Ensemble-Based Machine Learning with PSO and GA for Ship Coating Detection Using Portable Vis/NIR

ALI KHUMAIDI[1], RIDWAN RAAFI'UDIN[2], NUSA SETIANI TRIASTUTI[3], LINDA MARLINDA[1]

[1]Informatika, Universitas Nusa Mandiri, 17411 Jakarta, Indonesia
[2]Computer Science, Universitas Pembangunan Nasional Veteran, 12450 Jakarta, Indonesia
[3]Faculty of Engineering, Krisnadwipayana University, 13770 Jakarta, Indonesia

Corresponding author: Ridwan Raafi'udin (e-mail: raafiudin@upnvj.ac.id).

ABSTRACT This study presents a method for detecting ship coating quality using a portable Vis/NIR spectroscopy system combined with machine learning. To improve accuracy, we integrated spectral transformation (Nippy), feature selection methods (PSO and GA), and ensemble learning models. The experiments involved four coating quality levels, producing 148 spectral samples. Results show that the proposed approach consistently outperforms single baseline models and traditional feature selection methods such as PCA and IFS. The best performance was achieved by combining Nippy with PSO, where the LDA algorithm reached 99.33% accuracy, while GA also showed strong results with both single and ensemble models. We also examined ensemble results at different stages of preprocessing and feature selection, showing that the ensemble maintained stable performance throughout the process. These findings demonstrate that the integration of spectral transformation and metaheuristic feature selection can enhance model robustness, providing more reliable and accurate coating quality detection for maritime applications.

KEYWORDS Classification; Ensemble model; Feature Selection; Non-destructive; Ship Coating; Vis/NIR spectroscopy.

I. INTRODUCTION

Ship coatings are essential in the maritime industry for protecting ships from corrosion, abrasion, and biofouling caused by marine organisms [1]. High-quality coatings improve operational efficiency, extend ship service life, and reduce maintenance costs [2]. Conversely, corrosion, structural damage, and biofouling increase operational costs and safety risks when coating maintenance is inadequate [3]. International organizations such as the International Maritime Organization (IMO) and the Indonesian Classification Bureau (BKI) have established standards that require regular inspections as part of ship maintenance and certification procedures to mitigate these risks [4, 5].

The methods currently used for detecting coating conditions have limitations that reduce their efficiency and accuracy. Visual assessment, thickness measurement, and adhesion testing can identify visible defects such as rust, cracks, and peeling. However, these methods are subjective, depend on worker expertise, and are unable to detect micro or subsurface damage [6, 7]. Thickness and adhesion testing also suffer from measurement constraints, while visual inspection remains vulnerable to human error [8]. In contrast, the use of visible/near-infrared (Vis/NIR) spectroscopy as a non-destructive method has shown rapid performance and strong predictive accuracy in various fields [9–11], creating an opportunity for its application in the maritime industry. Vis/NIR-based detection enables the collection of chemical and physical information related to coating condition and quality indicators [12–14]. Nevertheless, its implementation faces challenges such as noise, scattering effects, and spectral shifts caused by environmental factors, sample characteristics, or device operation [15]. To address these issues, spectral transformation is required prior to modeling to improve machine learning performance [16, 17].

Efforts to improve the predictive performance of machine learning with Vis/NIR data have typically involved spectral transformation and feature selection [18]. The Nippy module, for example, enhances spectral data and has improved model performance by up to 12–15% in previous studies [19, 20]. Feature selection using techniques such as Principal Component Analysis (PCA) and Iterative Feature Selection (IFS) can also reduce data dimensionality and retain relevant

information [21, 22]. However, PCA and IFS are limited when dealing with high-dimensional Vis/NIR data, particularly in exploring complex feature spaces and achieving optimal feature subsets.

To overcome these limitations, metaheuristic-based feature selection methods such as Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) have been developed. These approaches are capable of searching broader solution spaces and identifying feature subsets that maximize model performance and accuracy [23, 24]. Prior research has demonstrated that combining spectral transformation with feature selection can enhance single-model performance [25]. However, single models often remain sensitive to noise and data imbalance. Ensemble methods, by integrating multiple base learners, can further improve stability and generalization. Studies on ensemble learning for classification tasks, such as wheat quality detection, have shown higher accuracy than single models [26]. By combining multiple algorithms, ensemble approaches reduce bias, variance, and the risk of overfitting [27, 28].

This study extends earlier work that applied spectral transformation and traditional feature selection (PCA and IFS) [25] by integrating spectral transformation with metaheuristic-based feature selection (PSO and GA) and ensemble machine learning. The proposed framework aims to overcome the limitations of conventional approaches by capturing nonlinear relationships more effectively and improving the robustness of classification results. The contributions of this study are threefold: (1) developing an integrated framework combining spectral transformation, metaheuristic feature selection, and ensemble learning for ship coating quality detection; (2) demonstrating improved accuracy compared to PCA- and IFS-based methods; and (3) highlighting the potential of portable Vis/NIR spectroscopy enhanced with advanced machine learning as a practical, reliable tool for maritime coating inspection.

II. MATERIALS AND METHOD

Fig. 1 explains the research stage, which begins with the determination of the ground truth of the classification level of ship coating quality, which is based on the guidelines of the condition assessment program (CAP) from PT Biro Klasifikasi Indonesia (BKI). We have grouped the level of coating quality into four categories: (1) Very Good Condition (VG), (2) Good Condition (G), (3) Class Condition (C), and (4) Poor Condition (P) [25, 29]. We carried out the development of portable VIS/NIR devices before the Vis/NIR data acquisition stage. We then analyzed and preprocessed the spectral data from the acquisition, utilizing the Nippy module for spectral transformation and a metaheuristic approach for feature selection. The next stage is the development of a ship coating classification model using the ensemble method. We evaluate the model using performance metrics to guarantee its accuracy. In addition, we designed the workflow so that results can be observed at intermediate stages (single model, feature selection only, and full ensemble), allowing us to compare how accuracy evolves across different steps of the pipeline.

A. SAMPLES

Samples were collected through spectral data in the field in partnership with PT Dok dan Perkapalan Kodja Bahari

Galangan Jakarta II. The samples used in the study were ships and the type of coating that is usually operated, namely Roll-on/Roll-off (RoRo) ships and ferries with coatings made of polyurethane and epoxy. Experts, numbering up to 40, determine the samples for each coating quality level (VG, G, C, P). Experts have marked a total of 160 sample points for Vis/NIR data collection using a developed portable device.
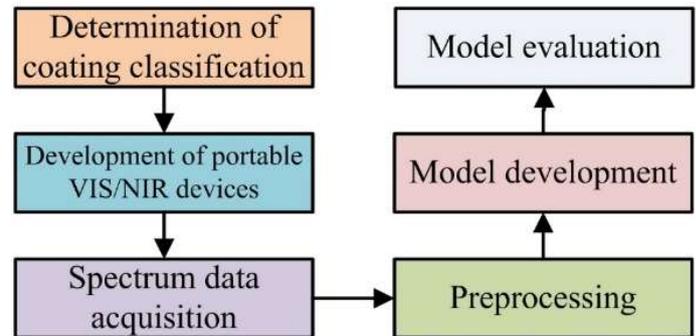

Figure 1. Research stages.

B. PORTABLE VIS/NIR DEVICES

Fig. 2 displays the results of the developed portable Vis/NIR device. The device consists of a Sparkfun AS7265x spectral sensor with a spectral range of 410–940 nm to assess the chemical and physical properties of the ship's coating [30]. The Raspberry Pi Zero W microcontroller integrates sensors via the I2C communication protocol [31]. This portable device draws its power from a battery power bank. Python-based software manages the sensor components, facilitates communication, and automates the data acquisition process.

Figure 2. Portable Vis/NIR devices and data acquisition.

C. DATA ACQUISITION

The data acquisition process on each sample was carried out three times, and the average value of the measurement was used for subsequent analysis. Data acquisition was carried out paying attention to environmental conditions with temperatures in the range of 25–30°C and humidity between 60–75%. Lighting is also taken into account when acquiring data, ensuring the sampling point operates under controlled lighting conditions. During data acquisition, the portable Vis/NIR device is placed perpendicular to the sample surface (see Fig. 2) with the aim of eliminating angular reflectance variations. The conditioning of external factors and the operation of the device is intended to avoid bias due to inconsistencies.

D. THE PROPOSED METHODS

Fig. 3 illustrates the proposed method. The proposed method builds on the previous approach [25] by expanding the feature

selection of metaheuristic approaches and machine learning ensemble models. Spectral transformations using nippy modules apply iterations to obtain the best combination of methods, operators, and values against spectral data. Feature selection uses population-based optimization algorithms, namely PSO and GA. This algorithm allows the identification of the most relevant wavelengths in high-dimensional data. PSO is able to converge data quickly and efficiently in computing, while GA is able to optimize global search so that it is effective in finding the optimal subset of features. Compared to conventional techniques such as PCA and IFS, this approach provides greater flexibility in exploring complex feature spaces, leading to more stable and accurate models when applied to Vis/NIR data.

Support Vector Classifier (SVC), Random Forest (RF), Multilayer Perceptron (MLP), Extreme Gradient Boosting (XGB), and Linear Discriminant Analysis (LDA) are some of the machine learning algorithms that are part of the ensemble model framework. The stacking method, using logistic regression (LR) as a meta-classifier and voting method, forms the ensemble model [32]. The proposed method is evaluated using accuracy metrics and K-fold cross-validation [33]. To ensure transparency, we not only report the final ensemble results but also present intermediate evaluations (single models and feature selection outputs) to show the contribution of each step in improving overall accuracy.
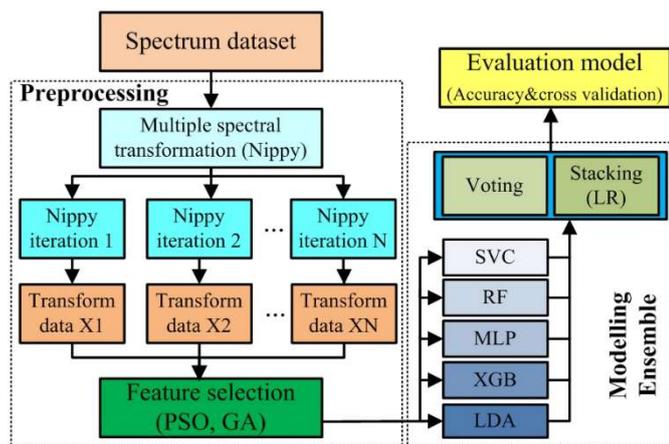

Figure 3. Proposed methods.

1. SPECTRAL TRANSFORMATION

Spectral transformation is an effective step in spectral data processing. The application of methods and operators in spectral transformation is carried out by trial and error because there is no standard guideline for achieving the best performance [34, 35]. Nippy modules with the use of machine learning have the ability to iteratively find the best methods and operators. This study applies the transformation methods of smoothing and derivative, baseline correction, normalization, and scatter correction. The smoothing and derivative technique effectively eliminates random noise and preserves key spectral features [36], the baseline correction method combats light scattering variation [37], the normalization method upholds amplitude consistency [38], and the scatter correction method reduces systematic shifts in the spectral baseline [39]. Table 1 shows the methods, operations, parameters, and values of spectral transformation used in this study.

Table 1. Methods, operations, parameters, and values of spectral transformation

Method	Operator	Parameter	Values
Smoothing and Derivative	SAVGOL	filter_win poly_order deriv_order	5, 7, 11 3 1, 2
Baseline Correction	BASELINE		
Normalization	NORML		
Scatter Correction	SNV		
	RNV	iqr	75–25, 90–10
	MSC		

2. FEATURE SELECTION

In several studies related to spectral Vis/NIR data, feature selection is a fairly important stage, it is proven that the application of PCA and IFS is able to improve model performance [25]. The selection of features with the metaheuristic approach used in this study is PSO and GA. These two population-based optimization methods have quite good abilities in exploring complex search spaces on spectral features. The concept of the PSO algorithm mimics the collective behavior of animals such as birds or fish. Potential solutions in a subset of spectral features are represented by each particle in the swarm, and the quality of these solutions is assessed using objective functions such as model accuracy [40]. The particle uses the best position of the individual (pbest) and global (gbest) of the swarm to adjust its position during repeated iterations. This iterative mechanism allows PSOs to coalesce quickly towards the best solution, which makes them more computationally efficient than other evolutionary approaches.

In contrast, GA mimics the natural selection process, in which populations and individuals constitute a randomly generated subset of characteristics. Genetic operations such as crossing, mutation, and selection drive evolution. These operations repeatedly refine the population, guaranteeing that the characteristic group that has the highest level of performance remains until the next generation [41]. Although GA is larger in terms of computing needs than PSO, the ability to explore the global search space is quite effective in finding the best combination of features.

The comparative advantages of PSO and GA have been emphasized in several studies. Faster convergence capabilities and lower computational costs, PSO is usually preferred [42, 43]. However, despite requiring higher compute costs, GA has shown outstanding performance on high-complexity data [44, 45]. The focus of this research is on two main criteria: the accuracy of the generated prediction model and the stability of the selected features across different paths. This comparative design allows us to clearly observe which algorithm contributes most to performance gains at both intermediate and final ensemble stages.

3. MODELLING

This study employs an ensemble model for machine learning. The ensemble approach is able to reduce bias and variance and improve model generalization. Voting and stacking techniques form the ensemble model. The voting approach combines individual model predictions with majority votes, while stacking uses RL as a meta-classifier to combine the predictions of the base model. The integration of this model reduces bias and variance while increasing model generalization.

This study will compare the ensemble and single models, namely the SVC, RF, MLP, XGB, and LDA algorithms. The SVC algorithm is able to optimally separate classes, with kernel tricks capable of mapping input data to higher dimensional spaces. The RF algorithm is chosen because of its robustness in capturing non-linear interactions between features and its ability to reduce overfitting. MLP is capable of modeling complex non-linear relationships in data using a flexible neural network architecture. XGB can minimize errors through iterations. LDA is able to reduce dimensions and maximize class separation in data sets. By comparing these single algorithms against ensemble strategies, we can provide step-by-step evidence of how ensemble integration yields higher and more consistent performance.

4. MODEL EVALUATION

Model evaluation is carried out to measure performance and reliability. The evaluation was carried out using the K-fold cross-validation technique with a value of K = 5 to ensure that the model has good generalization ability and is not very consistent with the training data. The cross-validation technique is able to make the model performance more accurate and stable because it uses all data as training and test data [46]. Accuracy metrics, which calculate the percentage of correct predictions against the entire test data, evaluate model performance comparisons. Cross-validation with K-fold ensures that the model's performance evaluation is free of bias. In addition to final accuracy, we also tracked performance metrics during feature selection and intermediate ensemble stages, providing a more detailed picture of model development.

III. RESULTS

The number of samples marked was initially 40 per coating quality level but due to an error in the spectral reading of the device, the number of spectral data acquired used in data processing was 148. The data is divided into four coating quality levels: 40 (VG), 33 (G), 37 (C), and 38 (P). The results of the visualization of spectral data before preprocessing can be seen in Fig. 4.
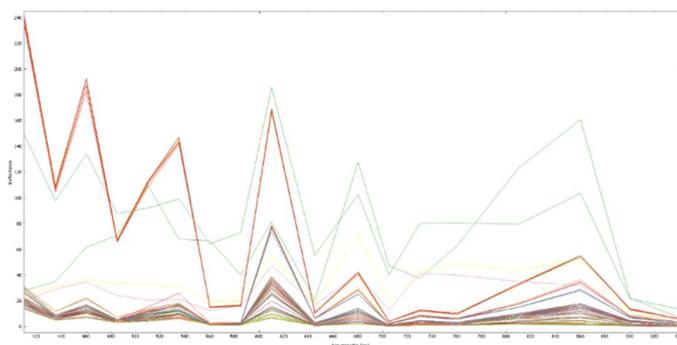

Figure 4. Spectral data acquisition results.

The results of data processing by applying spectral transform (Nippy), feature selection (PSO and GA), and a combination of both on single and ensemble models can be seen in Table 2. To provide transparency, we report the performance at each stage (no treatment, Nippy only, PSO only, GA only, and combined approaches), both for single classifiers and ensemble methods. This allows us to examine how accuracy evolves step by step before reaching the final model.

Table 2. Comparison of modeling results accuracy

Methods	SVC	RF	LDA	MLP	XGB	Ensemble	
						Stacking	Voting
None	97.29	97.29	97.98	96.62	97.29	97.98	97.29
Nippy	94.55	93.24	97.33	92.51	94.55	98	96.62
PSO	91.10	91.79	83.66	91.13	84.44	85.75	86.41
GA	98.67	97.95	97.98	98.64	98.64	98.64	98.64
Nippy-PSO	99.31	98.64	99.33	97.98	98.67	97.31	97.31
Nippy-GA	98.67	98.64	94.55	97.31	94.55	97.31	95.93

In the non-treatment scenario on a single model, the LDA algorithm achieves the highest accuracy of 97.98%, outperforming other algorithms, namely SVC, RF, and XGB, by 97.29%, and the MLP algorithm has a slightly lower accuracy of 96.62%. The ensemble model approach with stacking showed the same performance as LDA (97.98%) and with 97.29% voting. This indicates that even without preprocessing, the ensemble maintains stability comparable to the strongest single classifier.

The spectral transformation effect on a single model does not provide an increase in accuracy compared to no spectral treatment; the LDA algorithm still has the highest accuracy compared to other algorithms, with an accuracy of 97.33%. In the ensemble model with stacking, spectral transformation increases accuracy by 0.02% compared to no treatment's accuracy of 98%. This modest improvement suggests that transformation alone is insufficient, but it prepares the data for subsequent optimization.

The implementation of feature selection using GA yields superior performance. In both single and ensemble models, GA successfully enhances the accuracy of all algorithms when compared to no treatment or spectral transformation. The highest accuracy of the effect of applying GA to the SVC algorithm was 98.64%, and the ensemble model with voting and stacking had the same accuracy. These results confirm that GA consistently strengthens model performance across classifiers, demonstrating its robustness for complex Vis/NIR data.

The effect of feature selection using PSO shows poor performance, which proves that the ensemble and single models do not provide better accuracy than models without treatment (none). In the single model, the highest accuracy in the RF algorithm is 91.79%, and the ensemble model with voting has an accuracy of 86.41%. This suggests that PSO, although computationally efficient, may converge prematurely in this dataset, leading to weaker subsets of features.

The combination of spectral transform and feature selection yields the best results. Applying spectral transform and feature selection with PSO yields the best results. The Nippy-PSO combination is able to improve performance on all single and ensemble models. In a single model, the LDA algorithm has the highest accuracy of 99.33%, and the SCV algorithm has the highest accuracy increase of 2.02%. This demonstrates the synergy between preprocessing and metaheuristic optimization, especially when paired with classifiers sensitive to feature quality such as LDA and SVC. The effect of Nippy-PSO on the ensemble model, although it improves accuracy, is not as large as in the case of a single model.

In contrast, the Nippy-GA combination is able to improve accuracy performance compared to no treatment (None) only on SVC and RF algorithms with the highest accuracy of

98.67% on a single model. The accuracy in the ensemble model does not outperform that of the single model without treatment (None). Therefore, while the feature selection effect of Nippy-GA is quite effective for a single model, it is less effective in the ensemble model. Overall, the results highlight that different combinations bring different strengths: GA ensures broad stability, while PSO coupled with Nippy provides the highest peak accuracy. Presenting these intermediate outcomes gives a clearer picture of how each step contributes to the final performance.

IV. DISCUSSION

This study presents the findings from testing Vis/NIR data for ship coating quality detection, demonstrating how spectral transformation and feature selection can enhance the accuracy of the model when compared to no spectral treatment, both in single and ensemble model approaches. Fig. 5 illustrates that the performance of GA, Nippy-PSO, and Nippy-GA deployments is superior across all models. In a single model, the SCV algorithm has the highest accuracy of 98.67% when integrated with GA and Nippy-GA; the RF algorithm has the highest performance of 98.64% with Nippy-GA and Nippy-PSO; the LDA algorithm achieves its peak accuracy of 99.33% using the Nippy-PSO approach; the MLP algorithm achieves the highest accuracy of 98.64% with GA; and the XGB algorithm has the highest accuracy of 98.67% with Nippy-GA. The ensemble model has the highest accuracy of 98.64% with GA feature selection. These results confirm that both metaheuristic feature selection and preprocessing contribute significantly to performance improvement compared to untreated data.

The method designed in this study has proven its superior performance in both ensemble and single models compared to previous studies that integrate spectral transformation and PCA and IFS feature selection with the highest accuracy of only 97.33% [25]. This highlights the added value of metaheuristic-based feature selection, which is better suited for high-dimensional and nonlinear spectral data compared to linear reduction methods like PCA or stepwise elimination approaches such as IFS.
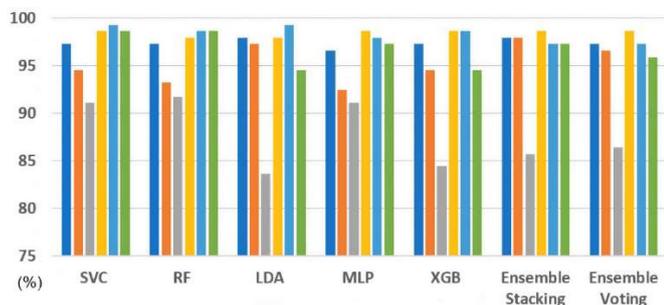

Figure 5. Comparison chart of modeling results accuracy.

The application of GA to Vis/NIR data has the advantage of accuracy in single models, namely SVC and MLP, as well as ensemble models, due to its ability in the selection of relevant features supported by strengths in the evolutionary base, both crossover and mutation, able to explore the solution space. This aligns with previous findings where GA was reported to be particularly effective in handling complex search spaces in spectroscopy-based studies [41, 44].

The application of Nippy and PSO also has the advantage of a single model on RF, LDA, and XGB algorithms, and even LDA has the highest accuracy compared to all algorithms with an accuracy of 99.33%. This indicates that swarm-based optimization benefits more from preprocessed data, where noise and scattering have already been minimized. Without transformation, PSO tends to converge prematurely, which explains its weaker performance in raw spectra. PSO is able to select more significant features from the data that has been transformed by Nippy with the power of PSO working with the principle of swarm intelligence based on sharing particle information in the solution space.

This is supported by the results of the standard deviation analysis (see Table 3), the implementation of GA feature selection and the integration of Nippy and PSO consistently show low standard deviations, namely 0.31 and 0.79. Lower deviation values indicate higher stability across folds, suggesting that the improvements are not only marginal but also robust and generalizable.

Table 3. Comparison of standard deviation methods

Methods	None	Nippy	PSO	GA	Nippy-PSO	Nippy-GA
Standard Deviation	0.44	1.94	3.22	0.31	0.79	1.61

From a practical standpoint, the proposed integration framework offers several advantages for the maritime industry. First, it allows coating quality to be assessed rapidly and non-destructively in the field, reducing reliance on subjective human inspection. Second, the combination of Vis/NIR spectroscopy with advanced feature selection ensures that micro-level coating degradation can be detected earlier, potentially lowering maintenance costs and extending ship service life. Finally, by demonstrating higher stability and accuracy compared to conventional approaches, the framework provides confidence for adoption in routine classification and certification processes, as recommended by IMO and BKI.

V. CONCLUSIONS

This research succeeded in developing a method for detecting the quality of ship coatings using portable Vis/NIR spectroscopy by integrating spectral transformation, feature selection, and ensemble models. The proposed framework demonstrates that combining preprocessing (Nippy), metaheuristic feature selection, and ensemble learning provides higher accuracy and stability compared to conventional approaches such as PCA and IFS. Among the feature selection methods, GA shows superior performance with the highest accuracy on single models with SVC and MLP algorithms as well as ensemble models. Although PSO alone produced weaker results, its integration with Nippy yielded significant improvements, with the Nippy-PSO–LDA combination achieving the overall best performance (99.33%). This confirms that preprocessing and metaheuristic-based feature selection complement each other in handling high-dimensional spectral data.

The integration of these methods reduces bias, variance, and subjectivity compared to traditional coating inspection practices, making the detection process more accurate, reliable, and suitable for field implementation. The proposed approach therefore offers both methodological innovation and practical relevance for coating quality assessment in the maritime

industry. Future research can explore hybrid metaheuristic strategies, deeper neural architectures, or real-time deployment on embedded systems to further enhance performance and computational efficiency. The findings of this study also open opportunities for broader applications of Vis/NIR spectroscopy, not only for ship coatings but also for other maritime materials that require non-destructive quality assessment.

References

[1] H. S. Sultan Aljibori, A. A. H. Al-Amiery, and A. A. H. Kadhum, "Advances in corrosion protection coatings: A comprehensive review," *Int. J. Corros. Scale Inhib.*, vol. 12, no. 4, 2023, https://doi.org/10.17675/2305-6894-2023-12-4-6.

[2] W. Pei, X. Pei, Z. Xie, and J. Wang, "Research progress of marine anti-corrosion and wear-resistant coating," *Tribol. Int.*, vol. 198, p. 109864, 2024, https://doi.org/10.1016/j.triboint.2024.109864.

[3] L. Tremps, B. Yeter, and A. Kolios, "Review and analysis of the failure risk mitigation via monitoring for monopile offshore wind structures," *Energy Reports*, vol. 11, pp. 5407-5420, 2024, https://doi.org/10.1016/j.egyr.2024.05.026.

[4] Y. A. Ahmed, G. Theotokatos, I. Maslov, L. A. L. Wennersberg, and D. A. Nesheim, "Regulatory and legal frameworks recommendations for short sea shipping maritime autonomous surface ships," *Mar. Policy*, vol. 166, p. 106226, 2024, https://doi.org/10.1016/j.marpol.2024.106226.

[5] A. W. Momber, D. Langenkämper, T. Möller, and T. W. Nattkemper, "The exploration and annotation of large amounts of visual inspection data for protective coating systems on stationary marine steel structures," *Ocean Eng.*, vol. 278, p. 114337, 2023, https://doi.org/10.1016/j.oceaneng.2023.114337.

[6] Z. Zhen-yu, Z. Qiu-yang, D. Cong, Y. Ju-yu, P. Guang-jian, and P. Zhong-yu, "Research on the promotion mechanism of surface burnishing process by two-dimensional ultrasonic vibration," *J. Mater. Res. Technol.*, vol. 13, pp. 1068-1082, 2021, https://doi.org/10.1016/j.jmrt.2021.05.038.

[7] I. Stojanović, A. Farkas, V. Alar, and N. Degiuli, "Evaluation of the corrosion protection of two underwater coating systems in a simulated marine environment," JOM, vol. 71, no. 12, pp. 4330-4338, 2019, https://doi.org/10.1007/s11837-019-03669-4.

[8] N. Hütten, M. Alves Gomes, F. Hölken, K. Andricevic, R. Meyes, and T. Meisen, "Deep learning for automated visual inspection in manufacturing and maintenance: A survey of open- access papers," *Appl. Syst. Innov.*, vol. 7, no. 1, p. 11, 2024, https://doi.org/10.3390/asi7010011.

[9] Z. Li, J. Wang, D. Xiao, Z. Gu, and H. Xie, "Iron ore rock classification and mine remote sensing inversion based on spectroscopy and improved extreme learning machine," *Infrared Phys. Technol.*, vol. 140, p. 105400, 2024, https://doi.org/10.1016/j.infrared.2024.105400.

[10] R. Visnupriyan, B. M. Flanagan, K. J. Harper, and D. Cozzolino, "Near infrared spectroscopy combined with chemometrics as tool to monitor starch hydrolysis," *Carbohydr. Polym.*, vol. 324, p. 121469, 2024, https://doi.org/10.1016/j.carbpol.2023.121469.

[11] M. Ceroni, F. S. Gobber, and M. Actis Grande, "Ultraviolet-Visible-Near InfraRed spectroscopy for assessing metal powder cross-contamination: A multivariate approach for a quantitative analysis," *Mater.*, vol. 242, p. 113023, 2024, https://doi.org/10.1016/j.matdes.2024.113023.

[12] S. Kishigami, Y. Mizokami, T. Sakata, Y. Ogawa, D. Shiozawa, and T. Sakagami, "Detection of heavy-duty anticorrosion coating deterioration using near-infrared spectral characteristics," *NDT E Int.*, vol. 132, p. 102708, 2022, https://doi.org/10.1016/j.ndteint.2022.102708.

[13] C. R. Petersen et al., "Non-destructive subsurface inspection of marine and protective coatings using near- and mid-infrared optical coherence tomography," *Coatings*, vol. 11, no. 8, p. 877, 2021, https://doi.org/10.3390/coatings11080877.

[14] E. Konadu-Yiadom, E. Bontrager, and A. Staerz, "The usefulness of infrared spectroscopy for elucidating the degradation mechanism of metal industrial heritage coatings," *Surfaces*, vol. 7, no. 4, pp. 846-863, 2024, https://doi.org/10.3390/surfaces7040056.

[15] A. Khumaidi, Y. A. Purwanto, H. Sukoco, and S. H. Wijaya, "Using fuzzy logic to increase accuracy in mango maturity index classification: Approach for developing a portable near-infrared spectroscopy device," *Sensors*, vol. 22, no. 24, p. 9704, 2022, https://doi.org/10.3390/s22249704.

[16] P. Koziol, K. Kosowska, P. Korecki, and T. P. Wrobel, "Scattering correction for samples with cylindrical domains measured with polarized infrared spectroscopy," *Anal. Chim. Acta*, vol. 1278, p. 341722, 2023, https://doi.org/10.1016/j.aca.2023.341722.

[17] S. Chakraborty, I. Patra, A. Pal, K. Mondal, and M. Pradhan, "Development of a semi-supervised machine learning based noise filter for quantum cascade laser-coupled mid-infrared spectrometer," *Infrared Phys. Technol.*, vol. 141, p. 105452, 2024, https://doi.org/10.1016/j.infrared.2024.105452.

[18] W. Zhang, L. C. Kasun, Q. J. Wang, Y. Zheng, and Z. Lin, "A review of machine learning for near-infrared spectroscopy," *Sensors*, vol. 22, no. 24, p. 9764, 2022, https://doi.org/10.3390/s22249764.

[19] J. Torniainen, I. O. Afara, M. Prakash, J. K. Sarin, L. Stenroth, and J. Töyräs, "Open-source python module for automated preprocessing of near infrared spectroscopic data," *Anal. Chim. Acta*, vol. 1108, pp. 1-9, 2020, https://doi.org/10.1016/j.aca.2020.02.030.

[20] A. Khumaidi and R. Raafi'udin, "Effects of oversampling smote and spectral transformations in the classification of mango cultivars using near-infrared spectroscopy," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 12, no. 3, p. 1047, 2022, https://doi.org/10.18517/ijaseit.12.3.16001.

[21] S. Ren and Y. Jia, "Near-Infrared data classification at phone terminal based on the combination of PCA and CS-RBFSVC algorithms," *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 287, p. 122080, 2023, https://doi.org/10.1016/j.saa.2022.122080.

[22] R. Raafi'udin, Y. A. Purwanto, I. S. Sitanggang, and D. A. Astuti, "Feature selection model development on near-infrared spectroscopy data case study of beef freshness quality prediction," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 1, pp. 645-653, 2024. https://doi.org/10.14569/IJACSA.2024.0150163.

[23] S. Zhang, Y. Yin, C. Liu, J. Li, X. Sun, and J. Wu, "Discrimination of wheat flour grade based on PSO-SVM of hyperspectral technique," *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 302, p. 123050, 2023, https://doi.org/10.1016/j.saa.2023.123050.

[24] Y. Qin, K. Song, N. Zhang, M. Wang, M. Zhang, and B. Peng, "Robust NIR quantitative model using MIC-SPA variable selection and GA-ELM," *Infrared Phys. Technol.*, vol. 128, p. 104534, 2023, https://doi.org/10.1016/j.infrared.2022.104534.

[25] A. Khumaidi, R. Raafi'udin, and N. S. Triastuti, "Enhancing ship coating quality detection via machine learning-optimized visible near-infrared spectroscopy," *Instrum. Mes. Métrologie*, vol. 23, no. 6, pp. 441-450, 2024, https://doi.org/10.18280/i2m.230604.

[26] H. Jiang, S. Zhang, Z. Yang, L. Zhao, Y. Zhou, and D. Zhou, "Quality classification of stored wheat based on evidence reasoning rule and stacking ensemble learning," *Comput. Electron. Agric.*, vol. 214, p. 108339, 2023, https://doi.org/10.1016/j.compag.2023.108339.

[27] A. K. Biswas, R. Seethalakshmi, P. Mariappan, and D. Bhattacharjee, "An ensemble learning model for predicting the intention to quit among employees using classification algorithms," *Decis. Anal. J.*, vol. 9, p. 100335, 2023, https://doi.org/10.1016/j.dajour.2023.100335.

[28] D. H. Parekh and V. Dahiya, "Early detection of breast cancer using machine learning and ensemble techniques," *Int. J. Comput.*, pp. 231-237, 2023, https://doi.org/10.47839/ijc.22.2.3093.

[29] B. K. Indonesia., "Guidelines for condition assessment program," vol. 11 Guidel. Cond. Assess. Progr. 2015 Ed., 2015.

[30] SparkFun, "AS7265x Datasheet," *AMS Datasheet*, vol. 1, pp. 1-63, 2018.

[31] J. Mankar, C. Darode, K. Trivedi, M. Kanoje, and P. Shahare, "Review of I2C protocol," *Int. J. Res. Advent Technol.*, vol. 2, no. 1, pp. 2321-9637, 2014.

[32] C. Rajathi and P. Rukmani, "Hybrid learning model for intrusion detection system: A combination of parametric and non-parametric classifiers," *Alexandria Eng. J.*, vol. 112, pp. 384-396, 2025, https://doi.org/10.1016/j.aej.2024.10.101.

[33] M. T R, V. K. V, D. K. V, O. Geman, M. Margala, and M. Guduri, "The stratified K-folds cross-validation and class-balancing methods with high-performance ensemble classifiers for breast cancer classification," *Healthc. Anal.*, vol. 4, p. 100247, 2023, https://doi.org/10.1016/j.health.2023.100247.

[34] L. A. C. Ribeiro, G. L. Menezes, T. Bresolin, S. I. A. Apelo, and J. R. R. Dórea, "Near-infrared spectroscopy analysis to predict urinary allantoin in dairy cows," *JDS Commun.*, vol. 6, pp. 212-216, 2024, https://doi.org/10.3168/jdsc.2024-0641.

[35] W. Yang, Y. Xiong, Z. Xu, L. Li, and Y. Du, "Piecewise preprocessing of near-infrared spectra for improving prediction ability of a PLS model," *Infrared Phys. Technol.*, vol. 126, p. 104359, 2022, https://doi.org/10.1016/j.infrared.2022.104359.

[36] J. Scheffel and K. Lindvall, "Temporal smoothing – A step forward for time-spectral methods," *Comput. Phys. Commun.*, vol. 270, p. 108173, 2022, https://doi.org/10.1016/j.cpc.2021.108173.

[37] X. Chen and D. Wang, "Baseline correction of near-fault ground motion records based on the Hilbert spectral analysis," *Soil Dyn. Earthq. Eng.*,

vol. 154, p. 107162, 2022, https://doi.org/10.1016/j.soildyn.2022.107162.

[38] Y. Wang and X.-D. Zhang, "On weighted spectral radius of unraveled balls and normalized Laplacian eigenvalues," *Discrete Math.*, vol. 345, no. 12, p. 113173, 2022, https://doi.org/10.1016/j.disc.2022.113173.

[39] N. McArdle, S. Cournane, J. McCavana, J. Lucey, and L. León Vintró, "Development of a scatter correction technique for planar 99mTc-MAA imaging to improve accuracy in lung shunt fraction estimation," *Phys. Medica*, vol. 99, pp. 94-101, 2022, https://doi.org/10.1016/j.ejmp.2022.05.017.

[40] M. Muljono and N. Adhipurna, "A weighted majority voting ensemble model for disease prediction boosted by PSO: The case of type 2 diabetes," *Int. J. Comput.*, vol. 22, issue 4, pp. 475-484, 2023, https://doi.org/10.47839/ijc.22.4.3354.

[41] C. Shambare, Y. Sun, and O. Imoru, "An improved QPSK based on a hybrid genetic algorithm for efficient OFDM transmission," *Int. J. Comput.*, vol. 23, issue 4, pp. 517-525, 2025, https://doi.org/10.47839/ijc.23.4.3751.

[42] N. Singh et al., "Integrating NIR spectroscopy with machine learning and heuristic algorithm-assisted wavelength selection algorithms for protein content prediction in rice bean (Vigna umbellata L.)," *Food Humanit.*, vol. 3, p. 100399, 2024, https://doi.org/10.1016/j.foohum.2024.100399.

[43] K. Zhang et al., "Optimization of visible transmission and NIR reflection in multilayer via PSO: Simulation and experimental validation," *Renew. Energy*, vol. 237, p. 121913, 2024, https://doi.org/10.1016/j.renene.2024.121913.

[44] B. Ferreiro, R. Leardi, E. Farinini, and J. M. Andrade, "Supervised classification combined with genetic algorithm variable selection for a fast identification of polymeric microdebris using infrared reflectance," *Mar. Pollut. Bull.*, vol. 195, p. 115540, 2023, https://doi.org/10.1016/j.marpolbul.2023.115540.

[45] X. Hao, Z. Chen, S. Yi, and J. Liu, "Application of improved Stacking ensemble learning in NIR spectral modeling of corn seed germination rate," *Chemom. Intell. Lab. Syst.*, vol. 243, p. 105020, 2023, ttps://doi.org/10.1016/j.chemolab.2023.105020.

[46] A. M. Peco Chacón, I. Segovia Ramírez, and F. P. García Márquez, "K-nearest neighbour and K-fold cross-validation used in wind turbines for false alarm detection," *Sustain. Futur.*, vol. 6, p. 100132, 2023, https://doi.org/10.1016/j.sftr.2023.100132.

DR. ALI KHUMAIDI, Doctor of Computer Science, Associate Professor of the doctoral program in Informatics, teaching disciplines: machine learning, artificial intelligence, data mining, computer vision. His research primary focuses on coding theory, internet of things (IoT), application of AI, application of machine learning on vis/NIR spectroscopy.

DR. RIDWAN RAAFI'UDIN, Doctor of Computer Science, lecturer in Informatics, teaching courses on algorithms, python programming, and machine learning. His expertise includes software development and artificial intelligence. His research focuses on IoT and AI, in processing Visible-Near Infrared (Vis-NIR) data for various applications.

DR. NUSA SETIANI TRIASTUTI, Doctor of civil engineering, lecturer specializing in structural engineering and ship construction. Her expertise includes foundation analysis, retrofitting techniques, and reinforced concrete applications. Her research focuses on alternative ship structures, reinforced concrete in fishing vessel construction, and vibration analysis in shipbuilding.

DR. LINDA MARLINDA, Doctor of Computer Science, Associate Professor of the doctoral program in Informatics, teaching disciplines: image processing, machine learning, artificial intelligence, data mining, computer vision. His research primary focuses on coding theory, internet of things (IoT), application of AI, and cultural heritage digitization.