Optimizing the Accuracy of Alzheimer's Detection Using Machine Learning and Intelligent Feature Selection Strategies

Suci Mutiara[1], Siti Nur Laila[2], Deppi Linda[3], Sri Lestari[4*], Jean Antoni[5], Christian Petrus Silalahi[6]
Department of Informatics Engineering, Computer Science, Institut Informatika Dan Bisnis Darmajaya, Lampung, Indonesia[1, 4, 5, 6]
Department of Information Technology Education, Computer Science, Institut Informatika Dan Bisnis Darmajaya, Lampung, Indonesia[2]
Department of Information Systems, Computer Science, Institut Informatika Dan Bisnis Darmajaya, Lampung, Indonesia[3]

***Abstract*—Alzheimer's disease is a progressive neurodegenerative disorder for which early detection remains a significant challenge due to the complexity of clinical features and the high dimensionality of medical data. This study aims to improve the accuracy and reliability of Alzheimer's disease detection by evaluating the performance of multiple machine learning algorithms integrated with intelligent feature selection strategies. Five classification models, Decision Tree, Naïve Bayes, Random Forest, Logistic Regression, and Deep Learning, were investigated under two experimental scenarios: without feature selection and with feature selection using Recursive Feature Elimination, Binary Particle Swarm Optimization, and Variance Threshold. Model performance was evaluated using K-fold cross-validation based on accuracy, precision, recall, and F1-score metrics. The results demonstrate that feature selection consistently enhances classification performance, particularly for conventional machine learning models such as Random Forest and Logistic Regression. Although the Deep Learning model achieves competitive accuracy, its reduced precision and F1-score indicate limitations when applied to reduced feature spaces. These findings highlight the importance of incorporating appropriate feature selection techniques to address data complexity and improve the effectiveness of early Alzheimer's disease detection.**

***Keywords*—Machine learning; Alzheimer; Random Forest (RF); logistic regression; deep learning**

I. Introduction

Alzheimer's disease (AD) constitutes a chronic neurological dysfunction that develops gradually and has a significant impact on the cognitive abilities and daily activities of those affected. Individuals with Alzheimer's slowly experience a decline in thinking, memory, and even in performing the most basic tasks [1]. The main problem in detecting Alzheimer's disease is that its symptoms often develop gradually and are challenging to recognize in the early stages. This usually results in many cases being diagnosed when the disease is already quite severe, thereby reducing the effectiveness of medical intervention. Based on this, early detection of the disease is critical as a mitigation step to slow its progression and improve patient outcomes through appropriate treatment and therapy [2]. Advancements in technology have made it possible to examine peripheral blood and accurately measure Alzheimer's biomarkers during the initial phases of the disease, using methods that are cost-effective, non-invasive, and safe. This technique contributes to the clinical and biological characterization of cognitive impairment at an early stage by measuring Alzheimer's biomarkers in blood and brain structures in groups of participants classified based on their clinical cognitive phenotypes. Accurate diagnosis at the early stages of Alzheimer's will minimize risk factors and facilitate preventive monitoring [3].

Recent advances in artificial intelligence (AI) and machine learning (ML) have provided promising solutions in medical diagnosis, including the detection of Alzheimer's disease [4]. Machine Learning techniques can analyze large datasets and identify complex patterns that can detect Alzheimer's disease. ML algorithms utilize medical imaging, clinical records, genetic data, and cognitive assessments to improve diagnostic accuracy and enable early intervention [5]. Several researchers have employed this method, one of which is described by Jimenez-Maggiora et al. Their work applies natural language processing (NLP) and artificial intelligence (AI) to streamline and unify the coding of adverse event data recorded by physicians in Alzheimer's disease (AD) clinical trials. The study involves creating a gold-standard dataset of adverse events in AD, testing the performance of NLP-driven models in classifying such events, and examining whether automated coding can surpass physician coding in terms of efficiency, accuracy, reliability, and consistency [6]. The International Neurodegenerative Disorder Research Center (INDRC) applies artificial intelligence and machine learning (AI/ML) to create advanced, multidisciplinary analytical approaches that integrate neuroscience, biophysics and biochemistry, computer science and engineering, mathematics, along with clinical and population studies. Through these AI/ML-driven strategies, researchers can gain deeper insights into the mechanisms underlying Alzheimer's disease and support the discovery of effective therapeutic options [7].

Wallensten conducted a subsequent study that integrated machine learning with medical records to facilitate the early detection of Alzheimer's disease (AD), thereby enabling timely intervention. The method used was Stochastic Gradient Boosting to identify predictive diagnoses of AD using primary healthcare data. By utilizing clinical data, including previous diagnoses and medical treatments, the sensitivity and specificity of diagnostic procedures will be improved [8]. Nilesh K and

*Corresponding author.

Vinayak B used Support Vector Machine (SVM), K-Nearest Neighbor (KNN) for classification in diagnosing Alzheimer's Disease [9]. Yue L, et al. built a predictive model using Deep Learning for MCI conversion. Deep Learning is applied to structural MRI to capture minor brain changes that characterize the pre-MCI phase. Predicting the conversion from normal cognition (NC) to MCI is essential for early detection and intervention [10].

In line with previous studies, we propose machine learning methods, including Decision Tree (DT), Random Forest (RF), Naïve Bayes (NB), Logistic Regression (LR), and Deep Learning (DL) to classify Alzheimer's disease cases. These methods were chosen because of their ability to handle complex medical datasets and provide interpretable results. In addition, DT and RF are highly effective in identifying non-linear relationships and essential features within the dataset [11]. This study proposes feature selection methods, including Binary Particle Swarm Optimization (BPSO), Recursive Feature Elimination (RFE), and Variance Threshold (VT), to develop a model with superior performance as a reference for early detection of Alzheimer's disease.

II. METHOD

This study was conducted through several systematic stages, namely: 1) data collection as the basis of the study, 2) Data pre-processing is performed to ensure that the available data remains consistent and meets quality standards, 3) feature selection to eliminate irrelevant attributes and retain significant features, 4) building a classification model using the selected machine learning algorithm, and (5) evaluating the model's performance. The detailed stages of the study were illustrated in Fig. 1.

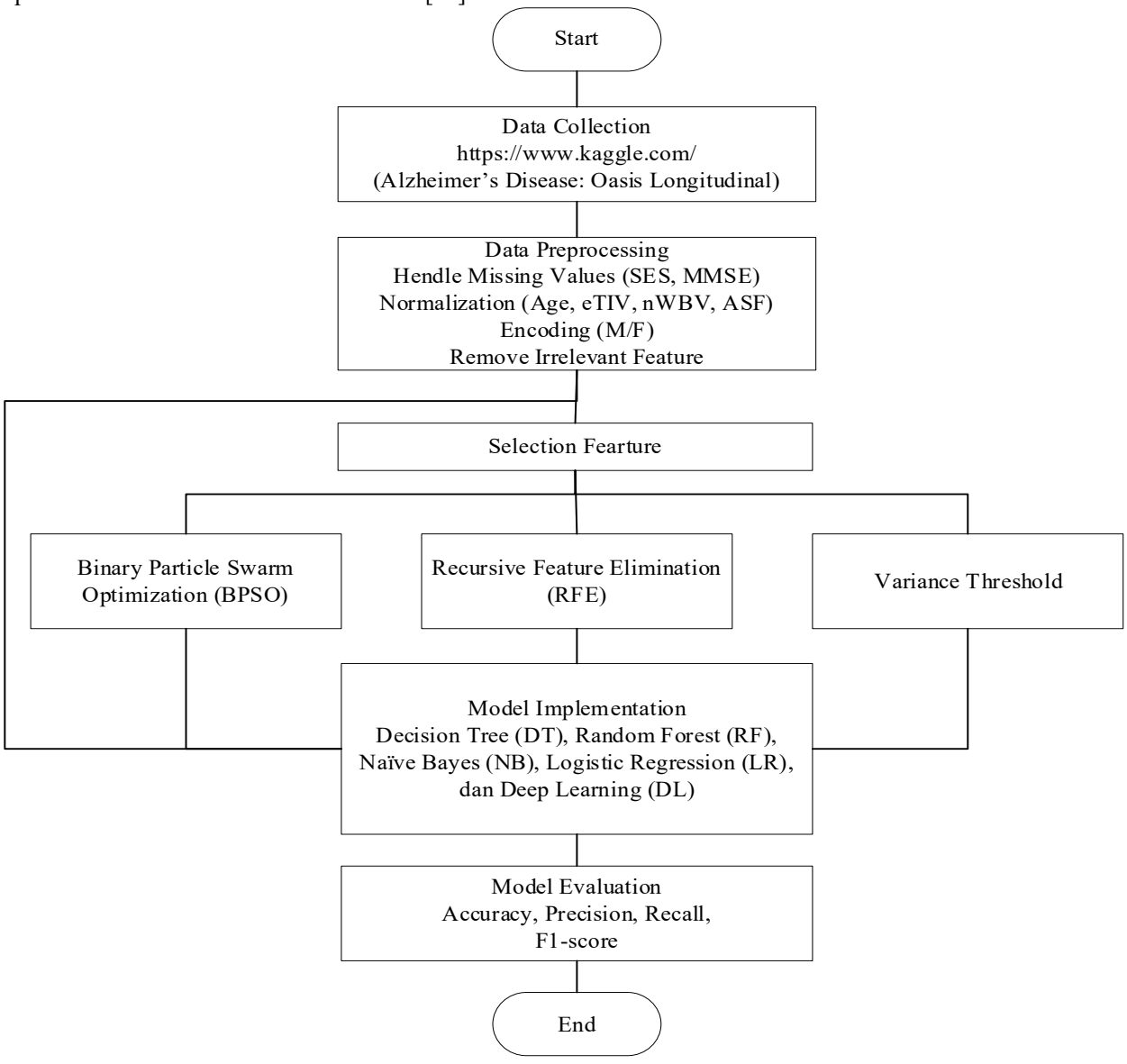

Fig. 1. Research stages classification method.

The dataset used in this study was sourced from the Kaggle platform (https://www.kaggle.com/) and specifically focuses on Alzheimer's Disease, utilizing the Oasis Longitudinal dataset. The dataset consists of 14 features: MRI ID, Group, Visit, MR

Delay, M/F, Hand, Age, EDUC, SES, MMSE, CDR, eTIV, nWBV, and ASF. The dataset was then preprocessed by handling missing values, normalization, encoding, and removing less relevant features. For feature selection, we used three approaches: Binary Particle Swarm Optimization (BPSO), Recursive Feature Elimination (RFE), and Variance Threshold. The next step was to build classification models using five algorithms: Decision Tree (DT), Random Forest (RF), Naïve Bayes (NB), Logistic Regression (LR), and Deep Learning (DL). The final stage evaluated the classification models using accuracy, precision, recall, and F1-score to assess prediction accuracy, positive case detection, and overall performance balance.

A. Decision Tree

The Decision Tree is a structured classification method characterized by internal nodes that evaluate features, branches that map to specific decision criteria, and leaf nodes that provide the final output. This algorithmic approach is widely adopted across diverse research fields, particularly for prediction and prognostic modeling [12]. For example, Naswin used Decision Tree algorithm to distinguish between normal and pneumonia-diagnosed X-ray images. Using pediatric X-ray data, this study applied a systematic data pre-processing strategy, including Canny segmentation and humoments feature extraction, to enhance performance [13].

The core of the Decision Tree algorithm was its ability to recursively split the data based on the features that provide the highest Information Gain (IG) or lowest Gini Impurity. The Decision Tree (DT) was a multilevel model that combines a series of basic tests in an efficient and integrated manner, where each test compares a numerical feature with a specific threshold value. The Information Gain metric determines the decrease in entropy when a dataset is divided. The computations for IG, Entropy, and Gain are carried out using Eq. (1)–(3) [14].

$$IG\ (D, A) = Entrophy\ (D) - \sum_{\mathrm{V} \in \mathrm{Values(A)}} \frac{|D|}{D}\ enthrophy\ (D_V) \quad (1)$$

Where Entropy (D) was calculated as:

$$Entrophy\ (D) = \sum_{I=1}^{K} P_I\ \log_2(P_I) \quad (2)$$

Alternatively, the Gini Impurity was calculated as:

$$Gain = Entrophy\ (S) - \sum_{I=1}^{K} P_2^I \quad (3)$$

B. Random Forest

Random forest was a machine learning model used for classification and prediction. It was built from multiple decision trees, each trained on a random subset of data generated through bootstrap sampling [15]. The overall model performance can be affected if the individual trees have high correlation or low accuracy [16], [17].

To partition the data at each node, the algorithm utilizes indicators such as Information Gain, which are computed using specific mathematical formulations. A lower Gini Split value indicated that a feature is more effective in separating data into more homogeneous classes. By selecting the most informative features, the algorithm can improve its classification performance optimally [18]. The formula for calculating entropy using Eq. (4).

$$Entrophy\ (D) = \sum_{I=1}^{K} P_I\ \log_2(P_I) \quad (4)$$

where, pi represented the probability of each class in the dataset, the Information Gain was calculated by comparing the entropy before and after splitting using Eq. (5).

$$IG\ (D, A) = Entrophy\ (D) - \sum_{\mathrm{V} \in \mathrm{Values(A)}} \frac{|D|}{D}\ enthrophy\ (D_V) \quad (5)$$

Gini Split was a measurement in decision trees to determine how well a feature (variable) divided data into classes. Gini Split measures the impurity of data division. The lower the Gini value, the better the feature was at separating data by class. Gini Split could be calculated using Eq. (6).

$$Gini\ Split = \sum_{i=1}^{c} \left(\frac{n_i}{n}\right) x\ Gini\ Indeks(S_i) \quad (6)$$

C. Naïve Bayes (NB)

The Naive Bayes algorithm was a probabilistic classification model rooted in Bayes' Theorem. Despite its "naive" assumption of conditional independence among features, an assumption that was not always valid in practice, the model often demonstrated robust performance, particularly in text classification tasks like spam filtering and sentiment analysis [19]. As a foundational approach in pattern recognition, the algorithm evaluates classification decisions by weighing the probabilities of each choice and its consequential implications [20][21].

In the context of Alzheimer's diagnosis, Naive Bayes computes the probability of a patient's class membership, Alzheimer's or non-Alzheimer's, based on a set of clinical and biometric features such as age, Mini-Mental State Examination (MMSE) score, Socioeconomic Status (SES), Atlas Scaling Factor (ASF), and Estimated Total Intracranial Volume [22]. These probability calculations are mathematically formulated based on the principles of Bayes' Theorem [23].

$$p\ (c|x) = \frac{\mathrm{p\ (C|\ X).p(c)}}{p(x)} \quad (7)$$

Since Naïve Bayes assumed that each feature was independent, the likelihood could be calculated using Eq. (8).

$$(C|X) = (X1|C) \times (X2|C)\ \ldots\ .\times P(XN|C) \quad (8)$$

For continuous features, the probability $(X1|C)$ was often modeled using a Gaussian (Normal) distribution, the Gaussian (Normal) distribution could be calculated using Eq. (9).

$$p(c|x) = \frac{1}{\sqrt{2\pi\sigma^2}}\ e^{-\frac{-(\mathrm{xi}-\mu)2}{2\sigma 2}} \quad (9)$$

D. Deep Learning

Deep learning was a computational paradigm that utilizes artificial neural networks to mimic the biological architecture of the human brain. This method was fundamentally capable of analyzing unstructured data and automatically identifying relevant features [24].

Unlike traditional classification algorithms such as Logistic Regression or Naive Bayes, Deep Learning excels at capturing complex nonlinear patterns thanks to its layered architecture. In the context of medical diagnosis, specifically Alzheimer's, this model processed patient data such as age, Mini-Mental State Examination (MMSE) score, Socioeconomic Status (SES), Atlas Scaling Factor (ASF), and Estimated Total Intracranial Volume (eTIV) through a series of layers, including the input,

hidden, and output layers, to produce accurate classification decisions [25], [26].

The Deep Learning model used in this study was implemented as a feed-forward multilayer perceptron consisting of one input layer, two hidden layers, and one output layer. Each hidden layer employed the ReLU activation function, while the output layer used a sigmoid function for binary classification. The training process was conducted using the Adam optimizer with a learning rate of 0.001 and the binary cross-entropy loss function. To reduce the risk of overfitting on the relatively limited dataset, dropout regularization was applied during training. Model performance was evaluated using a K-fold cross-validation scheme to ensure stable and reproducible results.

The prediction process in artificial neural networks was based on a linear combination of given features with certain weights in each neuron; the neuron calculation could use Eq. (10).

$$= WX + b \tag{10}$$

After the linear transformation, the result was passed to the activation function to add non-linearity. ReLU (Rectified Linear Unit) for hidden layers could be calculated using Eq. (11)[27][28].

$$(z) = (0, z) \tag{11}$$

During training, the model updated the weights W using an optimization algorithm, such as Stochastic Gradient Descent (SGD) or Adam Optimizer by decreasing the value of the loss function, which in binary classification was often Binary Cross-Entropy. Binary Cross-Entropy can be calculated using Eq. (12).

$$Loss = \sum_{N}^{1}[(Y_{I \log (\widehat{Yi})}] + (1 - Yi) \log (1 - \widehat{yi}) \tag{12}$$

E. Logistic Regression

Logistic Regression (LR) is a prevalent analytical tool employed when the outcome variable is restricted to two categories, representing binary outcomes such as decision success or failure, or clinical status (diseased or healthy). This study employed logistic regression as a classical statistical method to model binary outcomes, choosing it for medical applications due to its high interpretability. Although machine learning methods often yield better performance on high-dimensional data, their complexity makes them more challenging to understand and explain. In addition, on low-dimensional data, the performance of machine learning methods tends to be comparable to that of logistic regression [29].

In the context of Alzheimer's diagnosis, Logistic Regression calculated the probability of a patient being in the Alzheimer's (1) or Non-Alzheimer's (0) class based on features such as age, MMSE (Mini-Mental State Examination) score, SES (Socioeconomic Status), ASF (Atlas Scaling Factor), and eTIV (Estimated Total Intracranial Volume)[30].

This model used the following formula to determine the probability that a patient would fall into a particular class. The likelihood of a class was calculated using a sigmoid function, as shown in Eq. (13) [31]:

$$P(Y = 1|X) = \frac{1}{1+e^{-z}} \tag{13}$$

where z was defined as a linear combination of the input features using Eq. (14).

$$z = w0 + w1x1 + w2x2 + \cdots + wnxn \tag{14}$$

and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

F. Feature Selection

1) Recursive Feature Elimination (RFE): Recursive Feature Elimination (RFE) was employed to systematically select the most significant features, serving as the primary feature selection technique in this machine learning study. The goal was to enhance the model's accuracy by simplifying the data. RFE removed features or attributes that had the lowest ranking scores[32]. RFE selects features based on their importance ranking through a recursive elimination process[33]. In addition, feature selection improved the performance of the model, such as in the CatBoost regression model[34]. Improving prediction accuracy for heart disease by combining gradient boosting with Recursive Feature Elimination (RFE-GB)[35].

2) Binary Particle Swarm Optimization (BPSO): In general, the PSO algorithm was designed to solve continuous optimization problems. To address challenges in discrete optimization, Kennedy and Eberhart developed a special variant called Binary PSO (BPSO), where they encoded the particle positions in binary form[36]. BPSO is a special version of the PSO algorithm used to choose between two options, namely 0 or 1. In BPSO, the particle velocity did not indicate how far they move, but rather how likely the value 0 was to change to 1, or vice versa (bit flipping) [37].

3) Variance Threshold (VT): To ensure only significant attributes are retained for classification, the Variance Threshold (VT) method is employed to eliminate features whose low statistical variance indicates minimal contribution to the overall process. By eliminating irrelevant variables, the classifier operated more efficiently and rapidly without compromising overall prediction accuracy [38]. This method was effective for removing features with low variation or those that only contain noise, especially in data measured on a uniform scale [39].

4) Evaluation: This study conducted a confusion matrix as the primary tool in the model evaluation process. The confusion matrix allowed for a comprehensive analysis of the model's performance by comparing predictions to actual classes. Through this assessment framework, various metrics such as accuracy, precision, recall, and F1-score can be calculated to evaluate the model's effectiveness and accuracy [40], [41], [42]. Therefore, the utilization of the confusion matrix in this study provided a solid foundation for evaluating the performance of the designed classification algorithm. Model effectiveness was assessed through metrics formulated in Eq. (15)-(18).

$$Accuracy = \frac{1}{n} \sum_{m=1}^{n} \frac{TP_M + TN_M}{TP_M + TN_M + FP_M + FN_M} \tag{15}$$

$$Precision = \frac{1}{n}\sum_{m=1}^{n}\frac{TP_M}{TP_M+FP_M} \quad (16)$$

$$Recall = \frac{1}{n}\sum_{m=1}^{n}\frac{TP_M}{TP_M+FN_M} \quad (17)$$

$$F1-Score = \frac{1}{n}\sum_{m=1}^{n}\frac{2^*TP_M}{2^*TP_M+TP_M+FN_M} \quad (18)$$

where,

n = number of classes, m ∈ {1, 2, …n} = class index

TP (True Positives) = elements of class m that the model correctly classifies.

FN_m (False Negatives) = elements of class m that were not recognized by the model (predicted as another class).

TN_m (True Negatives) = elements that were not from class m, and correctly not predicted as class m.

FP_m (False Positives) = elements that were not from class m, but are incorrectly predicted as part of class m.

III. Results and Discussion

A. Results

This study had two scenarios. The first scenario employed the dataset directly without feature selection, while the second scenario used a dataset that had undergone feature selection. They employed feature selection methods, including Binary Particle Swarm Optimization (BPSO), Recursive Feature Elimination (RFE), and Variance Threshold, to identify the most relevant attributes for Alzheimer's classification. The goal of feature selection was to optimize the model's performance in detecting Alzheimer's. This was achieved by prioritizing pertinent features to lower complexity and improve prediction efficacy.

It processed the dataset using classification methods, including Decision Tree (DT), Random Forest (RF), Naïve Bayes (NB), Logistic Regression (LR), and Deep Learning (DL). They evaluated the model performance with K-Fold Cross-Validation. They chose K-Fold Cross-Validation because it provided more accurate and comprehensive results. This technique divided the data into several parts and then trained and tested the model alternately on each part. In this way, all data contributed to the evaluation process, reducing bias caused by uneven data splitting.

The evaluation results, which represented the average of all folds, made the model's performance more stable and representative. Therefore, this study selected K-Fold Cross-Validation as the primary method to ensure the quality and reliability of the model. This study compared model effectiveness using key evaluation metrics, each offering a distinct perspective on predictive capability. These metrics complemented each other in assessing accuracy, coverage, and balance of the model's predictions, particularly in the context of imbalanced data. By using these four metrics, the evaluation became more comprehensive and fairer in comparing the performance of different models. The evaluation results appeared in Fig. 2–5.

B. Performance Comparison

Fig. 2 presents a comparative evaluation of five ML algorithms, namely Random Forest, Naive Bayes, Decision Tree, , Deep Learning and Logistic Regression, that the study applied without using feature selection techniques. The study assessed each model using four standard evaluation metrics to achieve a thorough understanding of its classification effectiveness. The evaluation results showed that the Random Forest algorithm provided the best performance, with an accuracy and precision of 0.96 and a perfect recall of 1.00. This finding demonstrated the model's ability to identify all positive classes consistently. Logistic Regression also demonstrated stable and balanced performance across all metrics (0.95), which indicated reliability and consistency in the classification task.

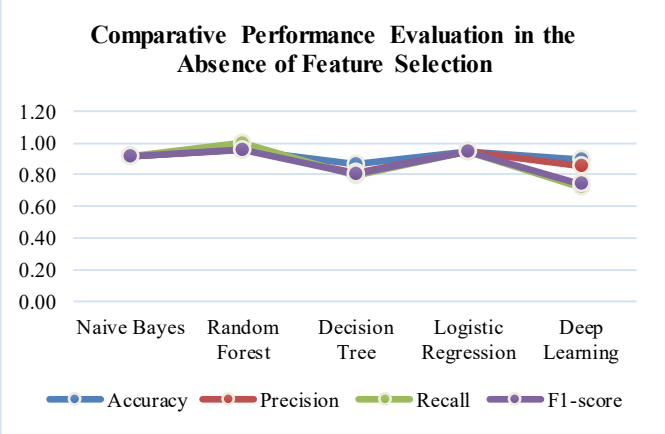

Fig. 2. Comparative performance evaluation in the absence of feature selection.

Conversely, Naive Bayes produced a uniform score of 0.91 across all metrics, which reflected moderate performance with low variability. The Decision Tree recorded the lowest scores, particularly in precision (0.81) and recall (0.79), which indicated limitations in handling complex data distributions. Meanwhile, Deep Learning achieved an accuracy of 0.90 but experienced a significant decline in recall (0.72) and F1-score (0.74), which demonstrated challenges in comprehensively detecting positive classes.

Fig. 3 outlines a comparative study of five classification approaches, namely: Random Forest, Naive Bayes, Decision Tree, Deep Learning and Logistic Regression that the study optimized through feature selection using the Binary Particle Swarm Optimization (BPSO) method. Feature selection played a crucial role in machine learning modeling, as it improved model generalization, reduced data dimensionality, and enhanced computational efficiency. The evaluation results showed that Naive Bayes and Random Forest, combined with BPSO, achieved the highest and most consistent scores across all metrics (0.96), which indicated strong and stable classification capabilities. Logistic Regression also demonstrated balanced performance with a value of 0.95 on all metrics. Decision Tree experienced a significant increase, with precision and recall at 0.94 and accuracy at 0.93. Conversely, Deep Learning achieved an accuracy of 0.89 but exhibited a sharp decline in precision (0.59) and F1-score (0.62), which suggested challenges in classifying positive classes despite

feature selection. Overall, the integration of BPSO-based feature selection contributed positively to the performance improvement of most models, especially in terms of stability and classification effectiveness. These findings emphasized the importance of applying optimization techniques in the feature selection process to produce models that were more efficient, accurate, and adaptable to high-dimensional data.

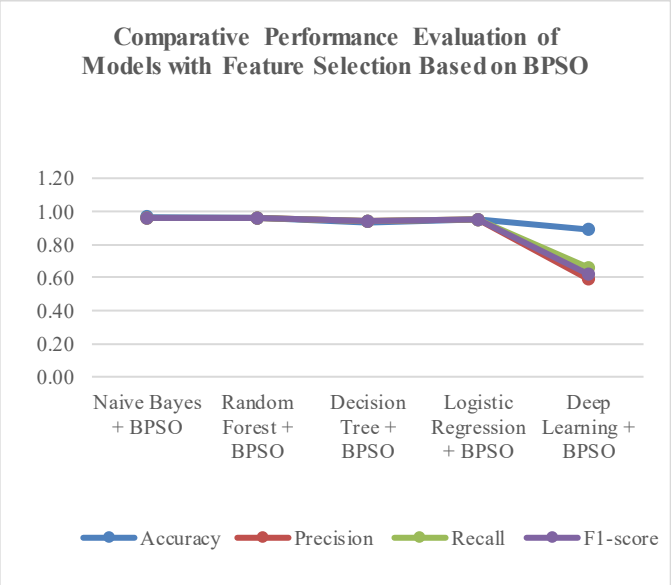

Fig. 3. Comparative performance evaluation of models with feature selection based on BPSO.

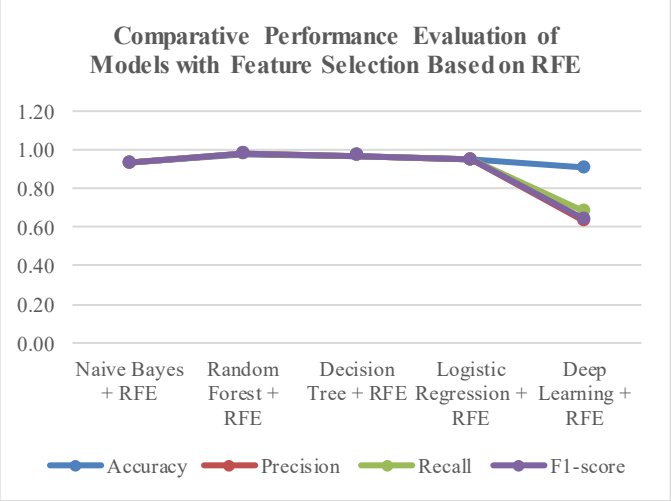

Fig. 4. Comparative performance evaluation of models with feature selection based on RFE.

The implementation of Recursive Feature Elimination (RFE) as a feature selection technique resulted in a consistent pattern of performance improvement across most of the tested classification algorithms. Based on the evaluation results shown in Fig. 4, RFE significantly contributed to increased accuracy and consistency in predictions. Random Forest and Decision Tree demonstrated the most prominent performance, recording accuracy scores of 0.98 and 0.97, respectively, with uniform precision, recall, and F1-score values. This indicated that both models responded highly to the feature simplification performed by RFE without losing their classification capacity. Logistic Regression also demonstrated stable performance, with a score of 0.95 across all metrics, which indicated that this model remained robust even when features were reduced. Naive Bayes consistently scored 0.93. It suggested that, although simple, this model still maintained competitive performance after feature selection was applied. Conversely, Deep Learning showed an imbalance among metrics, with an accuracy of 0.91 but much lower precision and F1-score (0.63 and 0.64). These findings suggested that neural network-based models required a more adaptive feature selection approach to match their internal architectural complexity. Overall, these results demonstrated that RFE enhanced the efficiency and accuracy of classical models. In contrast, its impact on complex models, such as Deep Learning, still required further exploration.

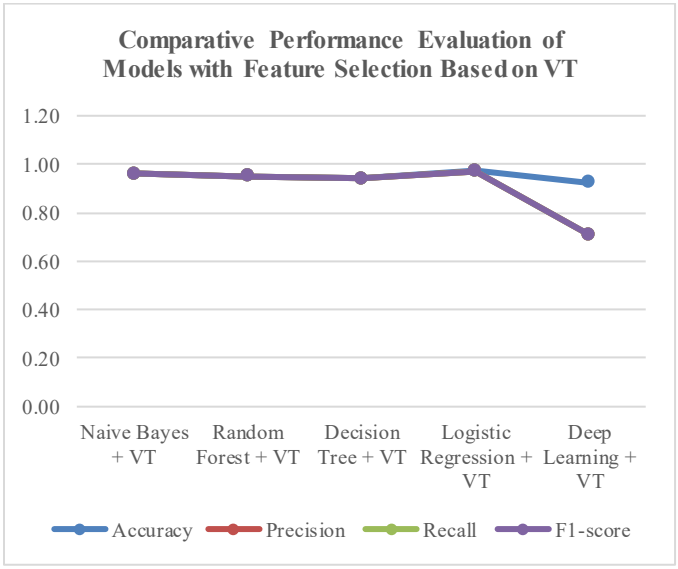

Fig. 5. Comparative performance evaluation of models with feature selection based on VT.

Fig. 5 presents the evaluation results of the impact of the Variance Threshold (VT)-based feature selection method on the performance of five classification algorithms: Random Forest, Naive Bayes, Decision Tree, Deep Learning and Logistic Regression. VT acted as a filter approach that eliminated features with low variance, thereby retaining only attributes that significantly contributed to the data distribution. The evaluation results showed that Logistic Regression combined with VT yielded the highest and most consistent performance across all metrics (0.97), followed by Naive Bayes (0.96), Random Forest (0.95), and Decision Tree (0.94). These four models demonstrated high classification stability, indicating that VT preserved statistically relevant features to improve prediction accuracy. Conversely, Deep Learning achieved fairly good accuracy (0.92) but experienced a significant decline in precision, recall, and F1-score (each at 0.71). This finding suggested that the VT approach was less optimal for complex architectures, such as neural networks, which tended to require richer and nonlinear feature representations.

Overall, this study demonstrated that the Variance Threshold served as a practical feature selection technique, enhancing the efficiency and performance of conventional classification

models. However, its application to deep learning models required a more adaptive approach to structural complexity and data representation needs.

C. Discussion

This study evaluated the efficacy of five machine learning algorithms in detecting Alzheimer's Disease (AD) by integrating intelligent feature selection strategies. The analytical results demonstrate that the application of feature selection techniques consistently enhances the predictive capability of the models compared to using the dataset without feature selection. This aligns with the principle that dimensionality reduction, through methods such as Recursive Feature Elimination (RFE) and Binary Particle Swarm Optimization (BPSO), is capable of eliminating irrelevant attributes, thereby strengthening the model's generalization in identifying AD pathological patterns more accurately.

The superiority of classical algorithms, particularly Random Forest (RF) and Logistic Regression (LR), is evident from their high-performance stability following feature optimization. Random Forest achieved peak performance with an accuracy of 0.98 in the RFE scenario due to its ability to handle non-linear relationships and prioritize high-importance features. Meanwhile, Logistic Regression reached an accuracy of 0.97 through the Variance Threshold (VT) combination, proving that on low-dimensional datasets, this statistical model is highly reliable because it provides excellent interpretability and efficiency in distinguishing binary classes between Alzheimer's and non-Alzheimer's patients.

On the other hand, this study revealed a significant weakness in the Deep Learning (DL) architecture when subjected to conventional feature selection. Although the DL model maintained a competitive accuracy rate (ranging from 0.89 to 0.92), there was a sharp decline in precision, recall, and F1-score metrics. This weakness stems from the fundamental characteristic of Deep Learning, which requires rich and complex feature representations to optimize weights within its hidden layers. The use of feature selection methods like VT or BPSO, which drastically reduce input, effectively hinders the model's ability to capture deeper non-linear patterns, resulting in less precise positive classifications. Consequently, a more adaptive feature selection approach is required to accommodate the architectural complexity of neural network models in medical diagnosis.

IV. Conclusion

Alzheimer's Disease (AD) is a gradually advancing neurodegenerative condition that substantially impairs cognitive abilities and adversely affects the overall quality of life of affected individuals. Early detection is a crucial aspect of mitigation strategies, as it allows for more effective medical interventions before extensive neurological damage occurs. Advances in technology, such as the use of artificial intelligence (AI) and machine learning (ML), have created novel avenues for more accurate and efficient diagnosis.

This study evaluates the performance of five classification algorithms: Decision Tree, Random Forest, Naïve Bayes, Logistic Regression, and Deep Learning in two scenarios: without feature selection and with feature selection using Binary Particle Swarm Optimization (BPSO), Recursive Feature Elimination (RFE), and Variance Threshold (VT). The evaluation was conducted using K-fold cross-validation and four main metrics: accuracy, precision, recall, and F1-score.

The results demonstrate that feature selection consistently enhances model performance, particularly in conventional classification algorithms. Random Forest and Logistic Regression demonstrate high accuracy and prediction stability after feature selection is applied. The RFE and BPSO methods make significant contributions to increasing accuracy and model efficiency. In contrast, VT is effective for simple models but less optimal for complex architectures, such as those used in deep learning. Although Deep Learning demonstrates competitive accuracy, the decline in precision and F1-score suggests the need for more adaptive feature selection approaches to accommodate the model's structural complexity.

This study confirms that integrating intelligent feature selection techniques into machine learning models significantly enhances the accuracy and efficiency of early Alzheimer's disease detection. The experimental results demonstrate that classical algorithms, such as Random Forest and Logistic Regression, achieve superior and stable performance when optimized with Recursive Feature Elimination (RFE) and Binary Particle Swarm Optimization (BPSO). These findings underscore the potential of AI-based diagnostic strategies as reliable, non-invasive tools for clinical intervention.

However, this research acknowledges several limitations that provide a balanced perspective on the findings. First, while conventional feature selection methods significantly benefit classical models, they proved less optimal for Deep Learning architectures, leading to a notable decline in precision and F1-scores. This suggests that the current feature selection approach may not fully accommodate the structural complexity and non-linear representation needs of neural networks. Furthermore, the study relied on the Oasis Longitudinal dataset from Kaggle; therefore, the generalizability of these models to more diverse, real-world clinical populations require further validation. Future research should explore more adaptive, deep-learning-specific feature selection methods and incorporate multi-modal data to further refine diagnostic reliability.

Funding Information

This research received financial support from the Beginner Lecturer Research Grant (Penelitian Dosen Pemula), administered by the Directorate of Research and Community Service, Directorate General of Research and Development, Ministry of Higher Education, Science, and Technology, under Contract Nos. 123/C3/DT.05.00/PL/2025 and 109/LL2/DT.05.00/PL/2025.

Author Contributions Statement

The journal adopts the Contributor Roles Taxonomy (CRediT) as a framework to acknowledge the specific roles of each author, mitigate potential disputes concerning authorship, and enhance collaborative efforts. See the Table I below:

TABLE I. AUTHORS' CONTRIBUTION

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Suci Mutiara	✓	✓			✓				✓			✓		
Deppi Linda		✓						✓	✓				✓	
Siti Nur Laila		✓			✓				✓	✓				
Sri Lestari	✓	✓		✓					✓	✓				
Jean Anton i			✓					✓			✓			
Christian Petrus Silalahi			✓	✓				✓	✓		✓			

CONFLICT OF INTEREST STATEMENT

The authors hereby declare that no conflicts of interest exist in relation to this manuscript.

DATA AVAILABILITY

This study utilizes a dataset obtained from the Kaggle platform (https://www.kaggle.com/) on the topic of Alzheimer's Disease, specifically the Oasis Longitudinal dataset.

REFERENCES

[1] P. P. Phanieswar, K. S. Harshitha, V. Marka, B. Srinivasa Rao, and M. Aparna, "An optimised deep learning approach for alzheimer's disease classification," *IAES International Journal of Artificial Intelligence*, vol. 13, no. 3, pp. 3364–3370, Sep. 2024, doi: 10.11591/ijai.v13.i3.pp3364-3370.

[2] J.-L. Hsu *et al.*, "Early Alzheimer's disease (mild cognitive impairment or mild dementia): Prevalence, diagnostics, treatment options, and guidelines in Asia, Australasia, and Pacific nations countries," *J Prev Alzheimers Dis*, p. 100362, Aug. 2025, doi: 10.1016/j.tjpad.2025.100362.

[3] M. Zurrón *et al.*, "Plasma and neurostructural biomarkers in the clinical-biological characterization of early stages of the Alzheimer's disease continuum: findings from the Compostela Aging Study," *J Prev Alzheimers Dis*, p. 100331, Aug. 2025, doi: 10.1016/j.tjpad.2025.100331.

[4] N. Franzmeier *et al.*, "Predicting sporadic Alzheimer's disease progression via inherited Alzheimer's disease-informed machine-learning," *Alzheimer's and Dementia*, vol. 16, no. 3, pp. 501–511, Mar. 2020, doi: 10.1002/alz.12032.

[5] S. Kwak *et al.*, "Machine learning prediction of the mechanical properties of γ-TiAl alloys produced using random forest regression model," *Journal of Materials Research and Technology*, vol. 18, pp. 520–530, May 2022, doi: 10.1016/j.jmrt.2022.02.108.

[6] G. A. Jimenez-Maggiora, M. C. Donohue, M. S. Rafii, R. Raman, and P. S. Aisen, "Artificial intelligence-enabled safety monitoring in Alzheimer's disease clinical trials," *J Prev Alzheimers Dis*, vol. 12, no. 1, p. 100002, Jan. 2025, doi: 10.1016/j.tjpad.2024.100002.

[7] A. S. Khachaturian, A. Dengel, V. Dočkal, P. Hroboň, and M. Tolar, "Accelerating Innovations for Enhanced Brain Health. Can Artificial Intelligence Advance New Pathways for Drug Discovery for Alzheimer's and other Neurodegenerative Disorders?," Jan. 01, 2023, *Serdi-Editions*. doi: 10.14283/jpad.2023.1.

[8] J. Wallensten *et al.*, "Machine learning to detect Alzheimer's disease with data on drugs and diagnoses," *J Prev Alzheimers Dis*, vol. 12, no. 5, p. 100115, May 2025, doi: 10.1016/j.tjpad.2025.100115.

[9] N. Kulkarni and V. Bairagi, "Classification Algorithms in Diagnosis of Alzheimer's Disease," in *EEG-Based Diagnosis of Alzheimer Disease*, Elsevier, 2018, pp. 61–71. doi: 10.1016/b978-0-12-815392-5.00005-8.

[10] L. Yue *et al.*, "Predicting cognitive decline: Deep-learning reveals subtle brain changes in pre-MCI stage," *J Prev Alzheimers Dis*, vol. 12, no. 5, p. 100079, May 2025, doi: 10.1016/j.tjpad.2025.100079.

[11] C. Wang *et al.*, "A high-generalizability machine learning framework for predicting the progression of Alzheimer's disease using limited data," *NPJ Digit Med*, vol. 5, no. 1, Dec. 2022, doi: 10.1038/s41746-022-00577-x.

[12] R. Riska, "Analysing Musculoskeletal Bone Fractures Using Decision Trees: A Deep Learning Approach with Canny Segmentation and Hu Moments Feature Extraction," *International Journal of Artificial Intelligence in Medical Issues*, vol. 2, no. 1, pp. 30–40, May 2024, doi: 10.56705/ijaimi.v2i1.140.

[13] Ahmad Naswin and Adityo Permana Wibowo, "Performance Analysis of the Decision Tree Classification Algorithm on the Pneumonia Dataset," *International Journal of Artificial Intelligence in Medical Issues*, vol. 1, no. 1, pp. 1–9, May 2023, doi: 10.56705/ijaimi.v1i1.83.

[14] B. T. Jijo and A. M. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 1, pp. 20–28, Jan. 2021, doi: 10.38094/jastt20165.

[15] A. Sopandi, "Soil Infiltration Rate Impact on Water Quality Modeled Using Random Forest Regression," *Journal of Applied Data Sciences*, vol. 2, no. 4, pp. 135–142, 2021.

[16] N. Alfiyyah, "Optimizing Cardiomegaly Detection: A Random Forest Approach to Processed Chest X-ray Imagery," *International Journal of Artificial Intelligence in Medical Issues*, vol. 2, no. 2, pp. 115–124, Nov. 2024, doi: 10.56705/ijaimi.v2i2.156.

[17] F. H. Moh'd, K. A. Notodiputro, and Y. Angraini, "Enhancing interpretability in random forest: Leveraging inTrees for association rule extraction insights," *IAES International Journal of Artificial Intelligence*, vol. 13, no. 4, pp. 4054–4061, Dec. 2024, doi: 10.11591/ijai.v13.i4.pp4054-4061.

[18] M. M. Mijwil, A. W. Abdulqader, S. M. Ali, and A. T. Sadiq, "Null-values imputation using different modification random forest algorithm," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 1, pp. 374–383, Mar. 2023, doi: 10.11591/ijai.v12.i1.pp374-383.

[19] Y. Hari, M. E. Yanggah, and A. S. Paramita, "Assessing Novice Voter Resilience on Disinformation During Indonesia Elections 2024 with Naïve Bayes Classifier," *Journal of Applied Data Sciences*, vol. 6, no. 1, pp. 299–310, Jan. 2025, doi: 10.47738/jads.v6i1.489.

[20] A. Latubessy, R. Wardoyo, A. Musdholifah, and S. Kusrohmaniah, "Fine-Tuned Attribute Weighted Naïve Bayes with Modified Partial Instances Reduction for Gaming Disorder Classification," *HighTech and Innovation Journal*, vol. 6, no. 1, pp. 67–80, Mar. 2025, doi: 10.28991/HIJ-2025-06-01-05.

[21] D. Kundana and Chairani, "Data-Driven Analysis of Borobudur Ticket Sentiment Using Naïve Bayes," *APTISI Transactions on Technopreneurship*, vol. 5, no. 2Sp, pp. 221–233, Sep. 2023, doi: 10.34306/att.v5i2sp.353.

[22] R. A. Widyanto, M. Hadi Avizenna, N. A. Prabowo, K. Alfata, and A. Ismanto, "Data Mining Predicts the Need for Immunization Vaccines Using the Naive Bayes Method," *Journal of Applied Data Sciences*, vol. 2, no. 3, pp. 93–101, 2021.

[23] A. Kesumawati and A. K. Thalib, "Hoax Classification with Term Frequency-Inverse Document Frequency Using Non-Linear SVM and Naïve Bayes," *Int. J. Advance Soft Compu. Appl*, vol. 10, no. 3, 2018.

[24] Fatoni, T. B. Kurniawan, D. A. Dewi, M. Z. Zakaria, and A. M. M. Muhayeddin, "Fake vs Real Image Detection Using Deep Learning Algorithm," *Journal of Applied Data Sciences*, vol. 6, no. 1, pp. 366–376, Jan. 2025, doi: 10.47738/jads.v6i1.490.

[25] M. El Ghazi and N. Aknin, "A systematic review of non-intrusive human activity recognition in smart homes using deep learning," *IAES*

International Journal of Artificial Intelligence, vol. 13, no. 3, pp. 3188–3202, Sep. 2024, doi: 10.11591/ijai.v13.i3.pp3188-3202.

[26] U. Usman, F. Yunita, and M. R. Ridha, "Improving Classification Accuracy of Local Coconut Fruits with Image Augmentation and Deep Learning Algorithm Convolutional Neural Networks (CNN)," *Journal of Applied Data Sciences*, vol. 6, no. 1, pp. 1–19, Jan. 2025, doi: 10.47738/jads.v6i1.389.

[27] A. Sharma and P. Tanwar, "Model for autism disorder detection using deep learning," *IAES International Journal of Artificial Intelligence*, vol. 13, no. 1, pp. 391–398, Mar. 2024, doi: 10.11591/ijai.v13.i1.pp391-398.

[28] S. M. Orebi and A. M. Naser, "Opinion Mining in Text Short by Using Word Embedding and Deep Learning," *Journal of Applied Data Sciences*, vol. 6, no. 1, pp. 526–535, Jan. 2025, doi: 10.47738/jads.v6i1.438.

[29] P. Schober and T. R. Vetter, "Statistical Minute Logistic Regression in Medical Research," 2021. [Online]. Available: www.anesthesia-analgesia.org365

[30] J. K. Harris, "Primer on binary logistic regression," *Fam Med Community Health*, vol. 9, Dec. 2021, doi: 10.1136/fmch-2021-001290.

[31] N. R. Panda, J. K. Pati, J. N. Mohanty, and R. Bhuyan, "A Review on Logistic Regression in Medical Research," Apr. 01, 2022, *MedSci Publications*. doi: 10.55489/njcm.134202222.

[32] N. C. M. , J. Katiravan, G. S. , and C. E. J. I. , "A novel dual optimized IDS to detect DDoS attack in SDN using hypertuned RFE and deep grid network," *Cyber Security and Applications*, vol. 2, Jan. 2024, doi: 10.1016/j.csa.2024.100042.

[33] N. S. Santhosh, M. Kannan, K. Keerthika, and S. Akshay, "AGXLG: Enhancing Hypothyroidism Prediction with Recursive Feature Elimination and Boosting Techniques," in *Procedia Computer Science*, Elsevier B.V., 2025, pp. 3457–3467. doi: 10.1016/j.procs.2025.04.602.

[34] A. Hadianto and W. H. Utomo, "CatBoost Optimization Using Recursive Feature Elimination," vol. 9, pp. 169–178, 2024, doi: 10.15575/join.v9i1.1324.

[35] P. Theerthagiri, "Predictive analysis of cardiovascular disease using gradient boosting based learning and recursive feature elimination technique," *Intelligent Systems with Applications*, vol. 16, Nov. 2022, doi: 10.1016/j.iswa.2022.200121.

[36] J. Wang, W. Q. Gu, X. C. Zhao, Y. N. Jiang, and K. Da Xu, "Designing broadband cross-polarization conversion metasurfaces using binary particle swarm optimization algorithm," *Mater Des*, vol. 247, Nov. 2024, doi: 10.1016/j.matdes.2024.113419.

[37] B. A. Dada, N. I. Nwulu, and S. O. Olukanmi, "Bayesian optimization with Optuna for enhanced soil nutrient prediction: a comparative study with genetic algorithm and particle swarm optimization," *Smart Agricultural Technology*, vol. 12, Dec. 2025, doi: 10.1016/j.atech.2025.101136.

[38] M. Al Fatih Abil Fida, T. Ahmad, and M. Ntahobari, "Variance Threshold as Early Screening to Boruta Feature Selection for Intrusion Detection System," in *Proceedings of 2021 13th International Conference on Information and Communication Technology and System, ICTS 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 46–50. doi: 10.1109/ICTS52701.2021.9608852.

[39] Y. Siti Ambarwati and S. Uyun, "Feature Selection on Magelang Duck Egg Candling Image Using Variance Threshold Method," in *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 694–699. doi: 10.1109/ISRITI51436.2020.9315486.

[40] M. Conciatori, A. Valletta, and A. Segalini, "Improving the quality evaluation process of machine learning algorithms applied to landslide time series analysis," *Comput Geosci*, vol. 184, Feb. 2024, doi: 10.1016/j.cageo.2024.105531.

[41] A. Humphrey *et al.*, "Machine-learning classification of astronomical sources: estimating F1-score in the absence of ground truth," *Mon Not R Astron Soc Lett*, vol. 517, no. 1, pp. L116–L120, Nov. 2022, doi: 10.1093/mnrasl/slac120.

[42] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, Jan. 2020, doi: 10.1186/s12864-019-6413-7.