JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 2, September 2025, Page 266-275
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i3.861

Soft Voting Based Optimized Ensemble for Migraine Type Classification

Titik Misriati[1], Riska Aryanti[2*], Henny Leidiyana[3]
[1,2,3]Faculty of Engineering and Informatics, Indonesia
[1]titik.tmi@bsi.ac.id, [2*]riska.rts@bsi.ac.id, [3]henny.hnl@bsi.ac.id

Abstract: The accurate classification of migraine subtypes is a complex challenge in neurology, hindered by symptomatic similarities between types. This complexity necessitates advanced computational tools to support diagnostic precision. This study aims to develop and evaluate an optimized soft voting ensemble classifier to automate this multi-class classification task effectively. The methodology involved training eight base models—including Neural Network, Random Forest, and Gradient Boosting—on a publicly available migraine dataset, with an 80-20 train-test split. The top three performers were integrated into a soft voting ensemble, which aggregates their predicted probabilities to enhance decision robustness. Model performance was rigorously assessed using accuracy, precision, recall, F1-score, and AUC-ROC metrics. The results demonstrated that the proposed ensemble achieved superior performance, with an accuracy of 91.67% and an F1-score of 91.50%, outperforming all constituent models. Furthermore, the ensemble attained near-perfect AUC-ROC values across multiple classes, confirming its strong discriminatory capability. The study concludes that the soft voting ensemble is a highly effective and reliable approach for migraine subtype classification, offering significant potential as a decision-support tool in clinical environments. Future work will focus on hyperparameter optimization, explainability, and validation with larger multi-centric datasets to facilitate clinical adoption.

Keywords: Migraine Classification; Ensemble Learning; Soft Voting; Machine Learning, Clinical Decision Support

1. INTRODUCING

Migraine is a neurological condition that affects approximately 14% of the world's population and ranks second as a leading cause of disability globally [1]. The prevalence is particularly high among women in productive age groups [2],[3]. Migraines can significantly reduce quality of life, impair daily activities, and impose economic burdens on society [4], [5]. Accurate identification of migraine types is essential because each subtype requires different management and treatment strategies[6],[7].

Despite this importance, diagnosis continues to rely primarily on clinical assessment, which can be subjective, time-consuming, and occasionally inconsistent. Overlapping symptoms with other types of headaches further increase the possibility of misdiagnosis and inappropriate therapy [8]. These challenges highlight the importance of approaches based on data and artificial intelligence to improve the accuracy and reliability of migraine classification [9], [10].

Recent Advances in artificial intelligence have introduced machine learning as a promising approach in medical decision-making, including migraine classification. Prior studies have utilized models such as Naïve Bayes [11] and Support Vector Machine (SVM), the latter achieving up to 84% accuracy in distinguishing migraine phases [12]. Other works have reported over 95% accuracy using Multi-Layer Perceptron (MLP) and Gradient Boosting [13],[10]. In addition, feature selection techniques have been shown to enhance model performance [15], further improving predictive capabilities in medical datasets.

Riska Aryanti: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 2, September 2025, Page 266-275
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i3.861

Although promising, most of these studies rely on single classifiers that are vulnerable to bias and overfitting when facing imbalanced data. Ensemble learning, which integrates multiple models to provide more robust predictions, has been less extensively explored in migraine research. Methods such as the Voting Classifier have demonstrated potential to improve outcomes by leveraging the strengths of diverse base learners [16], [17], [18], [19]. However, ensemble methods remain underexplored in migraine research.

Among ensemble techniques, soft voting ensembles are considered more flexible because they consider the prediction probabilities of each base model, rather than just the final prediction class. By combining three high performance algorithms Neural Networks, Random Forests, and Gradient Boosting soft voting ensembles can produce more accurate, generalizable, and reliable predictions [20], [21].

Thus, this study shows that the Soft Voting Classifier, which integrates Neural Network, Random Forest, and Gradient Boosting, is capable of producing more reliable classification performance for migraine subtypes compared to most single algorithms. These findings not only provide a comprehensive overview of the comparison between several algorithms, but also confirm the potential of the ensemble learning approach as an effective solution to support the development of migraine prediction and classification systems in the future.

2. RESEARCH METHODOLOGY

The research framework encompassed dataset preparation, preprocessing, model training, ensemble construction, and evaluation. An overview of the workflow is shown in Fig 1.

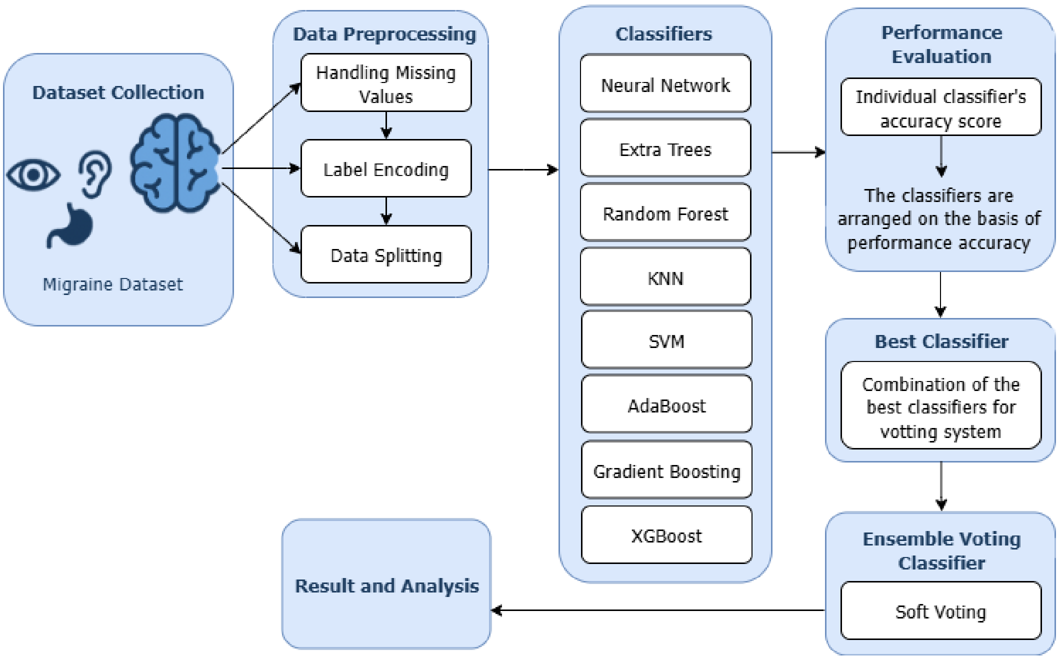

Figure 1. Proposed Research Workflow

2.1. Dataset

The dataset was obtained from Kaggle and includes 23 predictor attributes and one categorical target variable representing seven migraine subtypes. Table 1 summarizes the features.

2.2. Data Preprocessing

Prior to model training, the dataset underwent a series of preprocessing steps to ensure data quality and compatibility with the machine learning algorithms.

1. Handling Missing Values

Riska Aryanti: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 2, September 2025, Page 266-275
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i3.861

The presence of missing values was investigated. Given the nature of the dataset and after analysis, any instances with missing values were removed to maintain data integrity, as they constituted a negligible portion of the total data.

2. Label Encoding
   Categorical features within the dataset were converted into a numerical format using Label Encoding. This transformation assigns a unique integer to each category, making the data suitable for algorithms that require numerical input.
3. Data Splitting
   The preprocessed dataset was partitioned into a training set and an independent testing set using a stratified hold-out method. Specifically, 80% of the data was allocated for training the models, and the remaining 20% was reserved for testing and evaluating the final performance. Stratification was applied to ensure that the class distribution (proportions of migraine types) was preserved in both subsets, preventing sampling bias.

Table 1. Dataset Description

No	Feature	Type	Description
1	Age	Numeric	The age of the patient in years
2	Duration	Numeric	The duration of migraine attacks, categorized into 3 levels and encoded from 1 to 3
3	Frequency	Numeric	The frequency of migraine attacks, categorized into 3 levels and encoded from 1 to 3
4	Location	Numeric	The location of the headache pain, categorized into 3 levels and encoded from 1 to 3
5	Character	Numeric	The character of the pain, categorized into 3 types and encoded from 0 to 2
6	Intensity	Numeric	The intensity of the pain, categorized into 4 levels and encoded from 0 to 3
7	Nausea	Numeric	Symptoms of discomfort in the back of the throat or stomach (1: Yes, 0: No)
8	Vomit	Numeric	Actual vomiting (1: Yes, 0: No)
9	Phonophobia	Numeric	Hypersensitivity to sound, (1: Yes, 0: No)
10	Photophobia	Numeric	Hypersensitivity to light, (1: Yes, 0: No)
11	Visual	Numeric	Visual disturbances or aura, categorized into 5 conditions and encoded from 0 to 4
12	Sensory	Numeric	Sensory disturbances or aura, categorized into 3 conditions and encoded from 0 to 2
13	Dysphasia	Numeric	Language impairment
14	Dysarthria	Numeric	Motor speech difficulty
15	Vertigo	Numeric	Sensation of spinning or dizziness
16	Tinnitus	Numeric	Ringing in the ears
17	Hypoacusis	Numeric	Hearing loss
18	Diplopia	Numeric	Double vision
19	Defect	Numeric	Neurological deficit
20	Ataxia	Numeric	Lack of coordination
21	Conscience	Numeric	Impairment of consciousness
22	Paresthesia	Numeric	Sensation of tingling or "pins and needles
23	DPF	Numeric	A physical sign indicating a reduction in the distance between the upper and lower eyelids compared to the other eye, encoded as 0 (Absent) or 1 (Present). This is a rare sign associated with specific neurological conditions

Riska Aryanti: *Corresponding Author

24	Type	Categorical	This is the target variable for the multi-class classification task. It represents the diagnosed subtype of migraine. The label is nominally encoded into 7 distinct types

2.3. Model

A diverse suite of machine learning algorithms was selected for this study to ensure a comprehensive evaluation and to build a robust ensemble. The selection encompasses neural networks, tree-based ensembles, instance-based learners, and support vector machines, capturing a wide array of inductive biases.

1. Neural Network
   The Multi-Layer Perceptron (MLP) represents a type of feedforward artificial neural network (ANN). Its structure generally comprises an input layer, several hidden layers with non-linear activation, and an output layer. In this architecture, every neuron is fully connected to neurons in the next layer. Model training is carried out through backpropagation, combined with optimization methods such as Adam or Stochastic Gradient Descent (SGD), to minimize the loss function. The incorporation of non-linear activation functions enables the network to capture and model intricate, non-linear decision boundaries.
2. Extra Trees Classifier
   The Extremely Randomized Trees (Extra Trees) algorithm is a tree-based ensemble similar to Random Forest, but with additional randomization. Unlike Random Forest, which selects the best split based on criteria like Gini impurity or information gain, Extra Trees chooses split points randomly. Furthermore, it often grows trees on the full dataset rather than bootstrap samples. This design makes training faster and helps reduce variance, although sometimes at the cost of slightly higher bias.
3. Random Forest
   Random Forest is an ensemble approach based on bagging (Bootstrap Aggregating). It builds multiple decision trees, each trained on a randomly sampled subset of the dataset. At each split, a random selection of features is considered to diversify the trees. Final predictions are determined by majority voting in classification or by averaging in regression tasks. The combination of randomness in sampling and feature selection reduces overfitting and enhances predictive robustness.
4. KNN
   KNN is an instance-based learning method that predicts class labels based on the majority class of the k nearest neighbors in the feature space. It is simple and effective for smaller datasets but can become computationally demanding with larger datasets because it stores all training instances and requires distance calculations for each prediction.
5. SVM
   SVM is a powerful supervised learning algorithm that identifies an optimal hyperplane to separate classes with the maximum margin. For cases where data are not linearly separable, kernel functions map the input features into higher-dimensional spaces, enabling effective separation. SVM is particularly effective for problems with high-dimensional data.
6. AdaBoost
   Adaptive Boosting (AdaBoost) is a boosting technique that sequentially trains multiple weak classifiers, typically shallow decision trees. Each iteration adjusts weights by focusing more on misclassified samples from previous rounds. The final decision is made through a weighted vote, where classifiers with higher accuracy receive greater influence. This mechanism reduces bias and improves classification performance.
7. Gradient Boosting
   Gradient Boosting constructs an ensemble sequentially, where each new learner attempts to correct the errors of the preceding models. Instead of adjusting weights like AdaBoost, it builds learners to

Riska Aryanti: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 2, September 2025, Page 266-275
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i3.861

predict the residuals of the loss function's gradient. This flexibility allows optimization for a wide variety of differentiable loss functions. Gradient Boosting is known for its strong predictive accuracy and ability to balance bias and variance.

8. XGBoost
   XGBoost is an optimized extension of Gradient Boosting, designed for speed and scalability. It incorporates regularization to prevent overfitting and efficient handling of missing values. Parallelized tree construction and advanced optimization techniques make it significantly faster while maintaining high accuracy, which explains its popularity in machine learning competitions and applied research.

2.4. Soft Voting Classifier

Following a rigorous comparative analysis of eight candidate machine learning algorithms, the three top-performing models—Gradient Boosting, Random Forest, and a Neural Network—were selected for integration into an ensemble framework. To harness their collective predictive capability and methodological diversity, a Soft Voting Classifier was employed in Fig 2. This technique operates by aggregating the class probability estimates from each constituent model, as opposed to executing a simplistic majority (hard) vote. This sophisticated synthesis capitalizes on the distinct advantages inherent to each algorithm: the superior predictive accuracy of Gradient Boosting, the inherent robustness and variance reduction of Random Forest, and the capacity of the Neural Network to model intricate, non-linear decision boundaries. Consequently, the ensemble yields a final prediction that demonstrates enhanced accuracy, stability, and generalizability compared to the output of any singular model operating in isolation.

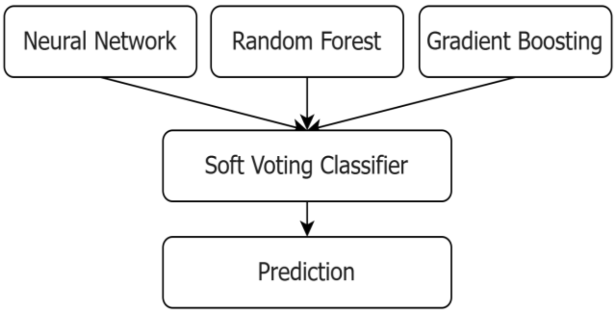

Figure 2. Soft Voting Classifier

2.5. Performance Evaluation Metrics

Model performance was measured using accuracy, precision, recall, F1-score, and AUC-ROC. These metrics provide a holistic understanding of model reliability in clinical contexts.

1. Accuracy
   The proportion of total correct predictions [22], [23].
2. Precision
   The proportion of true positive predictions among all positive predictions (Ability to not label a negative sample as positive).
3. Recall (Sensitivity)
   The proportion of true positives identified correctly among all actual positives (Ability to find all positive samples).
4. F1-Score
   The harmonic mean of precision and recall, providing a single metric that balances both concerns.
5. Area Under the Receiver Operating Characteristic Curve (AUC-ROC)
   A measure of the model's ability to distinguish between classes across all classification thresholds. The ROC curve plots the True Positive Rate (Recall) against the False Positive Rate.

Riska Aryanti: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 2, September 2025, Page 266-275
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i3.861

The results for each model across these metrics were tabulated and compared to demonstrate the effectiveness of the proposed soft voting ensemble for migraine type classification.

3. RESULT AND DISCUSSIONS

The dataset comprised 400 records, each representing a patient case. Table 2 presents the distribution of migraine subtypes. Performance evaluation across the nine models (eight base classifiers plus the ensemble) is reported in Table 3.

Table 2. Dataset Migraine

Age	Duration	Frequency	...	Intensity	...	DPF	Type
53	1	1	...	2	...	1	Typical aura with migraine
49	1	1	...	3	...	0	Migraine without aura
22	2	1	...	1	...	0	Sporadic hemiplegic migraine
19	1	1	...	3	...	1	Familial hemiplegic migraine
35	1	2	...	0	...	0	Typical aura without migraine
49	2	2	...	3	...	1	Other
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
37	1	1	...	3	...	1	Basilar-type aura

3.1. Performance Evaluation of Base Classifiers

The performance of all nine models, including the eight base classifiers and the soft voting ensemble, is summarized in Table 3. The results indicate a significant variance in performance across different algorithms.

The tree-based ensemble models (Random Forest, Gradient Boosting, XGBoost, Extra Trees) demonstrated strong overall performance, consistently achieving high accuracy and ROC-AUC scores. Notably, Gradient Boosting achieved the highest ROC-AUC score of 0.9820, indicating an excellent ability to distinguish between the seven migraine types. The Neural Network also performed exceptionally well, tying for the highest accuracy and achieving the highest precision score of 0.9282.

In contrast, simpler models like K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) exhibited substantially lower performance across all metrics. The SVM, in particular, showed a critical weakness with very low precision (0.3803), meaning a large proportion of its positive predictions were incorrect, despite a deceptively moderate recall. This performance disparity underscores the complexity of the feature space for this multi-class migraine classification problem and highlights the superiority of sophisticated, non-linear ensemble methods.

Table 3. Performance Comparison of All Classifiers

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Neural Network	0.9083	0.9282	0.9083	0.9130	0.9544
Extra Trees	0.8750	0.8934	0.8750	0.8807	0.9636
Random Forest	0.9083	0.9146	0.9083	0.9020	0.9742
KNN	0.6833	0.6215	0.6833	0.6066	0.8224
SVM	0.6167	0.3803	0.6167	0.4704	0.9485
AdaBoost	0.8250	0.7865	0.8250	0.7884	0.9146
Gradient Boost	0.9083	0.9160	0.9083	0.9050	0.9820
XGBoost	0.8833	0.8865	0.8833	0.8714	0.9715
Soft Voting Classifier	**0.9167**	**0.9267**	**0.9167**	**0.9150**	**0.9783**

Riska Aryanti: *Corresponding Author

Volume 6, Number 2, September 2025, Page 266-275
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i3.861

3.2. Performance of the Proposed Soft Voting Ensemble

The proposed Soft Voting Classifier, which aggregated the predictions from the top three performing models (Neural Network, Random Forest, and Gradient Boosting), achieved the best overall performance. It attained the highest accuracy (0.9167), recall (0.9167), and F1-Score (0.9150), while securing a near-top ROC-AUC score of 0.9783.

The confusion matrix for the Soft Voting Classifier (Fig 3) provides detailed insight into its predictive behavior. The model demonstrates perfect classification for the 'Migraine without aura' class (18 out of 18 instances). Most misclassifications occur within the rarer migraine types (e.g., Basilar-type, Familial hemiplegic, Sporadic hemiplegic) and the 'Other' category. For instance, one instance of 'Familial hemiplegic migraine' was misclassified as 'Typical aura with migraine', and several instances of 'Other' were confused with 'Typical aura with migraine'. This is a common challenge in medical datasets, where imbalanced class distributions and symptomatic similarities between rare conditions can lead to such confusions.

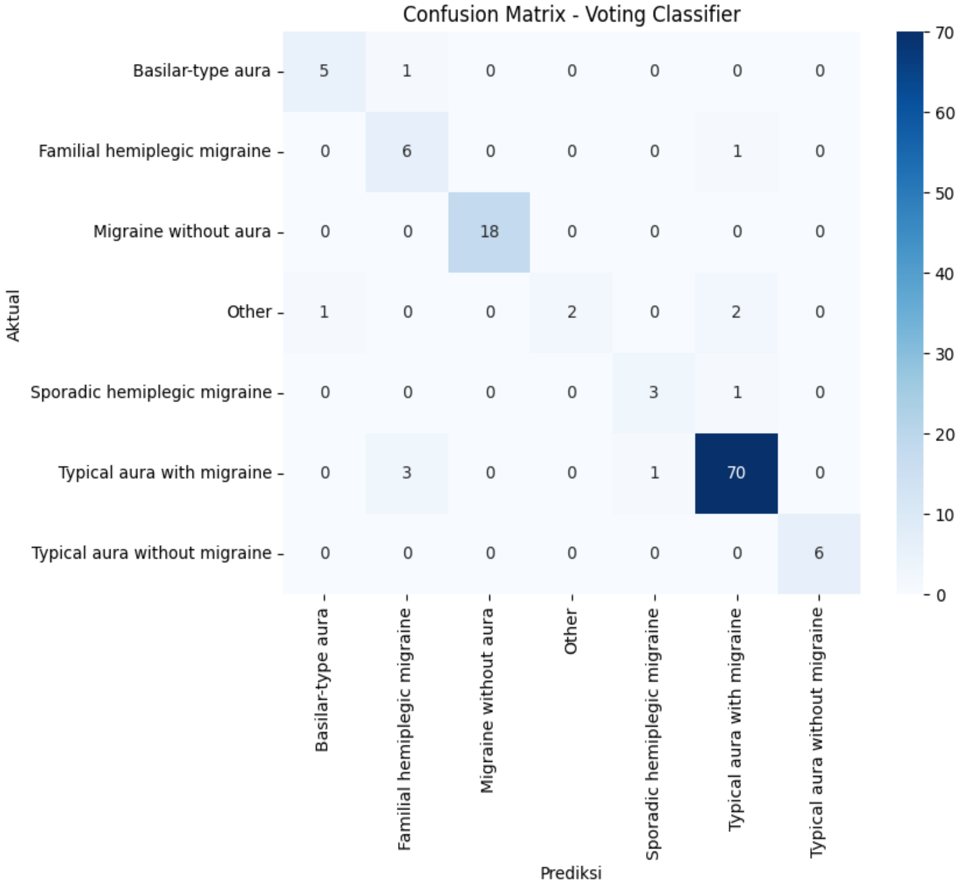

Figure 3. Confusion Matrix for the Soft Voting Classifier

The ROC curves for the ensemble, plotted for each class using a One-vs-Rest approach, are presented in Fig 4. The AUC values for all classes are exceptionally high, with three classes (0, 2, and 6) achieving a perfect AUC of 1.00, and the others scoring above 0.94. This confirms that the soft voting ensemble is highly capable of discriminating each individual migraine type from all others, with robust performance across the entire spectrum of classes.

Riska Aryanti: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 2, September 2025, Page 266-275
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i3.861

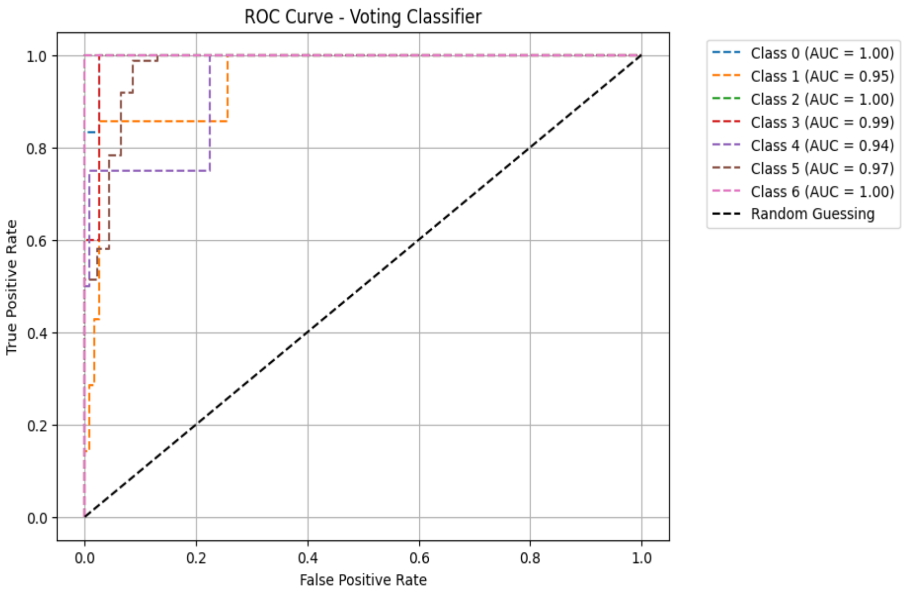

Figure 4. ROC Curves for the Soft Voting Classifier

3.3. Discussion

The experimental results strongly support the efficacy of the proposed soft voting ensemble strategy for migraine type classification. The superiority of ensemble learning is evident from the fact that the soft voting approach consistently outperformed its individual constituent models. This outcome illustrates the "wisdom of the crowd" principle in machine learning, where the combination of Neural Network, Random Forest, and Gradient Boosting successfully mitigates the weaknesses and biases of each model. By leveraging prediction probabilities rather than hard labels, the ensemble provides more nuanced decision-making, leading to a robust and generalized classifier.

Equally important, the high F1-Score and the balanced precision-recall values achieved by the ensemble demonstrate its strong potential for clinical application. High recall ensures that fewer patients are misclassified, minimizing the risk of missed diagnoses, while high precision reduces false alarms, ensuring that diagnoses are reliable. Moreover, the confusion matrix highlights patterns of misclassification, offering clinicians valuable insights into overlapping diagnostic criteria and potential areas for refinement. This interpretability further strengthens the practical relevance of the proposed method.

Another critical aspect of the findings is the model's ability to address class imbalance. The consistently high AUC-ROC values across all classes, including rare migraine subtypes, indicate strong discriminatory power beyond overall accuracy. This suggests that the ensemble does not merely optimize for the majority classes but also maintains reliable performance for minority categories, which are often overlooked in medical datasets. Such balanced effectiveness enhances the trustworthiness of the model as a clinical decision-support tool, particularly in diagnosing less common migraine subtypes.

4. CONCLUSION

This research confirms that a soft voting ensemble provides superior accuracy and reliability for migraine subtype classification compared to individual models. By integrating Neural Network, Random Forest, and Gradient Boosting, the ensemble achieved 91.67% accuracy and 91.50% F1-score, supported by near-perfect AUC-ROC values. The findings emphasize the strength of ensemble learning in mitigating individual model weaknesses and delivering consistent predictions. Future research will focus on model interpretability using techniques such as SHAP or LIME, hyperparameter optimization, and large-scale validation to strengthen clinical adoption.

Riska Aryanti: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 2, September 2025, Page 266-275
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i3.861

5. ACKNOWLEDGMENT (*optional*)

The authors gratefully acknowledge the financial support for this research provided by Bina Sarana Informatika University (UBSI) in the fiscal year 2025.

6. REFERENCES

[1] M. F. P. Peres *et al.*, "Migraine is the most disabling neurological disease among children and adolescents, and second after stroke among adults: A call to action," *Cephalalgia*, vol. 44, no. 8, Aug. 2024, doi: 10.1177/03331024241267309.
[2] T. J. Steiner, L. J. Stovner, R. Jensen, D. Uluduz, and Z. Katsarava, "Migraine remains second among the world's causes of disability, and first among young women: findings from GBD2019," *J. Headache Pain*, vol. 21, no. 1, p. 137, Dec. 2020, doi: 10.1186/s10194-020-01208-0.
[3] N. Magnavita, "Headache in the Workplace: Analysis of Factors Influencing Headaches in Terms of Productivity and Health," *Int. J. Environ. Res. Public Health*, vol. 19, no. 6, p. 3712, Mar. 2022, doi: 10.3390/ijerph19063712.
[4] M. Leonardi and A. Raggi, "A narrative review on the burden of migraine: when the burden is the impact on people's life," *J. Headache Pain*, vol. 20, no. 1, p. 41, Dec. 2019, doi: 10.1186/s10194-019-0993-0.
[5] T. J. Steiner *et al.*, "Migraine-attributed burden, impact and disability, and migraine-impacted quality of life: Expert consensus on definitions from a Delphi process," *Cephalalgia*, vol. 42, no. 13, pp. 1387–1396, Nov. 2022, doi: 10.1177/03331024221110102.
[6] A. K. Eigenbrodt *et al.*, "Diagnosis and management of migraine in ten steps," *Nat. Rev. Neurol.*, vol. 17, no. 8, pp. 501–514, Jun. 2021, doi: 10.1038/s41582-021-00509-5.
[7] A. A. Garstka *et al.*, "Accurate Diagnosis and Treatment of Painful Temporomandibular Disorders: A Literature Review Supplemented by Own Clinical Experience," *Pain Res. Manag.*, vol. 2023, pp. 1–12, Jan. 2023, doi: 10.1155/2023/1002235.
[8] J. A. Edlow, "Misdiagnosis of Acute Headache," *Emerg. Med. Clin. North Am.*, vol. 43, no. 1, pp. 67–80, Feb. 2025, doi: 10.1016/j.emc.2024.05.026.
[9] C. Baccouch and C. Bahar, "Advanced Machine Learning Approaches for Accurate Migraine Prediction and Classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 16, no. 1, 2025, doi: 10.14569/IJACSA.2025.0160101.
[10] L. Khan, M. Shahreen, A. Qazi, S. Jamil Ahmed Shah, S. Hussain, and H.-T. Chang, "Migraine headache (MH) classification using machine learning methods with data augmentation," *Sci. Rep.*, vol. 14, no. 1, p. 5180, Mar. 2024, doi: 10.1038/s41598-024-55874-0.
[11] M. Yuichi and Y. A. Susetyo, "Klasifikasi Penyakit Migrain dengan Metode Naïve Bayes pada Dataset Kaggle," *J. Indones. Manaj. Inform. dan Komun.*, vol. 6, no. 1, pp. 139–151, Jan. 2025, doi: 10.35870/jimik.v6i1.1150.
[12] A. Reddy and A. Reddy, "Migraine triggers, phases, and classification using machine learning models," *Front. Neurol.*, vol. 16, May 2025, doi: 10.3389/fneur.2025.1555215.
[13] C. Baccouch and C. Bahar, "Advanced Machine Learning Approaches for Accurate Migraine Prediction and Classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 16, no. 1, 2025, doi: 10.14569/ijacsa.2025.0160101.
[14] W. Lee, I. K. Min, K. I. Yang, D. Kim, C.-H. Yun, and M. K. Chu, "Classifying migraine subtypes and their characteristics by latent class analysis using data of a nation-wide population-based study," *Sci. Rep.*, vol. 11, no. 1, p. 21595, Nov. 2021, doi: 10.1038/s41598-021-01107-7.
[15] Y. Garcia-Chimeno, B. Garcia-Zapirain, M. Gomez-Beldarrain, B. Fernandez-Ruanova, and J. C. Garcia-Monco, "Automatic migraine classification via feature selection committee and machine learning techniques over imaging and questionnaire data," *BMC Med. Inform. Decis. Mak.*, vol. 17, no. 1, Apr. 2017, doi: 10.1186/s12911-017-0434-4.
[16] A. Rojarath and W. Songpan, "Cost-sensitive probability for weighted voting in an ensemble model for multi-class classification problems," *Appl. Intell.*, vol. 51, no. 7, pp. 4908–4932, Jul. 2021, doi: 10.1007/s10489-020-02106-3.
[17] S. Kumari, D. Kumar, and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier," *Int. J. Cogn. Comput. Eng.*, vol. 2, 2021, doi: 10.1016/j.ijcce.2021.01.001.
[18] K. A. Pratama, M. Nasution, and M. H. Munandar, "Sistem Informasi Pendaftaran Imunisasi Posyandu Anggrek Di Desa Tanjung Medan Berbasis Web," pp. 27–33, 2020.
[19] W. Zhang, D. Yang, and S. Zhang, "A new hybrid ensemble model with voting-based outlier detection and balanced sampling for credit scoring," *Expert Syst. Appl.*, vol. 174, p. 114744, Jul. 2021, doi: 10.1016/j.eswa.2021.114744.
[20] J. R. Campos, E. Costa, and M. Vieira, "Improving Failure Prediction by Ensembling the Decisions of Machine Learning Models: A Case Study," *IEEE Access*, vol. 7, pp. 177661–177674, 2019, doi:

Riska Aryanti: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 6, Number 2, September 2025, Page 266-275
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v6i3.861

10.1109/ACCESS.2019.2958480.
[21] I. D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," *IEEE Access*, vol. 10, pp. 99129–99149, 2022, doi: 10.1109/ACCESS.2022.3207287.
[22] P. Boozary, S. Sheykhan, H. GhorbanTanhaei, and C. Magazzino, "Enhancing customer retention with machine learning: A comparative analysis of ensemble models for accurate churn prediction," *Int. J. Inf. Manag. Data Insights*, vol. 5, no. 1, p. 100331, Jun. 2025, doi: 10.1016/j.jjimei.2025.100331.
[23] V. N. Thatha *et al.*, "Optimized machine learning mechanism for big data healthcare system to predict disease risk factor," *Sci. Rep.*, vol. 15, no. 1, p. 14327, Apr. 2025, doi: 10.1038/s41598-025-98721-6.

Riska Aryanti: *Corresponding Author