation subsets because validation is done iteratively in each fold. The reason for choosing 5 folds is because the dataset used includes data that is not too large, and to avoid the lack of quality of model evaluation if using more folds, the data selected is only 5 folds which is also based on previous studies. Not only that, small folds provide a fairly accurate evaluation while still being efficient in terms of computing time, especially if the dataset is small and the model used requires a considerable amount of training time. By dividing the data into 5 folds, each data will be used once as test data and 4 times as training data. This strategy is applied to reduce the bias that may occur if relying on only

Corresponding author: Daniel Febrian Sengkey, danielsengkey@unsrat.ac.id, Department of Electrical Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Jl. Kampus Unsrat, Bahu, Manado, Indonesia, 95115.
DOI: https://doi.org/10.35882/ijeeemi.v7i2.86

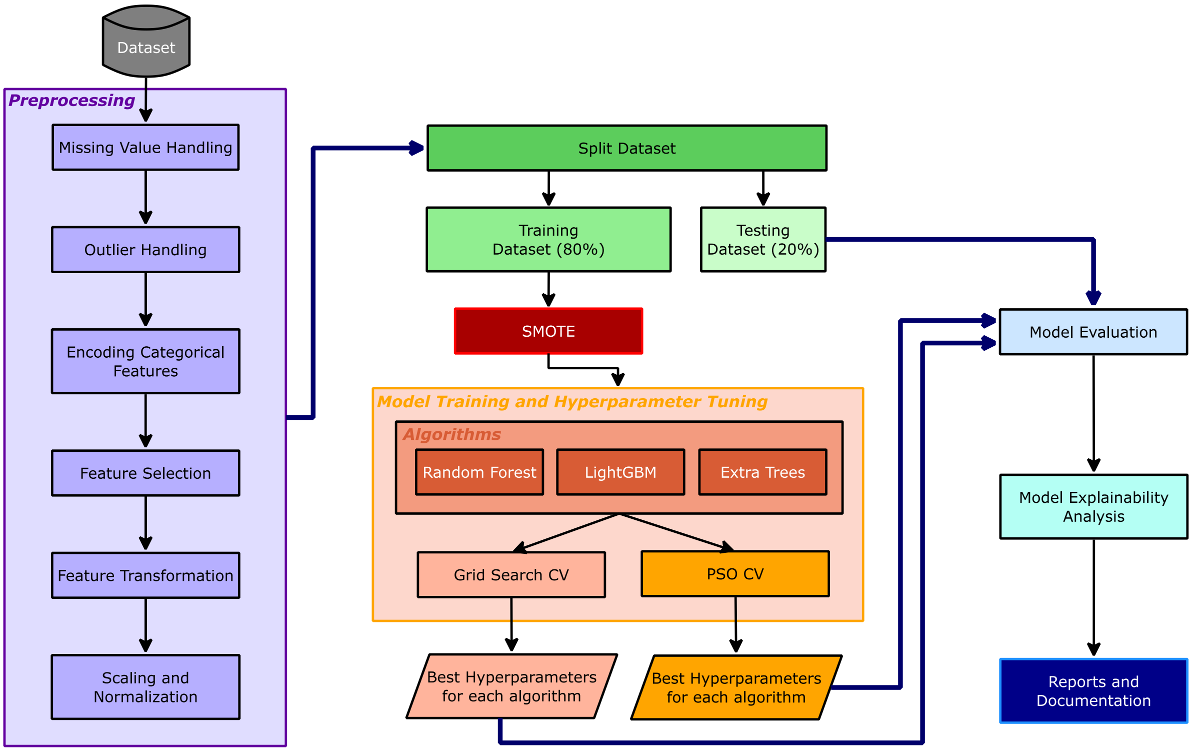

Fig. 1. The course of research. The heart of this work lies in the peach/orange shaded box, where Particle Swarm Optimization with Cross-Validation (PSO CV) is employed for hyperparameter tuning.

one set of test data, as well as provide more stable and reliable results..

Hyperparameter tuning is the process of finding the best and optimal parameters of a machine learning model to obtain better performance from the built model [32]. In the context of machine learning, a hyperparameter is a parameter whose value is set before starting the learning process, which has a difference from the parameter value obtained by default when training [19]. Regarding the choice of hyperparameters and their ranges, these are usually determined based on domain knowledge, previous research findings, and experiments [33]. The method in this study uses GSCV and PSO-CV, to thoroughly test each hyperparameter combination in the search space. The reason for the selection of optimization methods in this study is, for GSCV itself is chosen because it is simple and provides a comprehensive overview, suitable for small search spaces, and is often used in previous studies. And for PSO-CV is chosen because it is flexible and efficient in exploring a large search space so that it can save more time than GSCV. To ensure the generalization of the best-performing hyperparameters during training, 5-fold cross-validation is used [34]. Table lists the application of hyperparameter tuning to each algorithm with different models and the search space used for hyperparameter tuning.

D. Evaluation

Model evaluation is carried out using test data or data that has never been seen during training. The model evaluation stage uses evaluation metrics such as accuracy, precision, recall, and F1-score, using ROC-AUC and PR-AUC as an addition. In addition to calculations, analysis is also carried out to determine whether there is a tendency to misclassify classes with other classes using a confusion matrix [35]. In this research, we use:

1) Precision, to measure the number of correct positive predictions compared to all positive predictions made by the model. Measuring with the formula thus shows how reliable the model is in classifying the sample as positive without error, Eq. (1) ;

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}} \tag{1}$$

2) Recall, to measure how well the model identifies positive samples by comparing the number of true positives to the total actual positives. This metric is important in cases where negative misclassification has more impact than positive, such as in disease detection, Eq. (2) [36];

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \tag{2}$$

Corresponding author: Daniel Febrian Sengkey, danielsengkey@unsrat.ac.id, Department of Electrical Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Jl. Kampus Unsrat, Bahu, Manado, Indonesia, 95115.
DOI: https://doi.org/10.35882/ijeeemi.v7i2.86

3) F1-score, the harmonic mean of precision and recall, providing a balance between the two, especially on unbalanced data, Eq. (3) [37];

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

4) Receiver Operating Characteristic: Area Under the Curve: ROC-AUC is a metric that measures the classification model's ability to distinguish between positive and negative classes, with a value of 1 indicating perfect separation and 0.5 indicating performance equivalent to a random guess [38]. ROC-AUC is a graph used to show the performance of the classification model at various threshold values. This graph plots the following 2 things. True Positive Rate (TPR) or recall: this measures the proportion of positives that are correctly predicted by the model, also known as sensitivity, Eq. (4) [37];.

$$TPR = \frac{TP}{Actual\ Positive} = \frac{TP}{TP+FN} \quad (4)$$

False Positive Rate (FPR): measures the proportion of negatives that are predicted as positives. FPR can be calculated as, Eq. (5) [37];:

$$FPR = \frac{FP}{Actual\ Negative} = \frac{FP}{TN+FP} \quad (5)$$

where TP (True Positive) is the frequency of positive cases correctly predicted by the model, TN (True Negative) is the frequency of negative cases correctly predicted by the model, FP (False Positive) is the frequency of negative cases that the model incorrectly predicted as positive, and FN (False Negative) is the frequency of positive cases the model incorrectly predicted as negative.

5) Precision-Recall: Area Under the Curve: PR-AUC is a metric that measures the performance of the model on imbalanced data by calculating the area under the Precision-Recall curve, which is more informative than ROC-AUC in scenarios with minority classes. The higher the PR-AUC value, the better the model is at distinguishing positive from negative classes at various thresholds [38].

Table 2. Data type and descriptions for each feature in the dataset.

Feature	Type	Description
Id	Numeric	Unique identifier
Gender	Categorical	"Male", "Female" or "Other"
Age	Numeric	Age of the patient
Hypertension	Categorical	0 if the patient doesn't have hypertension, 1 if the patient has hypertension
Heart_disease	Categorical	0 if the patient doesn't have any heart diseases, 1 if the patient has a heart disease
Ever_married	Categorical	"No" or "Yes"
Work_type	Categorical	"children", "Govt_jov", "Never_worked", "Private" or "Self-employed"
Residence_type	Categorical	"Rural" or "Urban"
Avg_glucose_level	Numeric	The average glucose level in the blood
BMI	Numeric	Body mass index
Smoking_status	Categorical	"formerly smoked", "never smoked", "smokes" or "Unknown"
Stroke	Categorical	1 if the patient had a stroke or 0 if not

The previous paragraph discussed the evaluation metrics used in this study, where the selection of metrics was tailored to the needs of the study to assess the performance of the classification model. In this study, all metrics provided valuable information, but recall was chosen to be the main metric as the focus of analysis because it best reflects the clinical needs in the context of hypertension classification" [39].

From the previous discussion, it has been explained how the recall metric works, where recall will measure the extent to which the model is able to identify all cases of hypertension that actually exist. This is important because high false negatives can have a serious impact on patients, where patients can experience delays in diagnosis and treatment that can lead to complications.

Table 3. The hyperparameter search that used in the tuning process with grid search (GSCV) and Particle Swarm Optimization (PSO-CV).

Model	Hyperparameter	GSCV	PSO-CV
RF	n_estimators	[100, 300, 500, 800, 1000]	[100 ~ 500]
	max_depth	[10, 20, 30, 60, None]	[1 ~ 100]
	min_sample_split	[2, 5, 10]	[2 ~ 10]
	min_sample_leaf	[1, 2, 4]	[1 ~ 4]
	bootstrap	[True, False]	-
LGBM	learning_rate	[0.01, 0.05, 0.1]	[0.01 ~ 0.1]
	n_estimators	[100, 200, 300]	[100 ~ 300]
	num_leaves	[31, 50, 100]	[31, 100]
	min_data_in_leaf	[51, 55, 60]	[20 ~ 50]
ET	n_estimators	[100, 300, 500, 800, 1000]	[100 ~ 500]
	max_depth	[10, 20, 30, 60, 80, None]	[1 ~ 100]
	min_sample_split	[2, 5, 10]	[2 ~ 10]
	min_sample_leaf	[1, 2, 4]	[1 ~ 4]

Corresponding author: Daniel Febrian Sengkey, danielsengkey@unsrat.ac.id, Department of Electrical Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Jl. Kampus Unsrat, Bahu, Manado, Indonesia, 95115.
DOI: https://doi.org/10.35882/ijeeemi.v7i2.86

Therefore, keeping the recall value high is a priority in this study. In addition to recall, other metrics such as precision and F1-score are used to balance the performance of the model, where precision measures the accuracy of positive predictions, while F1-score provides a balance between precision and recall, especially useful when there is an imbalance between the number of positive and negative cases.
ROC-AUC and PR-AUC are then used to evaluate model performance more holistically under various thresholds. ROC-AUC provides an overview of the trade-off between true positive rate and false positive rate, while PR-AUC is considered more sensitive in situations of imbalanced data as in this study, as it focuses on the relationship between precision and recall.

Misclassification has a huge impact on patient outcomes, where a high false negative will lead to the neglect of treatment in patients who are actually hypertensive, whereas a high false positive will lead to unnecessary or unneeded treatment [40], although the clinical consequences tend to be less severe than a high false negative. Therefore, metrics such as recall are relevant to ensure that the model has a high sensitivity to hypertension cases even if other metrics are considered.

3. RESULT

A. Balanced Dataset

Table 4 shows the balanced labels in the training dataset after SMOTE is applied. The ratio of positive hypertension cases to negative hypertension cases has been improved, from 1:9.27 to 1:1.33. The balancing was not pushed further to 1:1 despite its theoretical plausibility to preserve the authenticity of the data fed to the algorithms.

Table 4. Comparison of the label distribution in the training before and after applying SMOTE.

Label	Before SMOTE	After SMOTE
0	3689	3689
1	398	2766

B. Model Performance

The model evaluation results in this study are presented as a whole in one section to facilitate comparative analysis of model performance. The baseline model is used as an initial reference to assess the performance of the algorithm before balancing techniques and hyperparameter optimization are performed. In the baseline model, the results show that the ET al algorithm has a more stable performance than RF and LGBM because it has a better trade-off between recall and F1-score. The PR-AUC of the ET model is also close to the PR-AUC of the RF algorithm, although the ROC-AUC is slightly lower.

Furthermore, the application of balancing techniques was carried out with two methods, namely with default parameters (class_weight for RF and ET and scale_pos_weight for LGBM) and SMOTE with the best sampling_strategy of 75%. The evaluation results showed that the RF model with SMOTE had a significant increase in recall from 0.06 to 0.45 and F1-score from 0.10 to 0.34, although the PR-AUC had a slight decrease from 0.25 to 0.23. The highest improvement was achieved by the LGBM model with scale_pos_weight, where recall increased from 0.07 to 0.49, so the F1-score also increased from 0.11 to 0.32. Meanwhile, the ET model with SMOTE experienced an increase in recall from 0.10 to 0.35, but precision decreased from 0.31 to 0.23, resulting in a relatively small increase in F1-score from 0.15 to 0.28.

Table 5. Metric evaluation results for all models for each algorithm.

Model	Precision	Recall	F1-score	ROC-AUC	PR-AUC
Baseline RF	0.32	0.06	0.10	0.77	0.25
RF + SMOTE	0.27	0.45	0.34	0.77	0.23
RF + GSCV	0.24	0.34	0.28	0.77	0.22
RF + PSO-CV	0.26	0.43	0.32	0.77	0.22
Baseline LGBM	0.24	0.07	0.11	0.78	0.22
LGBM + scale_pos_weight	0.23	0.49	0.32	0.77	0.22
LGBM + GSCV	0.25	0.43	0.32	0.79	0.24
LGBM + PSO-CV	0.22	0.63	0.33	0.79	0.24
Baseline ET	0.37	0.13	0.19	0.76	0.25
ET + SMOTE	0.23	0.35	0.28	0.75	0.22
ET + GSCV	0.50	0.01	0.02	0.75	0.22
ET + PSO-CV	0.21	0.36	0.27	0.76	0.22

Table 3. The best combination of hyperparameter values in each algorithm with the best balancing technique.

Model	HPs	GSCV	PSO-CV
RF	n_estimators	300	364
	max_depth	60	77
	min_sample_split	2	3
	min_sample_leaf	1	1
	bootstrap	False	-
LGBM	learning_rate	0.01	0.010
	n_estimators	200	207
	num_leaves	31	31
	min_data_in_leaf	55	46
ET	n_estimators	100	386
	max_depth	60	29
	min_sample_split	2	2
	min_sample_leaf	1	1

The average model has improved when balancing techniques are applied to each algorithm used, especially for recall which makes balancing techniques proven to have a major influence in improving model performance. The tuned model also proves to be more optimal, which makes the model not only in the sub-optimal configuration. The reason why the LGBM + PSO-CV model is superior to other models is because LGBM has

Corresponding author: Daniel Febrian Sengkey, danielsengkey@unsrat.ac.id, Department of Electrical Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Jl. Kampus Unsrat, Bahu, Manado, Indonesia, 95115.
DOI: https://doi.org/10.35882/ijeeemi.v7i2.86

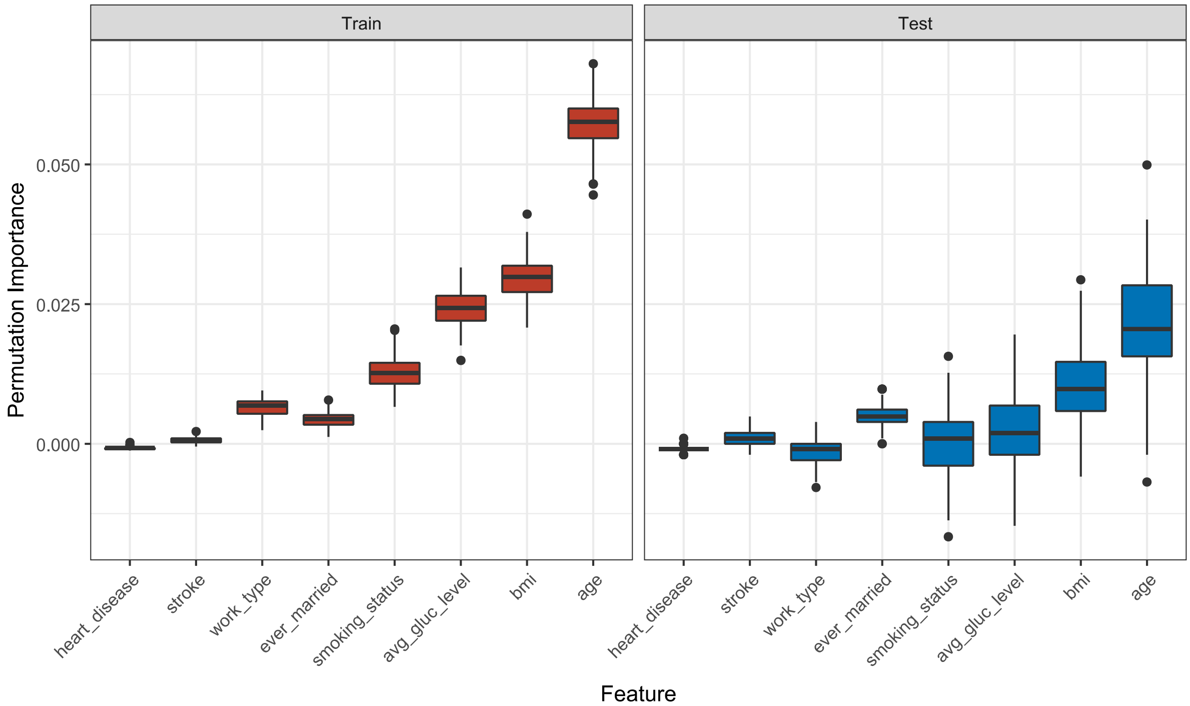

Fig. 2. Effects of shuffling a particular feature on the model performance (averaged). The x-axis indicates how much the performance degraded if the respective feature is shuffled.

a natural ability to handle unbalanced data by adjusting the weights using scale_pos_weight. Not only that, the characteristics of the dataset used also support boosting models so that LGBM, as a tree-based boosting model, is very suitable for handling complex interactions between features. For the optimization method itself, PSO-CV proved to be superior to GSCV because of the way it explores the hyperparameter space globally and efficiently and does not only search based on a grid like GSCV.

For the final stage, hyperparameter tuning is performed using Grid Search Cross-Validation (GSCV) and Particle Swarm Optimization Cross-Validation (PSO-CV) with the best balancing technique from the previous stage. The RF+PSO-CV model shows an increase in recall from 0.06 to 0.43 but is still lower than the RF model with SMOTE without hyperparameter tuning. The LGBM+PSO-CV model achieves the best performance, with a recall of 0.63, surpassing that of the LGBM+GSCV model, albeit with a slight decrease in precision. The ET model showed less stable results, where recall increased from 0.10 to 0.35 after the application of the best balancing, but decreased drastically in the ET+GSCV model to 0.01, despite remaining stable in the ET+PSO-CV model.

All models' performances are shown in Table which summarizes the performance of each method used in this study. Based on the performance, it can be seen that the combination of balancing techniques and hyperparameter optimization can significantly improve model performance, especially in the LGBM algorithm with the PSO-CV method. The last part is the result of the best hyperparameter combination obtained when training on each algorithm used in this study using GSCV and PSO-CV which can be seen in Table 6.

4. DISCUSSION

A. Effects of Hyperparameter Optimization on Model Performance

Hyperparameter optimization is an important stage in improving the performance of machine learning models, where in this study a comparison between Grid Search CV (GSCV) and Particle Swarm Optimization CV (PSO-CV) is conducted. Experimental results show that PSO-CV successfully produces models with better performance than GSCV, especially in improving recall and F1-score. This improvement can be attributed to PSO-CV's ability to explore a wider hyperparameter space and avoid local minimum traps, thus finding more optimal hyperparameter combinations. For example, the LGBM model optimized with PSO-CV shows a more significant recall improvement compared to the results from GSCV, demonstrating the effectiveness of swarm

Corresponding author: Daniel Febrian Sengkey, danielsengkey@unsrat.ac.id, Department of Electrical Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Jl. Kampus Unsrat, Bahu, Manado, Indonesia, 95115.
DOI: https://doi.org/10.35882/ijeeemi.v7i2.86

Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics e-ISSN: 2656-8624
Homepage: https://ijeeemi.org/; Vol. 7, No. 2, pp. 387-400, May 2025

intelligence-based optimization methods in handling imbalanced datasets.

In addition, the performance of the model optimized with PSO-CV is also more stable than that of GSCV, as indicated by a more consistent F1-score value across experiments. This indicates that PSO-CV is not only able to improve model performance in certain scenarios but also produces a more generalized model. This is in line with previous research which suggests that population-based approaches such as PSO can be more adaptive to complex dataset characteristics. Thus, the results of this study strengthen the evidence that using PSO-CV as a hyperparameter optimization method can be a more effective alternative to conventional approaches such as GS-CV, especially in the case of classification with unbalanced class distribution.

In this study, there are potential limitations of the hyperparameter tuning methods used, where for GSCV has challenges in time efficiency where it requires more time because of the way it works that tries all combinations of parameters in the specified grid, which can reduce the quality of tuning results because the best possibility that could not be evaluated [41]. And for PSO-CV itself does not have so many challenges in this study, only a little fluctuation in the results and is still reasonable because the performance gap is small. This is because PSO is scholastic or probability-based, which even though the parameters are the same, the final result can be slightly different between iterations because the particles are initialized randomly [42].

In the research entitled “Disease prediction via Bayesian hyperparameter optimization and ensemble learning” [43], it is said that the Bayesian Optimization (BO) optimization method is more stable and accurate than other optimization methods, such as grid search and random search. This is evidenced in the classification of breast cancer and cardiovascular disease, where BO manages to tune the XGBoost parameters efficiently and consistently with smaller variations in results in repeated tests. Not only that, there are also other studies comparing BO and PSO, where BO is often superior to PSO in terms of search efficiency and fewer iterations to achieve optimal performance, especially in low-dimensional search spaces and when model evaluation is expensive. However, PSO is faster in global exploration, more parallelizable, and excels in high-dimensional problems [44].

B. Model Explainability

An explainable model is essential, especially in clinical situations, where with an understanding of the inner workings of the model, physicians could make an informed decision about whether to use the model’s recommendation [45]. This can be achieved by using

Table 4. Results of the Shapiro-Wilk test for normality distribution of the feature importance for each dataset used in model development and evaluation.

Feature	Train		Test	
	W	p	W	p
age	0.985	0.326	0.990	0.670
avg_gluc_level	0.993	0.892	0.994	0.962
bmi	0.991	0.803	0.991	0.785
ever_married	0.985	0.363	0.965	*0.009*
heart_disease	0.844	*< 0.001*	0.702	*< 0.001*
smoking_status	0.987	0.499	0.987	0.453
stroke	0.962	*0.006*	0.925	*< 0.001*
work_type	0.977	0.088	0.979	0.112

Table 5. Results of the unpaired t-test for features with normally distributed importance.

Feature	group1	group2	t	p
age	Train	Test	33.662	*< 0.001*
avg_gluc_level	Train	Test	30.338	*< 0.001*
bmi	Train	Test	25.663	*< 0.001*
smoking_status	Train	Test	19.103	*< 0.001*
work_type	Train	Test	25.945	*< 0.001*

Table 6. The Wilcoxon Signed Rank test results for the features with partial or not normally distributed importance.

Feature	group1	group2	T	p
ever_married	Train	Test	4536	0.257
heart_disease	Train	Test	5197	0.619
stroke	Train	Test	3973	*0.011*

feature importance analysis, which measures how each feature affects the overall performance of the model [45], [46]. In this study, we apply the Permutation Feature Importance (PFI) [47]. Fig. 2 shows the importance of each feature in the best model, the LGBM+PSO-CV. It shows that by shuffling the age feature, the model performance is degraded by 0.2 on average, which is the largest of all features. This implies that age is the most influential feature for the LGBM+PSO-CV model. This can be seen as comparable to several studies in hypertension that, in the research design, divided the subjects based on age groups [48], [49]. The next two important features, the BMI and blood glucose level, have a far lower impact on the model. This can be attributed to the different impacts of each risk factor, depending on the age group [48], [50], [51], [52]. The last prominent feature is the smoking status, which is related to the patient’s smoking habit. Smoking is a known modifiable risk factor for hypertension [53]. A study found that medications for hypertension patients with a smoking habit have different effectiveness than those for non-smoking patients. By

Corresponding author: Daniel Febrian Sengkey, danielsengkey@unsrat.ac.id, Department of Electrical Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Jl. Kampus Unsrat, Bahu, Manado, Indonesia, 95115.
DOI: https://doi.org/10.35882/ijeeemi.v7i2.86

comparing the importance of each feature to the related medical studies, we can conclude that the best model we built in this study, LGBM+PSO-CV, complies with the findings in the medical field.

We further analyze the trends in the feature importance statistically. For inference statistics, distribution normality assumption has to be established before choosing the statistical method for comparing these quantities. The Shapiro-Wilk test is used to check this assumption on each feature, in both the training and testing datasets. The results are shown in Table 7. With α = 0.05, we can see that the importance values of heart_disease and stroke are not distributed normally in both datasets, while the same situation can also be observed with the ever_maried feature, but only in the testing dataset. Therefore, a non-parametric statistical test will be used to compare the importance of each particular feature in the training and testing datasets, while the other features will be analyzed with a parametric method. In Table 8, we present the results of the t-test of age, avg_gluc_level, bmi, smoking_status, and work_type, between training and testing. The p-values indicate that the feature importances are significantly different between the two groups of data. Meanwhile, for features analyzed with the Wilcoxon Signed Rank test, as shown in Table 9. Only the stroke feature shows a statistically significant difference between training and testing.

These statistical results should be interpreted independently, aside from their particular order, as shown in Fig. 2. The boxplots in Fig. 2 shows that for each particular feature, the importance quantity tends to be higher when evaluated with the training dataset, due to the internal parameters of the model itself being attuned to the training data. It can also be observed that the first three features with the highest importance quantities are consistent in training and testing, while smoking_status and ever_married, as well as stroke and work_type rankings are switched.

Improving model performance especially in terms of recall is important in the medical world, where in the context of hypertension a high recall indicates that the model is able to identify more patients who actually suffer from hypertension, thus reducing the possibility of false negative detection. This is crucial, because patients with hypertension who are not detected early can be at risk of developing serious complications and triggering other deadly diseases such as coronary heart disease, heart failure, aortic dissection, stroke, and chronic kidney disease [2], [3].

Thus, the application of an optimized model in this area, namely recall, can strengthen a more precise and responsive clinical decision-making process. Medical personnel can utilize the model output as an additional tool for screening, especially in times when resources are limited or when workload is very high. In addition, improving the model's ability to identify high-risk cases has the potential to expedite medical action, increase patient adherence to therapy, and improve long-term health outcomes.

Overall, these findings highlight the importance of developing predictive models that are not only accurate in general, but also sensitive in finding positive cases, especially in chronic diseases such as hypertension that require immediate treatment.

5. CONCLUSION

In this study, we implement one of the optimization methods, namely Particle Swarm Optimization (PSO) for hyperparameter tuning machine learning algorithms that focus on classification tasks, intending to see the performance of PSO by comparing it with other optimization methods, namely Grid Search. The class distribution of a dataset is crucial in machine learning, where the performance of the model built is based on the condition of the data used. The application of balancing techniques such as oversampling using SMOTE and weighting techniques using the default parameters of the algorithm used, such as class_weight for Random Forest and Extra Trees, and scale_pos_weight for Light Gradient Boosting (LGBM) has proven unable to produce models with satisfactory performance. Extreme data imbalance, limited sample datasets used, and weak feature correlation make the built model unable to classify data optimally. However, in this study, PSO has a big share in improving the performance of the model built, where in each experiment using PSO, it is proven that the minority class has an increase when compared to models that do not use PSO, which can be seen during model evaluation. For the best results in this study that focus on the recall value that adapts to medical needs, obtained from a model that uses the LGBM algorithm with the scale_pos_weight parameter as a balancing technique, as well as the use of PSO for hyperparameter tuning of the algorithm used, namely with a recall value of 0.63 which is slightly better than other models.

Moreover, this study also obtained the feature importance of the best model using Permutation Feature Importance, where the age and BMI features are the most influential features, followed by the avg_gluc_level and smoking_status features which have moderate influence, as well as the stroke and ever_married features which have little influence, to features that have no effect such as work_type and heart_disease.

Despite the problem of data imbalance, which is the main problem in this study, PSO can still be said to be the best optimization method that can improve the performance of the built model when compared to Grid

Corresponding author: Daniel Febrian Sengkey, danielsengkey@unsrat.ac.id, Department of Electrical Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Jl. Kampus Unsrat, Bahu, Manado, Indonesia, 95115.
DOI: https://doi.org/10.35882/ijeeemi.v7i2.86

Search, which is famous for its consistency and accuracy. Future research can consider using more optimization methods, such as random search, Bayesian optimization, genetic algorithm, and other optimization methods that have their advantages to compare with PSO [54]. Not only optimization methods, but also exploration of additional machine learning techniques, such as ensemble learning or deep learning, might provide further improvements in model accuracy and robustness [55], [56]. In addition, future research can also use datasets that are better distributed and have balanced classes, and can consider the integration of additional data such as medical history, lifestyle, or genetic data to improve predictive capabilities.

REFERENCES

[1] P. Muntner *et al.*, "Trends in Blood Pressure Control Among US Adults With Hypertension, 1999-2000 to 2017-2018," *JAMA*, vol. 324, no. 12, p. 1190, Sep. 2020, doi: 10.1001/jama.2020.14545.

[2] F. D. Fuchs and P. K. Whelton, "High Blood Pressure and Cardiovascular Disease," *Hypertension*, vol. 75, no. 2, pp. 285–292, 2020, doi: 10.1161/HYPERTENSIONAHA.119.14240.

[3] Y. Wu, B. Xin, Q. Wan, Y. Ren, and W. Jiang, "Risk factors and prediction models for cardiovascular complications of hypertension in older adults with machine learning: A cross-sectional study," *Heliyon*, vol. 10, no. 6, p. e27941, 2024, doi: 10.1016/j.heliyon.2024.e27941.

[4] G. A. Mensah, "Commentary – Hypertension Phenotypes: The Many Faces of a Silent Killer," *Ethn Dis*, vol. 29, no. 4, pp. 545–548, Oct. 2019, doi: 10.18865/ed.29.4.545.

[5] S. S. Franklin *et al.*, "The Cardiovascular Risk of White-Coat Hypertension," 2016. doi: https://doi.org/10.1016/j.jacc.2016.08.035.

[6] D.-Y. Zhang *et al.*, "Treatment of Masked Hypertension with a Chinese Herbal Formula," *Circulation*, vol. 142, no. 19, pp. 1821–1830, Nov. 2020, doi: 10.1161/CIRCULATIONAHA.120.046685.

[7] T. Ojangba *et al.*, "Comprehensive effects of lifestyle reform, adherence, and related factors on hypertension control: A review," *The Journal of Clinical Hypertension*, vol. 25, no. 6, pp. 509–520, Jun. 2023, doi: 10.1111/jch.14653.

[8] Y. Wu *et al.*, "Provider competence in hypertension management and challenges of the rural primary healthcare system in Sichuan province, China: a study based on standardized clinical vignettes," *BMC Health Serv Res*, vol. 22, no. 1, p. 849, Dec. 2022, doi: 10.1186/s12913-022-08179-9.

[9] O. Adejumo *et al.*, "Assessment of hypertension service availability in some primary health centres in Nigeria: a mixed-methods study," *BMJ Open*, vol. 13, no. 8, p. e073833, Aug. 2023, doi: 10.1136/bmjopen-2023-073833.

[10] M. A. Peters *et al.*, "Barriers to effective hypertension management in rural Bihar, India: A cross-sectional, linked supply- and demand-side study," *PLOS Global Public Health*, vol. 2, no. 10, p. e0000513, Oct. 2022, doi: 10.1371/journal.pgph.0000513.

[11] T. N. Hoang *et al.*, "Assessment of availability, readiness, and challenges for scaling-up hypertension management services at primary healthcare facilities, Central Highland region, Vietnam, 2020," *BMC Primary Care*, vol. 24, no. 1, p. 138, Jul. 2023, doi: 10.1186/s12875-023-02092-8.

[12] A. Barragán-Montero *et al.*, "Artificial intelligence and machine learning for medical imaging: A technology review," *Physica Medica*, vol. 83, no. December 2020, pp. 242–256, Mar. 2021, doi: 10.1016/j.ejmp.2021.04.016.

[13] M. Soori, B. Arezoo, and R. Dastres, "Artificial intelligence, machine learning and deep learning in advanced robotics, a review," *Cognitive Robotics*, vol. 3, pp. 54–70, Jan. 2023, doi: 10.1016/j.cogr.2023.04.001.

[14] B. Mahesh, "Machine Learning Algorithms - A Review," *Internation Journal of Science and Research (IJSR)*, vol. 9, no. 1, pp. 381–386, 2020, doi: 10.21275/ART20203995.

[15] N. Sharma, R. Sharma, and N. Jindal, "Machine Learning and Deep Learning Applications-A Vision," *Global Transitions Proceedings*, vol. 2, no. 1, pp. 24–28, Jun. 2021, doi: 10.1016/j.gltp.2021.01.004.

[16] M. Azad, T. H. Nehal, and M. Moshkov, "A novel ensemble learning method using majority based voting of multiple selective decision trees," *Computing*, vol. 107, no. 1, p. 42, Jan. 2025, doi: 10.1007/s00607-024-01394-8.

[17] Y. Wang, Y. Liu, J. Zhao, and Q. Zhang, "Low-Complexity Fast CU Classification Decision Method Based on LGBM Classifier," *Electronics (Switzerland)*, vol. 12, no. 11, 2023, doi: 10.3390/electronics12112488.

[18] N. H. Cahyana, Y. Fauziah, and A. S. Aribowo, "The Comparison of Tree-Based Ensemble Machine Learning for Classifying Public Datasets," *RSF Conference Series: Engineering and Technology*, vol. 1, no. 1, pp. 407–413, Dec. 2021, doi: 10.31098/cset.v1i1.412.

Corresponding author: Daniel Febrian Sengkey, danielsengkey@unsrat.ac.id, Department of Electrical Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Jl. Kampus Unsrat, Bahu, Manado, Indonesia, 95115.
DOI: https://doi.org/10.35882/ijeeemi.v7i2.86

[19] E. Elgeldawi, A. Sayed, A. R. Galal, and A. M. Zaki, “Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis,” *Informatics*, vol. 8, no. 4, p. 79, Nov. 2021, doi: 10.3390/informatics8040079.

[20] H. S. Salem, M. A. Mead, and G. S. El-Taweel, “Particle Swarm Optimization-Based Hyperparameters Tuning of Machine Learning Models for Big COVID-19 Data Analysis,” *Journal of Computer and Communications*, vol. 12, no. 03, pp. 160–183, 2024, doi: 10.4236/jcc.2024.123010.

[21] Y. Khourdifi and M. Bahaj, “Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization,” *International Journal of Intelligent Engineering and Systems*, vol. 12, no. 1, pp. 242–252, Feb. 2019, doi: 10.22266/ijies2019.0228.24.

[22] M. M. Amini, M. I. Mazdadi, Muliadi, M. R. Faisal, and T. H. Saragih, “Implementation of Extreme Learning Machine Method with Particle Swarm Optimization to Classify of Chronic Kidney Disease,” *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 6, no. 4, pp. 499–508, Sep. 2024, doi: 10.35882/JEEEMI.V6I4.561.

[23] R. E. Al Mamlook *et al.*, “Machine Learning Models Based on Grid-Search Optimization and Shapley Additive Explanations (SHAP) for Early Stroke Prediction,” in *2024 4th Interdisciplinary Conference on Electrics and Computer (INTCEC)*, IEEE, Jun. 2024, pp. 1–7. doi: 10.1109/INTCEC61833.2024.10602984.

[24] Q. Liu, C. Yang, S. Yang, C. F. Kwong, J. Wang, and N. Zhou, “Photoplethysmography-based non-invasive blood pressure monitoring via ensemble model and imbalanced dataset processing,” *Phys Eng Sci Med*, Dec. 2024, doi: 10.1007/s13246-024-01445-6.

[25] T. Darmawan, “Credit Classification Using CRISP-DM Method On Bank ABC Customers,” *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 6, pp. 2375–2380, 2020, doi: 10.30534/ijeter/2020/28862020.

[26] N. Razali, S. Ismail, and A. Mustapha, “Machine learning approach for flood risks prediction,” *IAES International Journal of Artificial Intelligence*, vol. 9, no. 1, pp. 73–80, 2020, doi: 10.11591/ijai.v9.i1.pp73-80.

[27] F. Soriano, “Stroke Prediction Dataset,” Kaggle Datasets. [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset/data

[28] A. Althnian *et al.*, “Impact of Dataset Size on Classification Performance: An Empirical Evaluation in the Medical Domain,” *Applied Sciences*, vol. 11, no. 2, p. 796, Jan. 2021, doi: 10.3390/app11020796.

[29] A. Tajali, T. H. Saragih, M. I. Mazdadi, I. Budiman, and A. Farmadi, “The Impactness of SMOTE as Imbalance Class Handling for Myocardial Infarction Complication Classification using Machine Learning Approach with Data Imputation and Hyperparameter,” *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 6, no. 4, pp. 227–239, Nov. 2024, doi: 10.35882/IJEEEMI.V6I4.13.

[30] M. Salmi, D. Atif, D. Oliva, A. Abraham, and S. Ventura, “Handling imbalanced medical datasets: review of a decade of research,” *Artif Intell Rev*, vol. 57, no. 10, p. 273, Sep. 2024, doi: 10.1007/s10462-024-10884-2.

[31] J. M. Johnson and T. M. Khoshgoftaar, “Survey on deep learning with class imbalance,” *J Big Data*, vol. 6, no. 1, p. 27, Dec. 2019, doi: 10.1186/s40537-019-0192-5.

[32] A. E. Minarno, M. Hazmi Cokro Mandiri, Y. Munarko, and H. Hariyady, “Convolutional Neural Network with Hyperparameter Tuning for Brain Tumor Classification,” *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, vol. 4, May 2021, doi: 10.22219/kinetik.v6i2.1219.

[33] H. S. Salem, M. A. Mead, and G. S. El-Taweel, “Particle Swarm Optimization-Based Hyperparameters Tuning of Machine Learning Models for Big COVID-19 Data Analysis,” *Journal of Computer and Communications*, vol. 12, no. 03, pp. 160–183, 2024, doi: 10.4236/jcc.2024.123010.

[34] D. Berrar, “Cross-Validation,” in *Encyclopedia of Bioinformatics and Computational Biology*, vol. 1–3, Elsevier, 2019, pp. 542–545. doi: 10.1016/B978-0-12-809633-8.20349-X.

[35] W. Abdun Naseer, S. Sarwido, and B. B. Wahono, “GRADIENT BOOSTING OPTIMIZATION WITH PRUNING TECHNIQUE FOR PREDICTION OF BMT AL-HIKMAH PERMATA CUSTOMER DATA,” *Jurnal Informatika Teknologi dan Sains (Jinteks)*, vol. 6, no. 3, pp. 719–727, Aug. 2024, doi: 10.51401/jinteks.v6i3.4702.

[36] H. Dalianis, “Evaluation Metrics and Evaluation,” in *Clinical Text Mining*, Springer International Publishing, 2018, pp. 45–53. doi: 10.1007/978-3-319-78503-5_6.

[37] P. M. Vieira and F. Rodrigues, “An automated approach for binary classification on imbalanced data,” *Knowl Inf Syst*, vol. 66, no. 5, pp. 2747–

Corresponding author: Daniel Febrian Sengkey, danielsengkey@unsrat.ac.id, Department of Electrical Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Jl. Kampus Unsrat, Bahu, Manado, Indonesia, 95115.
DOI: https://doi.org/10.35882/ijeeemi.v7i2.86

2767, May 2024, doi: 10.1007/s10115-023-02046-7.
[38] A. E. Maxwell, T. A. Warner, and L. A. Guillén, “Accuracy assessment in convolutional neural network-based deep learning remote sensing studies—part 1: Literature review,” *Remote Sens (Basel)*, vol. 13, no. 13, 2021, doi: 10.3390/rs13132450.
[39] S. Montagna *et al.*, “Machine Learning in Hypertension Detection: A Study on World Hypertension Day Data,” *J Med Syst*, vol. 47, no. 1, p. 1, Dec. 2022, doi: 10.1007/s10916-022-01900-5.
[40] S. Wyatt *et al.*, “Leveraging Machine Learning to Identify Subgroups of Misclassified Patients in the Emergency Department: Multicenter Proof-of-Concept Study,” *J Med Internet Res*, vol. 26, p. e56382, Dec. 2024, doi: 10.2196/56382.
[41] I. Syarif, A. Prugel-Bennett, and G. Wills, “SVM Parameter Optimization using Grid Search and Genetic Algorithm to Improve Classification Performance,” *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 14, no. 4, p. 1502, Dec. 2016, doi: 10.12928/telkomnika.v14i4.3956.
[42] Q. Li, S.-Y. Liu, and X.-S. Yang, “Influence of initialization on the performance of metaheuristic optimizers,” *Appl Soft Comput*, vol. 91, p. 106193, Jun. 2020, doi: 10.1016/j.asoc.2020.106193.
[43] L. Gao and Y. Ding, “Disease prediction via Bayesian hyperparameter optimization and ensemble learning,” *BMC Res Notes*, vol. 13, no. 1, p. 205, Dec. 2020, doi: 10.1186/s13104-020-05050-0.
[44] L. Tani and C. Veelken, “Comparison of Bayesian and particle swarm algorithms for hyperparameter optimisation in machine learning applications in high energy physics,” *Comput Phys Commun*, vol. 294, p. 108955, Jan. 2024, doi: 10.1016/j.cpc.2023.108955.
[45] J. Petch, S. Di, and W. Nelson, “Opening the Black Box: The Promise and Limitations of Explainable Machine Learning in Cardiology,” *Canadian Journal of Cardiology*, vol. 38, no. 2, pp. 204–213, Feb. 2022, doi: 10.1016/J.CJCA.2021.09.004/ASSET/09AA3827-2FF2-4031-A434-D1CC473DD0D3/MAIN.ASSETS/GR6.JPG.
[46] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, “Explainable AI: A Review of Machine Learning Interpretability Methods,” *Entropy 2021, Vol. 23, Page 18*, vol. 23, no. 1, p. 18, Dec. 2020, doi: 10.3390/E23010018.
[47] G. A. J. Satvika, I. N. Sukajaya, and I. G. A. Gunadi, “Improving k-nearest neighbor performance using permutation feature importance to predict student success in study,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 35, no. 3, pp. 1835–1844, Sep. 2024, doi: 10.11591/ijeecs.v35.i3.pp1835-1844.
[48] Y. Zhang *et al.*, “Distribution of risk factors of hypertension patients in different age groups in Tianjin,” *BMC Public Health*, vol. 21, no. 1, pp. 1–10, Dec. 2021, doi: 10.1186/S12889-021-10250-9/TABLES/4.
[49] J. G. Wang, W. Zhang, Y. Li, and L. Liu, “Hypertension in China: epidemiology and treatment initiatives,” *Nature Reviews Cardiology 2023 20:8*, vol. 20, no. 8, pp. 531–545, Jan. 2023, doi: 10.1038/s41569-022-00829-z.
[50] Y. Alhassan *et al.*, “Determinants of blood pressure and blood glucose control in patients with co-morbid hypertension and type 2 diabetes mellitus in Ghana: A hospital-based cross-sectional study,” *PLOS Global Public Health*, vol. 2, no. 12, p. e0001342, Dec. 2022, doi: 10.1371/JOURNAL.PGPH.0001342.
[51] H. Alsaadon *et al.*, “Hypertension and its related factors among patients with type 2 diabetes mellitus – a multi-hospital study in Bangladesh,” *BMC Public Health*, vol. 22, no. 1, Dec. 2022, doi: 10.1186/S12889-022-12509-1/TABLES/2.
[52] P. El Meouchy, M. Wahoud, S. Allam, R. Chedid, W. Karam, and S. Karam, “Hypertension Related to Obesity: Pathogenesis, Characteristics and Factors for Control,” *International Journal of Molecular Sciences 2022, Vol. 23, Page 12305*, vol. 23, no. 20, p. 12305, Oct. 2022, doi: 10.3390/IJMS232012305.
[53] M. Jareebi, “The Association Between Smoking Behavior and the Risk of Hypertension: Review of the Observational and Genetic Evidence,” *J Multidiscip Healthc*, vol. Volume 17, pp. 3265–3281, Jul. 2024, doi: 10.2147/JMDH.S470589.
[54] S. M. Almufti, A. Yahya Zebari, and H. Khalid Omer, “A comparative study of particle swarm optimization and genetic algorithm,” *Journal of Advanced Computer Science & Technology*, vol. 8, no. 2, pp. 40–45, Oct. 2019, doi: 10.14419/jacst.v8i2.29401.
[55] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, “Ensemble deep learning: A review,” *Eng Appl Artif Intell*, vol. 115, p. 105151, Oct. 2022, doi: 10.1016/j.engappai.2022.105151.
[56] F. Zhang, K. Li, and Z. Ren, “Improving Adversarial Robustness of Ensemble Classifiers by Diversified

Corresponding author: Daniel Febrian Sengkey, danielsengkey@unsrat.ac.id, Department of Electrical Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Jl. Kampus Unsrat, Bahu, Manado, Indonesia, 95115.
DOI: https://doi.org/10.35882/ijeeemi.v7i2.86

Feature Selection and Stochastic Aggregation," *Mathematics*, vol. 12, no. 6, p. 834, Mar. 2024, doi: 10.3390/math12060834.

AUTHOR BIOGRAPHY

Natalia Intan Suryani Lu'o was born in Tampemadoro, December 13, 2003. The author started her education at Tampemadoro Elementary School in 2009-2015, continued her education at SMPN 3 Lage in 2015-2018, and then in 2018-2021 continued her education at SMAN 2 Poso. In 2021, the author continued his undergraduate education in the Informatics Engineering Study Program, at the Department of Electrical Engineering, Faculty of Engineering, Sam Ratulangi University Manado. During lectures, the author actively participated in activities on campus and joined student organizations, namely the Electrical Student Association (HME) and the Unsrat IT Community (UNITY). Students have also taken part as teaching assistants in database courses for 4 months.

Daniel Febrian Sengkey is a lecturer at the Undergraduate Program in Informatics, Department of Electrical Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Manado, Indonesia. He graduated from the Undergraduate Program in Electrical Engineering of the same department in 2021, then in 2015 achieved his Master's degree in Electrical Engineering from the Department of Electrical Engineering and Information Technology, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia. His current teaching and research activities are mainly related to the Machine Learning fundamentals, as well as implementation, especially in Biomedical and Health Informatics. Besides his assignment in the Informatics program, he is also a member of the Bioinformatics team at the university's Biomolecular Laboratory.

Victor Florencia Ferdinand Joseph achieved his professional degree as a Medical Doctor from Universitas Sam Ratulangi in 2003. He completed his training as a cardiology and vascular medicine specialist and then as a subspecialist in cardiovascular prevention and rehabilitation in 2014 and 2020, respectively. His career as a clinician started in 2003, and since 2008, he has been involved in academic activities in the Department of Cardiology and Vascular Medicine, Faculty of Medicine, Universitas Sam Ratulangi. Dr. Victor also holds positions in several organizations, such as the Hypertension Working Group of the Indonesian Heart Association, the Code Blue Team, the Heart Surgery Team of Prof. R.D. Kandou Central General Hospital, and the Indonesian Medical Association. He is frequently invited as a speaker at various events, local and national, held by medical institutes, research centers, etc. He is also actively mentoring and supervising undergraduate students involved in competitions such as the Program Kreatifitas Mahasiswa (PKM).

Corresponding author: Daniel Febrian Sengkey, danielsengkey@unsrat.ac.id, Department of Electrical Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Jl. Kampus Unsrat, Bahu, Manado, Indonesia, 95115.
DOI: https://doi.org/10.35882/ijeeemi.v7i2.86