Overcoming imbalanced rice seed germination classification: enhancing accuracy for effective seedling identification

Muhlasah Novitasari Mara[1], Sidiq Syamsul Hidayat[1], Farika Tono Putri[2], Dwi Rahmawati[3], Sri Wahyuni[4], Muhamad Cahyo Ardi Prabowo[1], Noer Ni'mat Syamsul Kabir[5], Ragil Tri Indrawati[2]

[1]Department of Electrical Engineering, Politeknik Negeri Semarang, Semarang, Indonesia
[2]Department of Mechanical Engineering, Politeknik Negeri Semarang, Semarang, Indonesia
[3]Department of Agricultural Production, Politeknik Negeri Jember, Jember, Indonesia
[4]Research Centre for Food Corps, National Research and Innovation Agency (BRIN), Bogor, Indonesia
[5]Department of Research and Development, Monev Indonesia, Makassar, Indonesia

Article Info

Article history:

Received Nov 23, 2023
Revised Jul 17, 2024
Accepted Jul 26, 2024

Keywords:

Image classification
Logistic regression
Machine learning
Model performance
Unbalanced data

ABSTRACT

This study aimed to automatically classify rice seedling germination on day seven using image analysis. The categories included normal, abnormal, and dead seeds. Due to the rarity of abnormal seedlings, capturing their images resulted in imbalanced data. To address this, abnormal categories were combined into a single class. We compared logistic regression, random forest, and deep learning models (VGG19, VGG16, Alex Net) for classification. Surprisingly, logistic regression achieved the highest accuracy (93.89%) and F1-scores (0.96 normal, 0.81 abnormal) despite the imbalanced data and complex task. The effectiveness of logistic regression for rice seedling classification with imbalanced data has been demonstrated in this novel research. Historically, deep learning models dominate image recognition, but our findings suggest simpler models can excel in specific scenarios, especially with limited data availability. This highlights the importance of selecting models based on data characteristics. The urgency for this research stems from the need for efficient and accurate rice seedling evaluation. Improved classification can enhance agricultural practices and optimize resource allocation.

Corresponding Author:

Sidiq Syamsul Hidayat
Department of Electrical Engineering, Politeknik Negeri Semarang
Semarang 50275, Central Java, Indonesia
Email: sidiqsh@polines.ac.id

1. INTRODUCTION

Rice is one of the important food commodities that serve as a primary food source worldwide, including in Indonesia [1]. Meeting food needs sustainably is crucial in line with the continuous increase in population growth. The availability of rice in the community depends on cultivation efforts in the field. Seeds germination are one of the essential elements in plant cultivation that determine production efforts and yield quality [2]. Quality seeds, as indicated by certification of seeds germination, are necessary. Seed certification aims to maintain the genetic, physical, and physiological purity of seeds used by farmers through monitoring the seed production process in the field, seed processing, laboratory seed testing, and seed labelling [3]. To develop superior seeds and varieties, efforts are needed to detect the development of normal sprouts accurately, thus making the process of testing the quality of rice seeds and seed analysis faster, more effective, and more accurate. As of the current moment, the practice of observing and tallying the quantity of normal sprouts has been conducted manually by laboratory analysts. However, this manual procedure is susceptible to errors arising from fatigue factors, thus potentially compromising the integrity of the testing process. Furthermore, the method is time-consuming and vulnerable to subjectivity. Consequently, there arises a necessity to identify a more expeditious, precise, and straightforward approach to ascertain normal sprouts. Therefore, it is necessary to find a method of observing normal sprouts that is faster, more accurate, and simpler.

Genze *et al*. [4] indicate that the use of image analysis in seed quality testing shows a lower error rate and better performance than manual and conventional methods. The classification of rice seedlings was based on their germination levels, categorizing them as normal, abnormal, normal seedlings with decay, dead seeds, or fresh seeds that failed to germinate. During this research, the team encountered challenges in capturing images of the rare categories, including abnormal seedlings, dead seeds, normal seedlings with decay, and fresh seedlings that failed to germinate. Handling imbalanced class requires appropriate approaches such as oversampling, under sampling, to algorithmic techniques such as single class learning and ensemble learning. To overcome the problem of class imbalance in complex datasets, techniques such as deep learning are also used [5]. Therefore, this study aims to address the issue of class imbalance in the classification of rice seed germination and improve the model's performance by merging four minority classes into a single category called 'Poor'. The study utilized nine classification algorithms, including logistic regression, random forest, k-nearest neighbours (KNN), naïve bayesian, decision tree, artificial neural network (ANN), and convolutional neural network (CNN) with VGG16, VGG19, and Alex net architectures.

Previous research in the domain of seed classification using machine learning and deep learning methods has been conducted extensively by researchers globally. The selection of these methods in our study was influenced by their successful application in related research endeavours. For instance, Prabowo and Hidayat [6] propose a solution to assess the quality of rice grains by analyzing their shape through digital image processing techniques. Hidayat *et al.* [7] utilize CNN algorithms to classify two thousand images into two categories: superior seeds and non-superior seeds. These experimental results demonstrate that the system achieves a precision of 93% and a recall of 95% in classifying superior and non-superior seeds, respectively [8]. Created the rice seed germination evaluation system (RSGES), which was specifically designed to forecast rice seed germination using digital image processing and ANN techniques. RSGES achieved impressive precision rates, with only 7.66% false acceptance and 5.42% false rejection, while processing each image in just 8.31 seconds. Gulzar *et al.* [9] discussed the development of an identification and classification system for fourteen types of seeds using the VGG16 architecture. The classification accuracy of the model is exceedingly high, reaching 0.99. SVM used for the classification of three types of rice grains, namely basmati, bangs and brown rice [10]. The results showed a high SVM model accuracy of 0.92. Kiratiratanapruk *et al.* [11] classified rice seed using five machine learning techniques, i.e., logistic regression, linear discrimination analysis, KNN, support vector machine (SVM), and CNN. Those study resulted in best accuracy value at 95.15% for CNN method. However, all those previous studies [8]–[11] detected seed quality using only the seed's profile, i.e., colour, texture, and shape. The present study classified seed quality based on its germination development. In contrast, our study distinguishes itself by classifying seed quality based on germination development, addressing a significant gap in prior research. Sivakumar *et al.* [12] conducted a comprehensive analysis of seed germination rates by utilizing neural networks (NN) and comparing them to extensive datasets containing germinated and non-germinated samples. Their approach focused on analysing the germinated roots of seeds, leading to the development of the multi-level root metric ratio (MLRM) method. This method has shown promising results in accurately assessing seed germination rates. On the other hand, Huang *et al.* [13] explored the application of CNN and transfer learning for seed quality classification. Their study also involved a comparison with traditional machine learning algorithms. The experimental findings revealed that deep learning algorithms, particularly Google Net, outperformed machine learning algorithms significantly. Google Net achieved an impressive accuracy of 95%, while the machine learning algorithm (speeded up robust features (SURF)+SVM) achieved an accuracy of 79.2%. This showcases the superior capabilities of deep learning in seed quality classification tasks. Mohan and Raj [14] introduced an automated system to identify and classify rice seeds based on digital image processing techniques with SVM and NN classification methods. The results showed that SVM provided a higher accuracy of 91% compared to NN with an accuracy of 83%. Machine learning methods such as logistic regression, linear discriminant analysis (LDA), KNN, SVM, and deep learning techniques VGG16, VGG19, Xception, InceptionV3, InceptionResNetV2 used to classify fourteen varieties of *Oryza Sativa* rice [11]. The results showed the InceptionResNetV2 model gave the highest accuracy of 95.15%. Ahmed *et al.* [15] divide algorithms and techniques for image-based rice classification and gradation into five different approaches, namely: geometric, statistical, trained, unsupervised, and deep learning. Deep learning approaches have demonstrated more promising outcomes and are gaining popularity for future research. Machine learning methods such as logistic regression, decision tree, SVM, random forest, and naïve bayesian to classify five rice varieties namely Basmati, Arborio, Jasmine, Ipsala, and Karacadag [16]. The experimental results showed that random forest had the highest accuracy reaching 99.85%, not much different from the accuracy of decision trees which was 99.68%. Ruslan *et al.* [17] used image processing and machine learning techniques to classify wild rice seeds and cultivated rice based on morphology, colour, and texture. The machine learning methods used are decision tree, discriminant analysis, logistic regression, naïve bayesian, SVM, KNN, and essembler classifiers. The results showed that colour-based classification with the logistic regression model provided the highest accuracy and sensitivity, namely 0.98 and 0.85 respectively.

Research related to overcoming unbalanced classes in classification has also been widely conducted. Wu [18] explores the concept of class imbalance and how it can lead to misleading results like high overall accuracy despite poor performance on minority classes. Gosain and Sardana [19] conducted a comparison of different oversampling methods, including synthetic minority over-sampling technique (SMOTE), adaptive synthetic sampling (ADASYN), Borderline-SMOTE, and safe-level SMOTE, using various classifiers. Their research revealed that safe-level SMOTE, which generates minority instances more prominently around larger safe levels, outperforms SMOTE, ADASYN, and Borderline SMOTE in terms of accuracy performance. An overview of the literature [20] spanning the years 2000 to 2016, focusing on different approaches associated with random forest classification for addressing class imbalance. The scholarly review delineates that there are two effective methods, balanced random forest and weighted random forest, for addressing the issue of class imbalance. A literature review conducted [21] on the use of deep learning in handling unbalanced classroom data. The results show that deep learning methods, such as CNN and cost-sensitive learning, are effective in addressing classroom imbalance issues. Mahmudah *et al.* [22] took a pre-processing approach by converting binary features into numerical features using the feature extraction method to overcome the limitations of the oversampling method in generating diverse new samples of binary features. The results indicated that the combination of random forest and oversampling with the relocating-safe-level SMOTE (RSLS) method provided the highest performance on all datasets tested. However, there are variations in the best feature extraction method for each dataset, this is estimated because it is affected by the imbalance ratio. Although most performance improvements are not statistically significant, there is a significant improvement in accuracy with the use of t-distributed stochastic neighbor embedding (t-SNE) feature extraction. Kumar *et al.* [23] of various approaches presented to classifying unbalanced datasets and their applications. The results of the review show that the sampling method is easy to implement, but in real-world applications involving biased data distribution, a hybrid approach is also required. Unbalanced data classification is the subject of extensive research in machine learning. Feng *et al.* [24] provided guidance in choosing the right resampling technique and classification method for classification problems with class imbalance. The random forest algorithm used on the Google Earth Engine platform to classify rice phenology from landsat-8 satellite imagery [25]. Oversampling techniques are used to overcome the problem of data imbalance. The results showed the accuracy of random forest with over sampling is better than random forest without oversampling. Zheng *et al.* [26] provides guidance for selecting appropriate methods in overcoming the problem of unbalanced data classification. A comprehensive investigation conducted into the challenges associated with addressing imbalance issues in the field of object detection [27]. Their research reveals that at present, there is no single method that can be universally applied to all architectural setups to effectively mitigate objective imbalances.

The aforementioned literature presents valuable insights into methodologies for addressing class imbalance in different domains. It is notable that there is a scarcity of studies addressing the specific challenges posed by imbalanced datasets in rice seed germination classification. Despite the abundance of research on class imbalance within the broader fields of machine learning and deep learning, there has been no previous investigation into addressing class imbalance specifically in the realm of rice seed germination-based classification. This highlights a significant research gap that our study seeks to bridge.

2. METHOD

In this study, we classified rice seedlings based on their germination levels into five categories: normal, abnormal, dead seeds, normal seedlings with decay, and fresh seeds that failed to germinate. For the "Normal" classification, we identified several criteria that must be met by the rice seedlings. Firstly, the primary root must grow strong, accompanied by well-developed secondary roots, with a minimum of 2 to 3 well-developed secondary seminal roots. Additionally, the primary leaf must grow along the coleoptile and emerge from it in a healthy condition. The plumule, the part of the plant that will develop into the stem, must also be present without any signs of decay. Furthermore, the development of the hypocotyl, the part of the stem below the cotyledons, must be perfect without any damage to its tissues. "Normal" seedlings must also have one cotyledon for monocot seedlings and two cotyledons for dicot seedlings. For seeds with epigeal germination type, the length of the root should be at least four times the length of the seed and have a normal structural development. Lastly, seedlings that have been damaged due to infection by other seedlings are still considered normal as long as the essential parts of the seedlings remain intact. For the "Abnormal" classification, seedlings must meet the following criteria. Firstly, no primary or secondary roots grow, or if they grow, the roots are weak and short. Additionally, the seedlings do not grow their first leaf, and their coleoptile is colorless. Sometimes, the plumule may grow white or experience decay. Seedlings that are damaged, lack cotyledons, embryos, and short primary roots are also included in the "Abnormal" classification. Moreover, seedlings showing deformities, weak development, or imbalance in essential parts will also be categorized as "Abnormal". Seeds that decay before germination or fail to grow within the specified testing period, but are not in a dormant state, will be classified as "Dead seeds". The "Normal seedlings with decay" classification is used for seeds that remain hard at the end of the germination test due to not absorbing water, caused by an impermeable seed coat that hinders the germination process. Lastly, the classification of "Fresh seeds that failed to germinate" is given to seeds that do not germinate by the end of the testing period but still have the potential to grow normally. This type of seed can absorb water during the testing process but experiences obstacles in further development.

In this study, we tested the germination of rice seeds for 7 days on blotting paper. Subsequently, we captured images of each seedling. Figure 1 shows the data acquisition procedure in this study. To capture high-quality images of germinating rice seedlings, a Nikon D5600 digital camera with a resolution of 24.2 megapixels was used for data acquisition tool. This camera provides exceptional image quality and detail. Images taken in a 60×60 cm Minibox studio. The Minibox studio provides controlled lighting conditions and consistent backgrounds, ensuring uniformity across datasets. Diffused and uniform lighting was maintained in a studio setting to eliminate shadows and guarantee consistent lighting across all seedlings. Neutral, non-reflective backgrounds are used in the studio to minimize distractions and maintain focus on the seedlings. Each germinating rice seedling was photographed individually from above the studio box with a distance between the camera and the object as far as 60 cm (as high as the studio box).

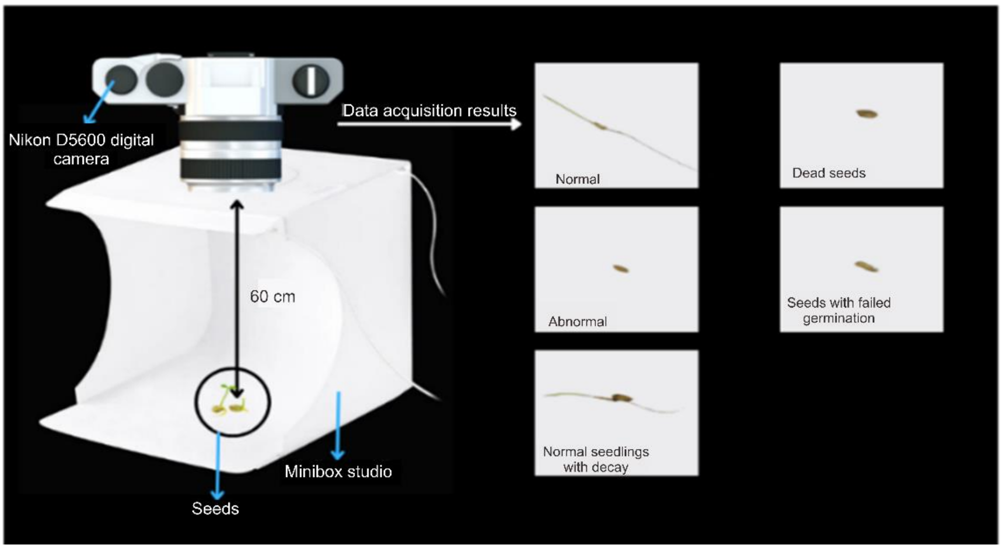

Figure 1. Data acquisition procedure

Based on their germination levels, rice seedlings were classified as normal, abnormal, normal seedlings with decay, dead seeds, or fresh seeds that failed to germinate. The image data was ensured to be of high quality to ensure accurate classification. The results of the image capture for each classification are as follows normal seedlings 753 images, abnormal seedlings: 68 images, dead seeds: 10 images, normal seedlings with decay: 50 images, and fresh seeds that failed to germinate: 13 images. Thus, the total number of images successfully collected for this study is 894 images, which will be used for the classification analysis.

The research conducted in this investigation faced difficulties in capturing images of abnormal seedlings, dead seeds, normal seedlings with decay, and fresh seedlings that failed to germinate due to their rarity. As a result, the number of images for these categories was considerably lower than that for normal seedlings. To address this issue, the four criteria were merged into a single category labeled "Poor." Consequently, the dataset consisted of 753 image data of normal rice seedlings classified as "Good" and 141 data categorized as "Poor." Figure 2 illustrates the classification of rice seedlings, depicting the distribution of images between the "Good" and "Poor" categories. This structured visualization allows for a clear understanding of the imbalance in the dataset, with a larger proportion of images falling under the "Good" category compared to the "Poor" category. By merging the rare categories into a single group, we aimed to enhance the robustness of our classification model while addressing the imbalance in the dataset.

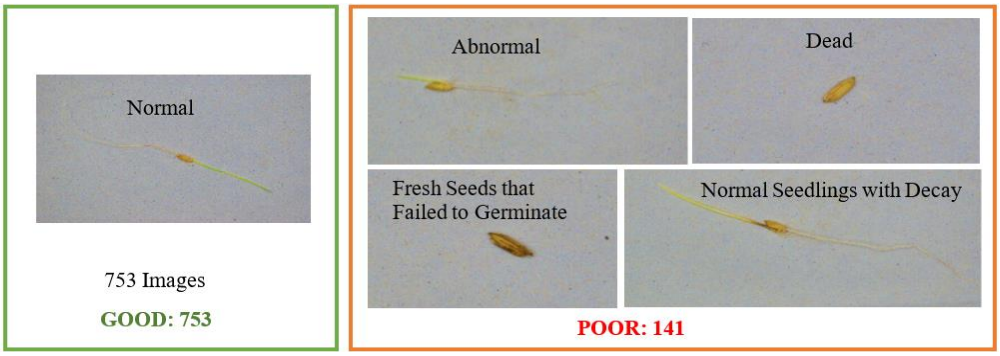

Figure 2. Classification of rice seedlings

The nine classification algorithms used in this study were logistic regression, random forest, KNN, naïve Bayesian, decision tree, ANN, VGG16, VGG19, and Alex Net were tested to classify rice seedling germination into good or bad. The results of each algorithm are given in the form of accuracy, confusion matrix, and classification report. Before the images were fed into the machine learning and deep learning models for classification, a series of preprocessing steps were applied to enhance the quality and consistency of the data. All images were resized to a uniform dimension of 227×227 pixels. This standardization was essential to ensure that the models received consistent input sizes. Each image was normalized by rescaling pixel values to the range [0, 1]. Normalization helps in speeding up the training process and allows the models to converge faster. The images were organized into batches to efficiently train the machine learning and deep learning models. In this case, a batch size of 32 was used. The dataset was then divided into two parts, with 80% of the data used for training and 20% for testing. So that the training data for the good class was 603 images data and for poor data as much as 111 images data. While 150 good class image data and 30 poor class image data were used as testing data. This splitting ensured that the models were trained on a representative portion of the data and evaluated on unseen samples to assess their generalization performance. The ImageDataGenerator function from the Keras library was utilized to efficiently load, preprocess, and augment image data.

Accuracy, precision, recall, and F1 score utilizing a confusion matrix were employed as intuitive performance indicators to assess the performance of the classification model. To generate a confusion matrix, the number of true positives, true negatives, false positives, and false negatives must be calculated. The following mathematical method is used to calculate the performance indicator [28].

$$Accuracy = (TP+TN)/(TP+TN+FP+FN)$$

$$Precision = TP/(TP+FP)$$

$$Recall = TP/(TP+FN)$$

$$F1\text{-}score = (2 \times Precision \times Recall\ Precision)/(Recall+Recall)$$

The confusion matrix illustrates the predictions provided by machine learning algorithms and the actual occurrences. Accuracy is the ratio of true positives to true negatives for each testing data set. The accuracy of the algorithm affects how well it predicts. Precision is defined as the proportion of accurately predicted positive test findings to all projected positive test results. When compared to all test data, recall indicates how often forecasts turn out to be correct. The F1 Score is computed by averaging precision and recall.

3. RESULTS AND DISCUSSION

Results from the logistic regression, random forest, KNN, naïve Bayesian, decision tree, VGG16, VGG19, and Alex Net algorithms showed different accuracy and performance. In the results of the analysis using logistic regression, an accuracy of 0.94 was obtained, indicating that this model succeeded well in classifying the germination of rice seedlings into good or poor categories. In the confusion matrix, out of 150 rice seed data that were good, 146 of them were correctly classified as good (true positive), while 4 seedlings were incorrectly classified as poor (false negative). Of the 30 poor rice seed data, 23 of them were correctly classified as poor (true negative), but 7 poor seedlings were incorrectly classified as good (false positive). The classification report shows that the model precision is 0.77 for the poor class, and 0.97 for the good class. While the recall for good and poor d classes is 0.95 and 0.85 respectively. In the F1-Score, which combines information about precision and recall, the good class has a score of 0.96 and the poor class has a score of 0.81. High F1-score in both classes indicate that the model performs well in classifying them. The execution time of the logistic regression model is about 16.7 seconds, which shows that this model has a relatively fast processing time in classifying rice seed data. This suggests its effectiveness for this classification task. For instance, Rahman and Yap [29] investigated the effect of imbalanced ratios on logistic regression, confirming that imbalance impacts parameter estimates and classification. Rahman *et al.* [30] provide evidence that logistic regression can still perform well with imbalanced data, although the imbalance ratio affects parameter estimates and performance metrics like precision and recall, which is similar to our reported results.

The results of the analysis using the random forest algorithm showed an accuracy of 0.91. However, this accuracy value is dominated by the accuracy of the model predicting the majority class. This can be seen from the confusion matrix which shows that out of a total of 150 data on rice seeds that are actually good, all of them are classified correctly as good (true positive), so there is no error in identifying good rice seeds However, from 30 poor rice seed data, only 14 of them are correctly classified as poor, while 16 poor seeds are incorrectly classified as good (false positive). In the F1-Score, which combines information about precision and recall, the good class has a very high score, which is 0.95, while the poor class has a score of 0.64. These results indicate that the random forest model has difficulty in identifying poor rice seedlings. The execution time of the random forest model is about 11.9 seconds, which indicates that this model has a relatively fast processing time in classifying rice seed data. Shu *et al*. [31] have identified similar constraints of random forest algorithms in managing imbalanced data, recommending the generation of minority class samples using generative adversarial network (GAN) to enhance classification performance. More and Rana [32] review various techniques like weighted random forest and balanced random forest to tackle the class imbalance problem, acknowledging the limitations of standard random forests in imbalanced scenarios. Khoshgoftaar *et al.* [33] provide an extensive empirical evaluation of random forests built from imbalanced data, highlighting the challenges and limitations faced by the algorithm, particularly its tendency to favor the majority class.

The results of the analysis using the KNN algorithm showed a low accuracy of 0.36, indicating that this model had difficulty in classifying rice seed germination into good or poor categories. The confusion matrix shows that out of a total of 150 rice seed data that are actually good, only 36 are correctly classified as good (true positive), while 114 good seedlings are incorrectly classified as poor (false negative). Of the 30 poor rice seed data, only 29 of them were correctly classified as poor (true negative), while only 1 poor seed was incorrectly classified as good (false positive). The classification report shows that the model has a high recall, which is 0.97, for the good class but for the poor class, the low precision is 0.20. In the F1-score, which combines information about precision and recall, the good class has a score of 0.39, while the poor class has a score of 0.34. The execution time of the KNN model is very fast, which is about 0.21 seconds, but the low performance in classifying rice seedlings requires improved model performance to produce better and more accurate classification results. Related to this result, Sun and Chen [34] provide a comprehensive survey of KNN algorithms and highlight the difficulty KNN faces with imbalanced data, which often leads to poor performance. In their thorough analysis of KNN algorithms, [34] point out that KNN frequently performs poorly when faced with unbalanced data.

The results of the analysis using the naïve Bayesian algorithm showed an accuracy of 0.54, indicating that this model has a low performance in classifying rice seed germination into good or poor categories. The confusion matrix shows that out of a total of 150 rice seed data that are actually good, only 72 are correctly classified as good (true positive), while 78 good seedlings are incorrectly classified as poor (false negative). Of the 30 poor rice seed data, only 26 of them were correctly classified as poor (true negative), while 4 poor seedlings were incorrectly classified as good (false positive). The classification report shows that the model has a high recall rate, which is 0.95, for the good class however, for the poor class the recall value is very low at 0.25. In the F1-score, which combines information about precision and recall, the good class has a score of 0.64, while the poor class has a score of 0.39. This indicates that there is difficulty in identifying poor rice seedlings. The execution time of the naïve Bayesian model is about 0.78 seconds, which indicates that this model has a fairly fast processing time. Although the accuracy is still above 0.5, the classification results in poor grades need to be improved to improve the model’s performance in better identifying poor rice seedlings. Highlight similar difficulties of naïve Bayes faces in imbalanced datasets and the impact on classification accuracy [35].

The results of the analysis using the decision tree algorithm showed an accuracy of 0.74, which indicates that this model has a fairly good level of accuracy in classifying rice seed germination into good or poor categories. The confusion matrix shows that out of a total of 150 data on rice seeds that are actually good, 112 of them are correctly classified as good (true positive), while 38 good seedlings are incorrectly classified as poor (false negative). Of the 30 poor rice seed data, 21 of them were correctly classified as poor (true negative), but 9 poor seedlings were incorrectly classified as good (false positive). The classification report shows that the model has a high recall rate, which is 0.93, for the good class but for the poor class, the recall value is low at 0.36. In the F1-score, which combines information about precision and recall, the good class has a score of 0.83, while the poor class has a score of 0.47 signaling the difficulty in identifying poor rice seedlings. The execution time of the decision tree model is approximately 77.73 seconds, which indicates that this model takes a considerable amount of time for processing. Although the accuracy is quite good, the classification results in poor grades need to be improved to improve the performance of the model in better identifying poor rice seedlings. In addition, models also need to be optimized to require shorter processing times. Truică and Leordeanu [36] discuss various decision tree algorithms applied to imbalanced datasets and their respective performance improvements.

The ANN, VGG19, and Alex Net architectures provide the same test results. The ANN architecture consists of a sequential model with a flattening layer to convert input images into a one-dimensional vector, a hidden layer with 128 neurons and rectified linear unit (ReLU) activation, and an output layer with a sigmoid activation for binary classification. The model is compiled with the Adam optimizer and binary cross-entropy loss. The Alex Net architecture commences with convolution layer 1, featuring 96 filters with an 11×11 receptive field and ReLU activation. A 4×4 stride is applied to accommodate 227×227 pixel RGB input images. Subsequently, max pooling layer 1 with a 3×3 window and 2×2 stride reduces spatial dimensions. Convolution layer 2 follows, comprising 256 5×5 filters with ReLU activation, and is succeeded by max pooling layer 2. Convolution layers 3, 4, and 5 introduce 384, 384, and 256 3×3 filters with ReLU activation, progressively capturing hierarchical features. Max pooling layer 3 continues spatial reduction. The data is flattened and processed through two fully connected layers with 512 units, applying ReLU activation and dropout of 0.5 for regularization. The architecture concludes with a single-unit output Layer with sigmoid activation for binary classification. Model training employs the Adam optimizer and binary cross-entropy loss.

The foundational element within the architecture of the VGG19 base model is the VGG19 CNN, renowned for its adeptness in extracting intricate hierarchical features from images. In this specific instantiation, its role is that of a feature extractor. Subsequent to the VGG19 convolutional layers, a flatten layer is introduced, serving the purpose of reshaping the multi-dimensional output into a singular one-dimensional vector. Two densely connected layers, each comprising 4096 units, are incorporated into the architecture. These layers establish comprehensive interconnections with units from the preceding layer, and ReLU activation functions are applied to modulate their outputs. The terminal layer consists of a solitary unit and employs a sigmoid activation function, conforming to the customary configuration for binary classification tasks. Subsequently, the model is meticulously assembled, integrating the Adam optimizer and binary cross-entropy loss function.

The accuracy value of ANN, VGG19, and Alex Net models is 0.83 or 83%. However, when looking at the confusion matrix, the model cannot distinguish poor classes (minority classes) perfectly because it has a very low recall value and F1-score for that class. The confusion matrix results showed that out of a total of 30 data that should belong to class 1, the model failed in classifying any of them, resulting in 30 false negatives. In contrast, for either class (majority class), the model managed to classify all data correctly, resulting in 150 true positives. This causes the precision and recall for class 1 to zero. From the precision, recall, and F1-score generated, it can be seen that the model has problems in recognizing minority classes well, and this can be caused by an imbalance in the amount of data between the majority class and the minority class. ANN, VGG19, and Alex Net models achieved similar accuracy (83%) but suffered from data imbalance, as shown by their high F1-scores for the majority class and an F1-Score of zero for the minority class, highlighting this limitation as seen in [21], [37].

The architecture for VGG16 closely parallels the structure and rationale of VGG19, as previously described. In the context of testing, the VGG16 architecture yielded an accuracy of 0.65 or 65%. However, when looking at the confusion matrix, the model has problems classifying both classes. The model correctly classified 113 data into good class but misclassified 37 data from good class into false positives. In addition, the model also managed to classify 4 data from class 1 correctly but failed to classify 26 data from class 1 as class good (false negatives). From the precision, recall, and F1-score generated, the model has problems recognizing both classes well. Precision for good classes is quite high (0.75), but precision for poor classes is very low (0.13). This indicates that the model tends to classify data into majority classes well but has problems classifying data into minority classes. Similarly, recall for good classes is quite high (0.81), but recall for poor classes is also low (0.10). Low recall indicates that the model failed to identify most of the data from minority classes. Shah and Manjula [38] emphasized that while VGG16 is effective for balanced datasets, its performance drops in imbalanced scenarios.

To enhance the clarity and comprehensiveness of our evaluation, we have thoughtfully compiled the key performance metrics for each method in Table 1 and represented the method's performance graphically in Figure 3. The results of our assessments reveal logistic regression as the leading performer, achieving an impressive accuracy of 0.94 (94%), with F1-scores of 0.96 for the "Good" class and 0.81 for the "Poor" class. Random forest follows closely behind with an accuracy of 0.91 (91%). However, a more detailed examination of precision and F1-scores, particularly for the minority classes (the "Poor" class), uncovers a significant limitation in random forest, characterized by low precision and F1-Scores, indicating challenges in effectively identifying these minority classes. Importantly, this limitation extends beyond random forest to encompass more complex models like ANN, VGG19, and Alex Net. Despite their high overall accuracy, these sophisticated models exhibit an F1-score of 0 for the minority class. This observation highlights the substantial impact of data imbalance, where the minority class is notably underrepresented compared to the majority class. Similar findings were reported by Luo *et al*. [39], who compared the performance of random forest and logistic regression on imbalanced datasets, noting that logistic regression often outperforms random forest in terms of handling imbalanced data.

This study underscores that while complex models such as ANN, VGG19, and Alex Net have the capability to discern intricate image features. However, they might not be the most suitable choice for our rice sprouts image classification tasks involving imbalanced data. In such scenarios, simpler models like logistic regression not only yield superior results but also offer the advantage of relatively shorter execution times, rendering them a more efficient and effective option. These findings emphasize the critical importance of judiciously selecting models based on the specific characteristics of the dataset and the objectives at hand.

Table 1. Comprehensive comparison of classification results

Model	Accuracy	Precision (Good)	Precision (Poor)	Recall (Good)	Recall (Poor)	F1 score (Good)	F1 score (Poor)	Execution time (s)
Logistic regression	0.94	0.97	0.77	0.95	0.85	0.96	0.81	16.69
Random forest	0.91	1.00	0.47	0.90	1.00	0.95	0.64	11.93
KNN	0.36	0.24	0.97	0.97	0.20	0.39	0.34	0.21
Naïve Bayesian	0.54	0.48	0.87	0.95	0.25	0.64	0.39	0.78
Decision tree	0.74	0.75	0.70	0.93	0.36	0.83	0.47	77.73
ANN	0.83	1.00	0.00	0.83	0.00	0.91	0.00	11013.70
VGG 19	0.83	1.00	0.00	0.83	0.00	0.91	0.00	8124.89
Alex Net	0.83	1.00	0.00	0.83	0.00	0.91	0.00	8021.24
VGG 16	0.65	0.75	0.13	0.81	0.10	0.78	0.11	4720.69

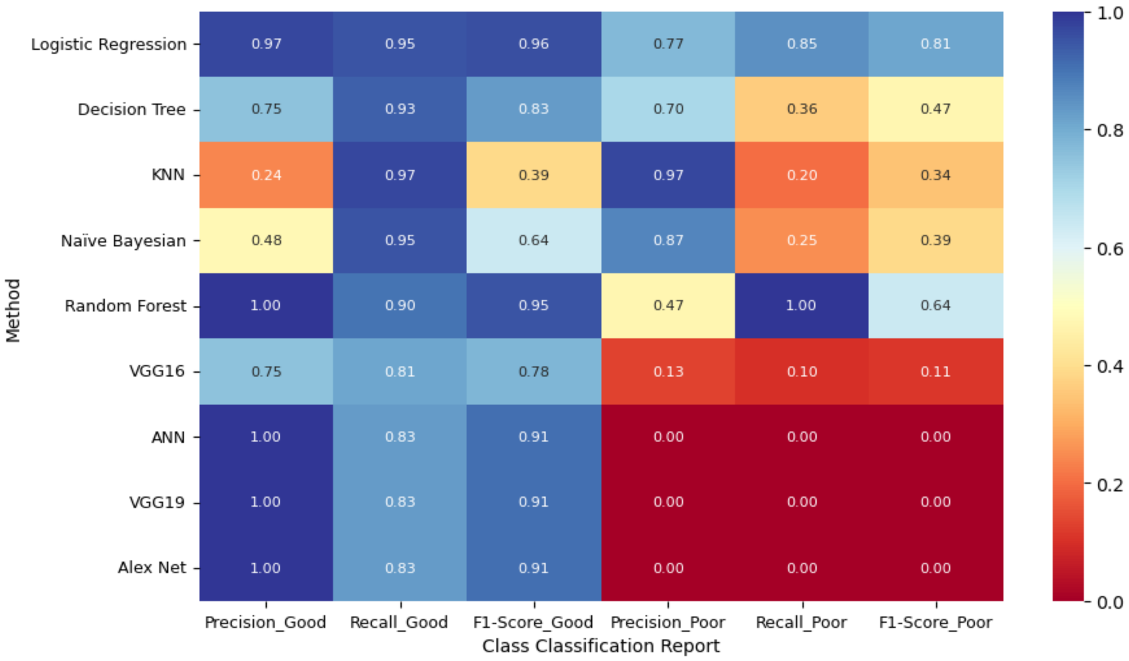

Figure 3. Method’s performance

4. CONCLUSION

In conclusion, our evaluation of various classification methods for rice sprout image classification has provided valuable insights. Logistic regression emerged as the top performer with an impressive accuracy of

94%, demonstrating its effectiveness in this task. Random forest followed closely with an accuracy of 91%, although it exhibited limitations in recognizing minority classes due to low precision and F1-scores. This limitation extended to more complex models like ANN, VGG19, and Alex Net, which, despite their high overall accuracy, struggled with the minority class. This underscores the impact of data imbalance in classification tasks. Our study highlights the importance of model selection based on dataset characteristics and classification goals. While complex models have the capacity to analyze intricate image features, they may not be the most suitable choice for tasks involving imbalanced data, such as rice sprout classification. In such scenarios, simpler models like logistic regression offer superior results and shorter execution times, making them a more efficient option. These findings emphasize the need for a thoughtful approach to model selection, considering both accuracy and the specific challenges posed by imbalanced data. Ultimately, the choice of model should align with the objectives of the classification task and the characteristics of the dataset to achieve optimal results. Moving forward, future research should focus on developing novel techniques to address imbalanced data challenges effectively. Additionally, exploring ensemble methods that combine the strengths of different models could offer improved performance in handling imbalanced datasets. Moreover, investigating techniques for generating synthetic data to balance class distributions and refining feature engineering processes specific to rice sprout images could further enhance classification accuracy. Overall, continued exploration in these directions will contribute to advancing the field of rice sprout image classification and addressing real-world challenges in agriculture and food security.

ACKNOWLEDGEMENTS

We would like to express our deepest gratitude to the Indonesian National Research and Innovation Agency for the funding provided in this research. The funding was provided in collaboration with the Education Fund Management Agency of the Ministry of Finance under the Research and Innovation for Advance Indonesia funding scheme, contract number 84/IV/KS/11/2022. We also extend our gratitude to Politeknik Negeri Semarang for the facilitation provided. In addition, we would also like to express our gratitude to Politeknik Negeri Jember for their assistance in providing facilities for data collection of rice seed sprouts.

REFERENCES

[1] F. Rozi *et al.*, "Indonesian market demand patterns for food commodity sources of carbohydrates in facing the global food crisis," *Heliyon*, vol. 9, no. 6, Jun. 2023, doi: 10.1016/j.heliyon.2023.e16809.
[2] P. A. Tuan, M. Sun, T. N. Nguyen, S. Park, and B. T. Ayele, "Molecular mechanisms of seed germination," in *Sprouted Grains: Nutritional Value, Production, and Applications*, Elsevier, 2018, pp. 1–24, doi: 10.1016/B978-0-12-811525-1.00001-4.
[3] L. O. Copeland and M. B. McDonald, "Seed Enhancements," in *Principles of Seed Science and Technology*, Boston, MA: Springer US, 2001, pp. 277–296, doi: 10.1007/978-1-4615-1619-4_13.
[4] N. Genze, R. Bharti, M. Grieb, S. J. Schultheiss, and D. G. Grimm, "Accurate machine learning-based germination detection, prediction and quality assessment of three grain crops," *Plant Methods*, vol. 16, no. 1, Dec. 2020, doi: 10.1186/s13007-020-00699-x.
[5] S. S. Rawat and A. K. Mishra, "Review of methods for handling class-imbalanced in classification problems," *arXiv-Computer Science*, pp. 1-15, 2022, doi: 10.48550/arXiv.2211.05456.
[6] M. C. A. Prabowo and S. S. Hidayat, "Edge Detection technique for rice quality analysis using digital image processing," in *AIP Conference Proceedings*, 2023, vol. 2431, no. 1, doi: 10.1063/5.0117510.
[7] S. S. Hidayat, D. Rahmawati, M. C. A. Prabowo, L. Triyono, and F. T. Putri, "Determining the rice seeds quality using convolutional neural network," *International Journal on Informatics Visualization*, vol. 7, no. 2, pp. 527–534, Jun. 2023, doi: 10.30630/joiv.7.2.1175.
[8] B. Lurstwut and C. Pornpanomchai, "Image analysis based on color, shape and texture for rice seed (*Oryza sativa L.*) germination evaluation," *Agriculture and Natural Resources*, vol. 51, no. 5, pp. 383–389, Oct. 2017, doi: 10.1016/j.anres.2017.12.002.
[9] Y. Gulzar, Y. Hamid, A. B. Soomro, A. A. Alwan, and L. Journaux, "A convolution neural network-based seed classification system," *Symmetry*, vol. 12, no. 12, pp. 1–18, Dec. 2020, doi: 10.3390/sym12122018.
[10] S. Ibrahim, N. A. Zulkifli, N. Sabri, A. A. Shari, and M. R. M. Noordin, "Rice grain classification using multi-class support vector machine (SVM)," *IAES International Journal of Artificial Intelligence*, vol. 8, no. 3, pp. 215–220, Dec. 2019, doi: 10.11591/ijai.v8.i3.pp215-220.
[11] K. Kiratiratanapruk *et al.*, "Development of paddy rice seed classification process using machine learning techniques for automatic grading machine," *Journal of Sensors*, vol. 2020, pp. 1–14, Jul. 2020, doi: 10.1155/2020/7041310.
[12] D. Sivakumar, K. Suriyakrishnaan, P. Akshaya, G. V. Anuja, and G. T. Devadharshini, "Computerized growth analysis of seeds using deep learning method," *International Journal of Recent Technology and Engineering*, vol. 7, no. 6, pp. 1885–1892, 2019.
[13] S. Huang, X. Fan, L. Sun, Y. Shen, and X. Suo, "Research on classification method of maize seed defect based on machine vision," *Journal of Sensors*, vol. 2019, pp. 1–9, Nov. 2019, doi: 10.1155/2019/2716975.
[14] D. Mohan and M. G. Raj, "Quality analysis of rice grains using ANN and SVM," *Journal of Critical Reviews*, vol. 7, no. 1, pp. 395–402, Jan. 2020, doi: 10.31838/jcr.07.01.79.
[15] S. B. Ahmed, S. F. Ali, and A. Z. Khan, "On the frontiers of rice grain analysis, classification and quality grading: a review," *IEEE Access*, vol. 9, pp. 160779–160796, 2021, doi: 10.1109/ACCESS.2021.3130472.
[16] P. Saxena, K. Priya, S. Goel, P. K. Aggarwal, A. Sinha, and P. Jain, "Rice varieties classification using machine learning algorithms," *Journal of Pharmaceutical Negative Results*, vol. 13, 2022, doi: 10.47750/pnr.2022.13.S07.479.
[17] R. Ruslan, S. Khairunniza-Bejo, M. Jahari, and M. F. Ibrahim, "Weedy rice classification using image processing and a machine learning approach," *Agriculture*, vol. 12, no. 5, Apr. 2022, doi: 10.3390/agriculture12050645.
[18] O. Wu, "Rethinking class imbalance in machine learning," *arXiv-ComputerScience*, 2023.

[19] A. Gosain and S. Sardana, "Handling class imbalance problem using oversampling techniques: A review," in *2017 International Conference on Advances in Computing, Communications and Informatics*, Sep. 2017, pp. 79–85, doi: 10.1109/ICACCI.2017.8125820.
[20] S. J. Basha, S. R. Madala, K. Vivek, E. S. Kumar, and T. Ammannamma, "A review on imbalanced data classification techniques," in *2022 International Conference on Advanced Computing Technologies and Applications, ICACTA 2022*, Mar. 2022, pp. 1–6, doi: 10.1109/ICACTA54488.2022.9753392.
[21] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0192-5.
[22] K. R. Mahmudah, F. Indriani, Y. Takemori-sakai, Y. Iwata, T. Wada, and K. Satou, "Classification of imbalanced data represented as binary features," *Applied Sciences*, vol. 11, no. 17, Aug. 2021, doi: 10.3390/app11177825.
[23] P. Kumar, R. Bhatnagar, K. Gaur, and A. Bhatnagar, "Classification of imbalanced data:review of methods and applications," *IOP Conference Series: Materials Science and Engineering*, vol. 1099, no. 1, Mar. 2021, doi: 10.1088/1757-899x/1099/1/012077.
[24] Y. Feng, M. Zhou, and X. Tong, "Imbalanced classification: A paradigm-based review," *Statistical Analysis and Data Mining*, vol. 14, no. 5, pp. 383–406, Oct. 2021, doi: 10.1002/sam.11538.
[25] H. Suryono, H. Kuswanto, and N. Iriawan, "Rice phenology classification based on random forest algorithm for data imbalance using Google Earth engine," *Procedia Computer Science*, vol. 197, pp. 668–676, 2021, doi: 10.1016/j.procs.2021.12.201.
[26] M. Zheng, F. Wang, X. Hu, Y. Miao, H. Cao, and M. Tang, "A method for analyzing the performance impact of imbalanced binary data on machine learning models," *Axioms*, vol. 11, no. 11, Nov. 2022, doi: 10.3390/axioms11110607.
[27] K. Oksuz, B. C. Cam, S. Kalkan, and E. Akbas, "Imbalance problems in object detection: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3388–3415, Oct. 2021, doi: 10.1109/TPAMI.2020.2981890.
[28] J. Ha, M. Kambe, and J. Pe, *Data mining: concepts and techniques*, Burlington: Morgan Kaufmann, 2011, doi: 10.1016/C2009-0-61819-5.
[29] H. A. A. Rahman and B. W. Yap, "Imbalance effects on classification using binary logistic regression," in *Communications in Computer and Information Science*, vol. 652, 2016, pp. 136–147, doi: 10.1007/978-981-10-2777-2_12.
[30] H. A. A. Rahman, Y. B. Wah, and O. S. Huat, "Predictive performance of logistic regression for imbalanced data with categorical covariate," *Pertanika Journal of Science and Technology*, vol. 28, no. 4, pp. 1141–1161, Oct. 2020, doi: 10.47836/pjst.28.4.02.
[31] Q. Shu, T. Hu, and S. Liu, "Random forest algorithm based on GAN for Imbalanced data classification," *Journal of Physics: Conference Series*, vol. 1544, no. 1, 2020, doi: 10.1088/1742-6596/1544/1/012014.
[32] A. S. More and D. P. Rana, "Review of random forest classification techniques to resolve data imbalance," in *1st International Conference on Intelligent Systems and Information Management*, Oct. 2017, pp. 72–78, doi: 10.1109/ICISIM.2017.8122151.
[33] T. M. Khoshgoftaar, M. Golawala, and J. Van Hulse, "An empirical study of learning from imbalanced data using random forest," in *International Conference on Tools with Artificial Intelligence*, Oct. 2007, vol. 2, pp. 310–317, doi: 10.1109/ICTAI.2007.46.
[34] B. Sun and H. Chen, "A Survey of k nearest neighbor algorithms for solving the class imbalanced problem," *Wireless Communications and Mobile Computing*, vol. 2021, no. 1, Jan. 2021, doi: 10.1155/2021/5520990.
[35] W. N. L. W. H. Ibeni, M. Z. M. Salikon, A. Mustapha, S. A. Daud, and M. N. M. Salleh, "Comparative analysis on bayesian classification for breast cancer problem," *Bulletin of Electrical Engineering and Informatics*, vol. 8, no. 4, Dec. 2019, doi: 10.11591/eei.v8i4.1628.
[36] C. O. Truică and C. A. Leordeanu, "Classification of an imbalanced data set using decision tree algorithms," *UPB Scientific Bulletin, Series C: Electrical Engineering and Computer Science*, vol. 79, no. 4, pp. 69–84, 2017.
[37] S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng, and P. J. Kennedy, "Training deep neural networks on imbalanced data sets," in *Proceedings of the International Joint Conference on Neural Networks*, vol. 2016, pp. 4368–4374, Jul. 2016, doi: 10.1109/IJCNN.2016.7727770.
[38] H. Shah and V. Manjula, "A performance analysis of deep convolutional neural networks using kuzushiji character recognition," in *2020 International Conference on Decision Aid Sciences and Application*, pp. 1068–1071, Nov. 2020, doi: 10.1109/DASA51403.2020.9317188.
[39] H. Luo, X. Pan, Q. Wang, S. Ye, and Y. Qian, "Logistic regression and random forest for effective imbalanced classification," in *International Computer Software and Applications Conference*, vol. 1, pp. 916–917, Jul. 2019, doi: 10.1109/COMPSAC.2019.00139.

BIOGRAPHIES OF AUTHORS

Muhlasah Novitasari Mara holds both bachelor and Master of Mathematics degree from Universitas Gadjah Mada, Indonesia. She is currently a lecturer at Department of Telecommunication Engineering Science in Politeknik Negeri Semarang, Indonesia. Her research prowess extends to machine learning, data mining, and time series forecasting. She can be contacted at email: muhlasah@polines.ac.id.

Sidiq Syamsul Hidayat notable contributions in the development of a robot for primary school learning, the design of high-frequency transmission lines using microstrip technology, and the creation of a smart E-Health application for monitoring patients' daily activities. His bachelor's in electrical engineering from Universitas Diponegoro, Indonesia, master's in electrical engineering from Institut Teknologi Bandung, Indonesia, and Doctor in Intelligent Interaction Technology from the University of Tsukuba, Japan. He can be contacted at email: sidiqsh@polines.ac.id.

Farika Tono Putri is currently based at Politeknik Negeri Semarang, Indonesia. His academic journey commenced at Universitas Diponegoro, Semarang, where she earned his bachelor's and master degree in mechanical engineering. She can be contacted at email: farika.tonoputri@polines.ac.id.

Dwi Rahmawati is a respected academic and lecturer in agribusiness. She earned a bachelor's degree in Seed Technology from Institut Pertanian Bogor, Indonesia, researching upland rice genotypes from East Kalimantan, and a master's degree in Agribusiness from Universitas Jember, studying the impact of government policies on the cocoa industry. She can be contacted at email: rahmawati@polije.ac.id.

Sri Wahyuni has dedicated her career to the advancement of agricultural sciences. Currently affiliated with the Center for Rice Plant Research, she serves as a beacon of knowledge and innovation in the field. Her academic journey commenced at the renowned Institut Pertanian Bogor, Indonesia, where she pursued her bachelor's degree in Agricultural Cultivation. Subsequently, she continued her educational pursuit by earning a master's degree in Seed Technology from the University Putra Malaysia. She can be contacted at email: sri.wahyuni.bbpadi@gmail.com.

Muhamad Cahyo Ardi Prabowo currently serving at Politeknik Negeri Semarang. He earned his bachelor's degree in electrical engineering from Universitas Tidar, Indonesia and master's degree in Telecommunication Engineering at Politeknik Negeri Semarang, Indonesia. He actively contributes in Digital Engineering and Artificial Intelligence subject. He can be contacted at email: m.cahyoardi.p@polines.ac.id.

Noer Ni'mat Syamsul Kabir earned her bachelor's degree in Statistics from Universitas Hasanuddin, Indonesia. She actively participating in various scholarship programs that showcase her prowess in data science and artificial intelligence. Notable among these are the Digital Talent Scholarship 2021: PROA Data Analyst from DQLAB and the Ministry of Communication and Information of the Republic of Indonesia in 2021. She can be contacted at email: nikmatnoer@gmail.com.

Ragil Tri Indrawati is a proficient professional in the field of naval and mechanical engineering. Her academic journey commenced at Universitas Indonesia, where she pursued her bachelor's degree in Shipping Techniques. She continued her education at Universitas Indonesia, completing her master's degree in mechanical engineering. She can be contacted at email: ragil.tri@polines.ac.id.