*International Journal of Advanced Technology and Engineering Exploration, Vol 11(111)*
*ISSN (Print): 2394-5443 ISSN (Online): 2394-7454*
*http://dx.doi.org/10.19101/IJATEE.2023.10102060*

**Research Article**

# Multi-classifier models to improve the accuracy of fish landing application

**Rosaida Rosly[1*], Mustafa Man[2], Amir Ngah[2] and Nor Saidah Abd Manan[3]**
Lecturer, Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu (UMT), Malaysia[1]
Associate Professor, Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu (UMT), Malaysia[2]
Research Assistant, Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu (UMT), Malaysia[3]

Received: 03-August-2023; Revised: 10-February-2024; Accepted: 13-February-2024


## Abstract
***Despite the numerous fish classification systems developed over the years, they often suffer from poor prediction accuracy, necessitating further improvement. This study addresses this issue by comparing the performance of different classifiers on fish landing datasets (2005-2019) obtained from the Department of Fisheries Malaysia (DOFM). The focus is on the East Coast of Peninsular Malaysia. The classifiers evaluated include Sequential minimal optimization (SMO), naïve Bayes (NB), multi-layer perception (MLP), instance-based for k-nearest neighbor (IBK), and random forest (RF). The performance of each classifier is assessed using classification accuracy and confusion matrix metrics, employing a 10-fold cross-validation method. Additionally, a multi-classification technique is applied to enhance the accuracy of individual classifiers and determine the most effective approach for generating an accurate dataset. The study reveals that the combinations RF+SMO+NB+MLP and SMO+RF+NB+MLP outperform single classifiers and other fusion methods, achieving the highest accuracy at 80.95%. This indicates that a multi-classifier approach can significantly enhance the performance of individual classifiers. The findings highlight the effectiveness of the multi-classifier approach in improving prediction accuracy for fish classification. The identified combinations, RF+SMO+NB+MLP and SMO+RF+NB+MLP, demonstrate superior performance and can serve as a robust methodology for fish landing classification in the context of the East Coast of Peninsular Malaysia. Further research and implementation of such multi-classifier approaches could contribute to more accurate and reliable fish classification systems.***

## Keywords
***Fish landing dataset, Feature selection, Classification performance, Multi-classifier.***

## 1.Introduction
The ocean, serving as a vast reservoir of resources crucial for the economy and human sustenance, plays a pivotal role in influencing the economies of specific countries. This impact is particularly evident through the expansion of the fisheries sector and related marine industries [1]. To strategically develop and ensure the sustainable growth of these industries, the application of data mining, classification, and analyses becomes indispensable. Data mining, a set of techniques focused on extracting pertinent information from extensive databases across diverse business domains, stands as a key tool in informed decision-making [2]. However, the existing literature in this field faces challenges that warrant careful consideration.

The current state of knowledge lacks a comprehensive understanding of the complexities associated with fish landing classification, a critical aspect for effective resource management in the marine domain. These challenges underscore the necessity for further research and exploration in this domain, aiming to bridge the existing gaps in knowledge.

Motivated by the critical role of fisheries in economic development and sustainability, this paper seeks to address the aforementioned gaps by employing both single and multi-classifier algorithms to study fish landing classification. This research extends beyond conventional approaches by conducting a comparative analysis of the East Coast fish landings dataset using various classifiers and classifier fusions. The Waikato Environment for Knowledge Analysis (WEKA) tool is employed to assess the performance of five selected classifiers: Sequential minimal

*Author for correspondence

optimization (SMO), naïve Bayes (NB), multi-layer perception (MLP), instance-based for k-nearest neighbor (IBK), and random forest (RF). The objectives of this paper are to outline the specific contributions of our research, which involve the application of innovative single and multi-classifier algorithms to enhance the accuracy and efficiency of fish landing classification.

Challenges in the existing literature are rooted in the lack of a comprehensive understanding of fish landing classification complexities [3]. These complexities are crucial for developing effective resource management strategies. As the fisheries sector continues to be a cornerstone of economic growth, addressing these challenges becomes imperative for sustainable development [4]. The motivation for this work arises from the need to fill these knowledge gaps and contribute to the advancement of fish landing classification methodologies. By applying state-of-the-art algorithms and conducting a thorough comparative analysis, we aim to provide insights that can inform decision-making processes in the fisheries and marine industries. Objectives of the paper include to employ single and multi-classifier algorithms for fish landing classification, to conduct a comparative analysis of the East Coast fish landings dataset using various classifiers and classifier fusions and to assess the performance of selected classifiers using the WEKA tool. Contributions of this research involve the application of innovative algorithms to enhance the accuracy and efficiency of fish landing classification. This paper contributes to the existing body of knowledge by providing insights into the effectiveness of different classifiers and classifier fusions in the context of fish landing classification.

The organization of this paper is presented as follows: Section 2 reviews related literature and recent studies on classification algorithms and performance evaluation criteria. Section 3 outlines the proposed fish landing classification algorithm, including data acquisition, data pre-processing, and feature selection. Section 4 reports the results obtained from six experiments, while section 5 discusses the findings and presents the study's limitations. Finally, section 6 concludes the study and outlines avenues for future research.

## 2.Literature review

To manage the data, several data mining techniques can be used, one of which is the classification method. There are two steps to implement the classification function. In the first step, the classification model is built for describing a predetermined set of classes or concepts, while the second step consists of the model being used for classification.

### 2.1Single classifier

A single classification pertains to individual classifiers employed in the creation of a multi-classifier. Some of these classifiers, including support vector machines (SVM), k-nearest neighbors (k-NN), decision trees, RF, and NB classifiers, are commonly used as base classifiers [5]. In the realm of supervised learning, prevalent methods encompass artificial neural networks (ANN), SVM, and decision trees. Additional primary classification techniques comprise decision tree induction, Bayesian networks, k-NN classifiers, case-based reasoning, genetic algorithms, and fuzzy logic techniques [6]. It is noteworthy that a specific classifier may outperform others for a particular dataset, while a different classifier may excel for various datasets [7].

Within the scope of this research, only five classifiers have been selected: SMO, NB, MLP, IBK and RF. SMO is a potent method widely applied across diverse applications, capable of classifying both linear and non-linear data [8]. Additionally, NB operates as a supervised and probabilistic learning method, relying on conditional probability principles articulated by Bayes' theorem. MLPs are capable of learning and modelling complex non-linear relationships in data [9]. The network architecture, which includes multiple layers of nodes (neurons) and non-linear activation functions, enables MLPs to capture intricate patterns and representations in the data. IBK stands out as one of the simplest and earliest classification algorithms [10]. Finally, RF stemming from the decision tree, facilitates the aggregation of numerous weak or weakly-correlated classifiers into a robust classifier [11].

### 2.2Multi-classifier

The method of multi-classification combines the results of individual classification techniques, leading to enhanced performance compared to using a single classifier. This approach has rapidly gained popularity in the field of machine learning. Drawing on insights from prior studies, this method focuses on aggregating the outputs of various machine-learning-based classifiers (such as SVM, MLP, and decision tree) trained to predict the bug-proneness of software components [12]. In essence, the multi-classifier integrates predictions from diverse models to produce a final prediction, and its performance improves with the inclusion of more models. Importantly, the primary goal of the multi-classifier method is to enhance machine learning outcomes by combining multiple models. This technique results in superior predictive performance and increased accuracy compared to using a single model [13]. *Figure 1* illustrates the flow diagram of the multi-classifier system.

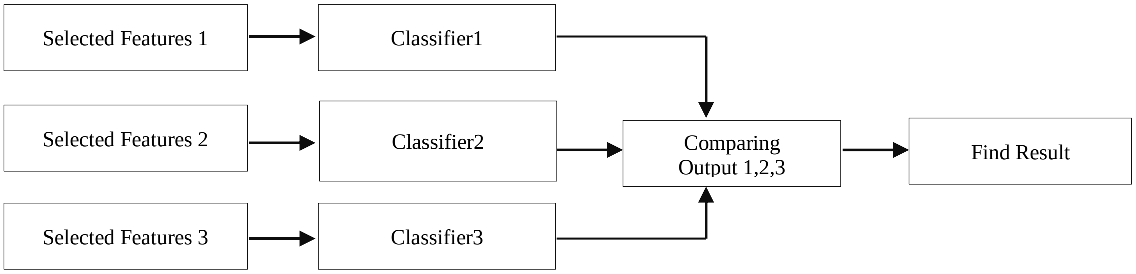


**Figure 1** Block diagram of multi-classifier system

## 2.3 Performance evaluation criteria

The number of correct/incorrect predictions in the confusion matrix was calculated using the evaluation method. *Table 1* shows the confusion matrix. The entries in the confusion matrix have the following meanings:

- A is the number of correct predictions for *Kebasi* (K) class,
- B, C are the numbers of incorrect predictions for *Kebasi* (K) class,
- E is the number of correct predictions for *Biji Nangka* (BN) class,
- D, F are the numbers of incorrect predictions for BN class,
- I is the number of correct predictions for *Siakap* (S), and
- G, H are the numbers of incorrect predictions for *Siakap* (S) class.

**Table 1** Confusion matrix

| | | **Predicted** | | |
|---|---|---|---|---|
| | | **K** | **S** | **BN** |
| Actual | K | A | B | C |
| | S | D | E | F |
| | BN | G | H | I |

The accuracy (Acc) is the measurement of the total number of accurate predictions. Subsequently, the accumulation of correct predictions was divided by the total dataset. It is determined using Equation 1 [14]:

$$\text{Acc} = \frac{A+E+I}{A+B+C+D+E+F+G+H+I} \qquad (1)$$

In addition, the model performance can be expressed using the error rate by dividing the total incorrect predictions by the total number of the dataset. It is given by the following Equation 2:

$$Error\ rate = \frac{Number\ of\ wrong\ predictions}{Total\ number\ of\ prediction} = \frac{B+C+D+F+G+H}{A+B+C+D+E+F+G+H+I} \qquad (2)$$

## 2.4 Related work

Numerous studies have delved into the realm of fish classification and prediction, employing diverse machine learning algorithms. *Table 2* reviews key works, highlighting the methods, results, advantages, and limitations. Moreover, the final review analysis reveals that besides the high accuracy demonstrated by SVM and SMO algorithms in fish species identification and water quality prediction, NB provides a straightforward probability-based classification. In addition, MLP exhibits efficiency in supervised training for rainbow trout. Furthermore, k-NN demonstrates flexibility, and RF successfully tracks fishing activity through decision tree aggregation. Ensemble methods, combining RF and Gradient Boosting, not only show improved accuracy but also highlight the effectiveness of multi-classification in enhancing predictive abilities. In summary, the studies collectively contribute valuable insights into fisheries management using diverse ML techniques. While each algorithm has shown promise in specific contexts, there is a need for comprehensive studies that explore their scalability, generalizability, and robustness across varied datasets. Furthermore, the transition from single classifiers to multi-classification methods is identified as a key trend for improved predictive performance, although challenges in decision function selection and classifier grouping need careful consideration. Future research should aim to address these gaps for the development of more robust and versatile fish classification models.

**Table 2** Review analysis based on classification models applied to fish dataset

| Ref. | Classification model | Method | Results | Advantages | Limitations |
|---|---|---|---|---|---|
| [15] | SMO algorithm in water quality prediction for fish farming | Utilized SMO algorithm via library for support vector machine (LIBSVM) for solving quadratic programming in SVM training | Reported SMO as the best machine learning algorithm with an error rate of 0. | Efficient resolution of quadratic programming issues during SVM training. | Limited discussion on the scalability and generalizability of SMO. |
| [16] | NB for fish freshness identification | Applied NB algorithm for fish freshness identification based on eye images. | Successful identification of fish freshness using Bayes' theorem. | Straightforward probability-based classification. | Potential limitations in handling complex datasets with implicit independence assumptions. |
| [17] | Efficiency of MLP in rainbow trout classification | Demonstrated the efficiency of MLP with three layers for rainbow trout classification. | Satisfactory prediction of rainbow trout based on supervised training. | Ability to solve complex problems through supervised training. | Lack of exploration on the sensitivity of MLP to hyperparameters. |
| [18] | IBK learning algorithm for k-NN in fish classification (2023) | Applied IBK learning algorithm k-NN for fish classification based on nearest neighbors. | Successful identification of fish classes through distance weighting. | Flexibility in specifying the number of nearest neighbors. | Potential sensitivity to the choice of k and computational intensity for large datasets. |
| [19] | RF algorithm for tracking fishing activity | Used RF algorithm to distinguish fishing and non-fishing vessels in the North Sea. | Successful tracking of fishing activity by averting overfitting with multiple decision trees. | Reduced overfitting through the aggregation of decision trees. | Lack of discussion on the interpretability of the RF model. |
| [20] | Multi-classification method for enhanced performance | Employed multi-classification method combining weak and strong learners. | Demonstrated enhanced performance compared to individual classifiers. | Improved accuracy in training and testing data. | Challenges in selecting decision functions and optimal group of classifiers. |
| [21] | Single vs. multi-classification for improved prediction | Compared single machine learning with multi-classification. | Advocated the necessity of combining classifiers for improved prediction. | Highlighted the importance of combining classifiers for enhanced predictive abilities. | Limited exploration of specific scenarios where single classifiers may outperform multi-classifiers. |
| [22] | SVM for fish species identification | SVM applied to identify fish species | Achieved high accuracy in species identification based on morphological features | Robust performance in handling complex feature spaces | Sensitivity to noise in the data, potential challenges with scalability |
| [23] | Deep learning for fish recognition in underwater imagery | Convolutional neural networks (CNNs) used for fish recognition in underwater images | Demonstrated state-of-the-art accuracy in identifying diverse fish species | Ability to learn hierarchical features; effective in image-based classification | High computational requirements, dependence on large labelled datasets |
| [24] | Ensemble learning for fish abundance prediction | Ensemble of RF and Gradient Boosting for predicting fish abundance | Improved accuracy compared to individual models; robust predictions across varied conditions | Mitigates overfitting; handles nonlinear relationships | Challenges in interpretability; sensitivity to hyperparameter tuning |
| [25] | Feature selection in fish classification using Genetic algorithms | Genetic algorithms employed for feature selection in fish classification | Identified a subset of features leading to improved accuracy | Automatic identification of relevant features; reduced dimensionally | Computationally intensive; effectiveness depends on the representation of the Genetic algorithm |
| [26] | Time series analysis for fish migration prediction | Time series analysis applied to predict fish migration patterns | Successfully forecasted migration timing and routes | Captures temporal dependencies; aids in fisheries management | Limited applicability to non-migratory species; reliance on accurate time-stamped data |
| [27] | Transfer learning in fish classification | Transfer learning using pre-trained models for fish classification in new environments | Achieved competitive accuracy with reduced training data | Leverages knowledge from pre-trained models; enhances generalization | May not perform well if source and target domains are too dissimilar |
| [28] | Bayesian networks for fish behavior modelling | Bayesian networks applied to model and predict fish behavior patterns | Provided insights into complex interactions among environmental factors affecting fish behavior | Probabilistic modeling captures uncertainties; interpretable representation of dependencies | Relies on accurate prior knowledge; may struggle with modeling highly dynamic behaviors |

| Ref. | Classification model | Method | Results | Advantages | Limitations |
|---|---|---|---|---|---|
| [29] | Clustering-based approach for fish habitat prediction | Utilized clustering algorithms to predict suitable fish habitats | Identified distinct habitat clusters, aiding in habitat conservation efforts | Unsupervised learning approach; can reveal hidden patterns in habitat preferences | Sensitivity to initial conditions; challenges in defining optimal cluster numbers |
| [30] | Gaussian processes for fish stock prediction | Implemented Gaussian Processes to model and predict fish stock dynamics | Accurate predictions of fish stock fluctuations over time | Provides uncertainty estimates; flexibility in handling non-linear relationships | Computationally demanding for large datasets; challenging for high-dimensional input spaces |
| [31] | Explainable artificial intelligence (AI) in fish classification | Employed explainable AI techniques to enhance interpretability of fish classification models | Improved understanding of model decisions, aiding in user trust and acceptance | Transparent decision-making; facilities regulatory compliance | May sacrifice some accuracy for interpretability; challenges in explaining complex deep learning models |
| [32] | Spatial-temporal modelling for fish distribution | Developed spatial-temporal models to predict the distribution of fish populations | Accurately predicted changes in fish distribution over different seasons and geographical regions | Captures both spatial and temporal dependencies; valuable for fisheries management | Requires detailed spatial-temporal data; potential challenges in extrapolating to unseen conditions |
| [33] | Evolurionary algorithms for feature engineering | Applied evolutionary algorithms to automatically engineer features for fish classification | Identified non-trivial features, improving classification accuracy | Automatic feature generation; reduces manual feature engineering efforts | Computational complexity; effectiveness dependent on the representation of the evolutionary algorithm |
| [34] | Meta-learning for cross-dataset generalization | Implemented meta-learning techniques to improve generalization across diverse fish datasets | Enhanced model adaptability to new datasets with minimal retraining | Fast adaptation to new domains; addresses dataset shift issues | Requires careful selection of meta-learning algorithms; may struggle with highly dissimilar datasets |
| [35] | Hybrid model for fish species recognition | Developed a hybrid model combining rule-based systems and deep learning for fish species recognition | Achieved high accuracy while leveraging domain-specific knowledge | Synergy between rule-based and data-driven approaches; interpretable decision-making | Dependent on the availability and accuracy of domain rules; potential challenges in handling highly complex datasets |
| [36] | Ensemble of time series models for fish migration prediction | Ensemble approach combining various time series models for predicting fish migration patterns | Improved accuracy in capturing the nuances of migratory behavior | Robust predictions by leveraging different model strengths; accommodates temporal dependencies | Increased computational demands; potential challenges in integrating diverse time series models |
| [37] | Self-supervised learning for fish feature representation | Utilized self-supervised learning techniques for unsupervised feature representation learning | Discovered meaningful representations without explicit labels, improving downstream classification | Reduces reliance on labeled data; learns hierarchical features | May still require labeled data for fine-tuning; effectiveness dependent on data characteristics |
| [38] | Transferable knowledge distillation for small-scale fisheries | Employed transferable knowledge distillation techniques to train compact models suitable for small-scale fisheries | Successfully transferred knowledge from large-scale datasets, enabling accurate predictions in resource-constrained settings | Efficient model deployment in data-limited scenarios; leverages insights from larger datasets | Potential challenges in distilling complex knowledge; sensitivity to differences in data distributions |
| [39] | Multi-modal fusion for improved fish species identification | Utilized multi-modal fusion, combining information from images and acoustic data, to enhance fish species identification | Improved accuracy by leveraging complementary information from different sensor modalities | Comprehensive information integration; robust to variations in environmental conditions | Increased data preprocessing complexity; reliance on synchronized multi-modal data |
| [40] | Semi-supervised learning for fish population estimation | Applied semi-supervised learning techniques to utilize both labelled and unlabelled data for fish population estimation | Achieved accurate population estimates with limited labelled data, reducing the need for extensive annotation | Exploits unlabelled data effectively; cost-effective approach | Performance dependent on the quality of unlabelled data; potential challenges in noisy label scenarios |
| [41] | Hyperparameter optimization for improved fish classification | Investigated the impact of hyperparameter optimization techniques on fish classification | Identified optimal hyperparameter configurations leading to enhanced model | Fine-tuned model performance; improved generalization | Computational demands; may require extensive search in high-dimensional |

| Ref. | Classification model | Method | Results | Advantages | Limitations |
|---|---|---|---|---|---|
| | | performance | accuracy | | hyperparameter spaces |
| [42] | Domain adaptation for fish behavior prediction in unseen environments | Explored domain adaptation techniques to enhance the generalization of fish behavior prediction models to unseen environmental conditions | Improved model robustness when deployed in diverse ecological settings | Adapts to changes in environmental characteristics; enhances model transferability | Requires carefully annotated domain-specific data for adaptation; sensitivity to domain shift magnitude |
| [43] | Explainable fish classification using rule-based systems | Developed a rule-based system for fish classification to enhance interpretability | Provided transparent decision rules for each classification, aiding in understanding model predictions | Increased trust in decision-making; insights into decision rationale | Limited expressiveness for complex relationships; may struggle with capturing non-linear patterns |
| [44] | Meta-analysis of fish classification models | Conducted a meta-analysis of various fish classification models to identify common trends and performance benchmarks | Synthesized findings from multiple studies, offering a comprehensive overview of state-of-the-art techniques | Highlights consistent trends and best practices; aids in benchmarking | Subject to publication bias; challenges in harmonizing diverse experimental setups |
| [45] | Dynamic ensemble learning for adaptive fisheries management | Introduced a dynamic ensemble learning approach for adaptive fisheries management, adjusting classifier weights based on real-time environmental changes | Improved adaptability to fluctuating conditions, enhancing the reliability of fish predictions | Real-time responsiveness; addresses temporal variations in fish behavior | Requires continuous monitoring; potential challenges in rapid environmental shifts |

# 3.Methods

*Figure 2* depicts the functional block diagram of the proposed fish landing classification algorithm consisting of data acquisition and pre-processing, feature selection, and single and multi-classification. Information on instances and attributes was obtained in the data acquisition phase, while the data pre-processing step screened incomplete, noisy, and uncertain data. Class conditional probability was then applied during the feature selection step. This study compared five classification algorithms (SMO, NB, MLP, IBK, and RF) at the single classification. Subsequently, the classifiers were combined in the multi-classification phase to improve the prediction accuracy. Finally, the classifier performance was measured based on the accuracy and error rate.

Here is a simple representation of the fish landing classification algorithm.

Algorithm: Fish Landing Classification

Step 1. Input:
- Raw data containing instances and attributes related to fish landing.

Step 2. Data Acquisition and Pre-processing:

a. Acquire information on instances and attributes.

b. Pre-process the data:
- Screen for incomplete, noisy, and uncertain data.

Step 3. Feature Selection:

a. Apply class conditional probability for feature selection.

Step 4. Single Classification:

a. Compare five classification algorithms:
- SMO
- NB
- MLP
- IBK
- RF

b. Evaluate the performance of each algorithm individually:
- Measure accuracy and error rate.

Step 5. Multi-Classification:

a. Combine the classifiers:
- RF+SMO+NB+MLP (or any desired combination)
- SMO+RF+NB+MLP (or any desired combination)

b. Use combined classifiers to improve prediction accuracy.

Step 6. Performance Measurement:

a. Evaluate the performance of the multi-classification system:
- Measure accuracy and error rate.

Step 7. Output:
- Final classification results based on the chosen algorithm(s).

This algorithm outlines the key steps involved in the fish landing classification process, including data acquisition, pre-processing, and feature selection, single classification using various algorithms, multi-classification for improved accuracy, and the final performance measurement.

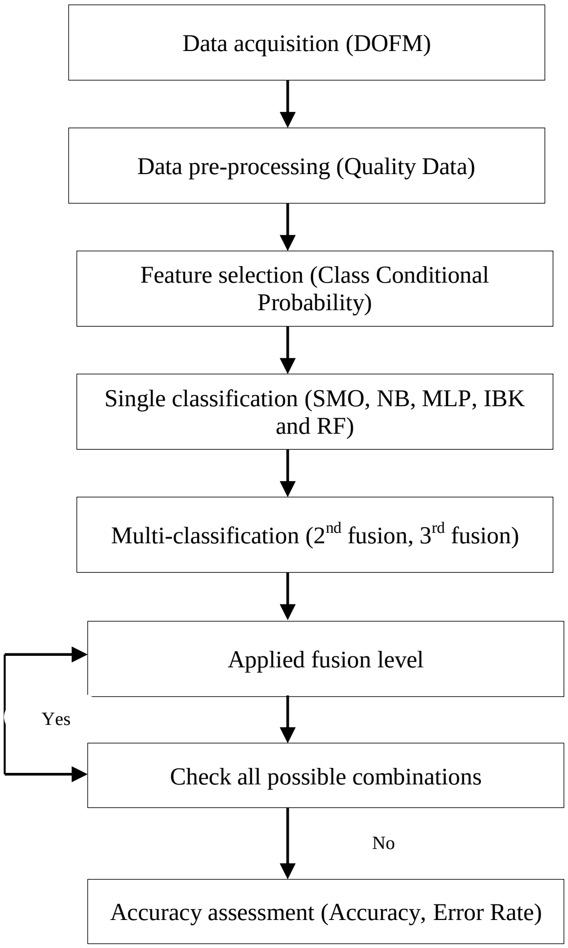


**Figure 2** Proposed fish landing classification algorithm

## 3.1Data acquisition

The dataset used is the fish landing dataset in East Coast of Peninsular Malaysia from 2014 until 2019. The dataset was sourced from the Department of Fisheries Malaysia (DOFM) website. It comprises 42 fish landing records for three fish species (K, BN, and S). The dataset includes the numbers of fish landing monthly.

## 3.2Data pre-processing

Data pre-processing is important since the dataset contains noisy, inconsistent, missing, and outdated values [46]. Data pre-processing is vital prior to dataset classification to improve data quality by identifying and removing noisy data. The data will be filtered to identify and remove the noisy data to improve the quality of the data. This step is applied to the dataset before the classification. Since the fish landing dataset is numerical, 'Descritize' was applied to the dataset. It is an instance filter built in the WEKA to discretize numerical into nominal attributes. This filter can be applied by weka > filters > supervised > attributes > Discretize.

The algorithm integrates data acquisition and pre-processing seamlessly. Instead of treating these steps as separate entities, this approach recognizes the interconnectedness of obtaining instances and attributes with the need for thorough data screening. The holistic treatment of data acquisition and pre-processing ensures that the algorithm starts with a refined dataset, reducing the impact of incomplete, noisy, and uncertain data on subsequent stages.

## 3.3Feature selection

Implementing feature selection on the fish landing dataset improves classification accuracy by removing noisy features and disregarding unnecessary features. After testing various feature selections on the fish landing dataset, class conditional probability consistently yielded the most promising results. Class conditional probability is particularly relevant in the context of fish classification because it takes into account the dependencies between attributes given the class labels. In the domain of fish landing, where attributes represent monthly collected data, it is essential to consider how attributes interrelate, especially in relation to different fish classes. The relevance of class conditional probability to the specific characteristics of the fish landing dataset, where attributes display significant interrelatedness, makes it a suitable choice. The method aligns with the nature of the data, ensuring that feature selection is tailored to the intricacies of fish classification. The analysis was conducted using WEKA, which assesses dependencies among features. The dependence of two attributes is calculated using the conditional probabilities of the class attribute [47]. Given its aptitude for evaluating feature dependencies, the class conditional probability was selected as the preferred feature selection method. The attributes represent monthly collected data, displaying significant interrelatedness.

The feature selection phase incorporates class conditional probability. This strategic integration leverages class-specific information to guide the selection of relevant features, emphasizing the importance of attributes specific to fish classification. Many existing approaches overlook the incorporation of class conditional probability during feature selection. This novel aspect enhances the algorithm's ability to focus on attributes crucial for discriminating between different fish classes.

Feature selection contributes to improved classification accuracy by eliminating noisy and unnecessary features. By focusing on the most relevant attributes, the algorithm can discern patterns and relationships more effectively. The enhanced accuracy resulting from feature selection signifies a more robust and efficient classification algorithm, which is crucial for reliable predictions in the domain of fish landing. Besides that, feature selection not only improves prediction accuracy but also enhances the interpretability of the model. By focusing on a subset of relevant features, it becomes easier to understand the factors contributing to classification decisions. While feature selection enhances performance, there are trade-offs to consider. Aggressive feature reduction may lead to information loss, especially if certain attributes contain valuable insights. Striking the right balance is crucial. The trade-off involves finding a subset of features that optimally balances accuracy improvements with the retention of critical information. Careful consideration is needed to avoid excluding attributes that may contribute to a comprehensive understanding of fish classification. Trade-off can be managed by perform cross-validation to assess the generalization performance of the algorithm with the selected features. This helps ensure that the feature subset is not overfitting to the training data and is applicable to unseen instances.

### 3.4Single classification

Classification is a two-stage process that involves training and testing. The training stage is a learning stage where the classification model is constructed based on the input data. Meanwhile, the classifier model is measured in the testing stage. A 10-fold cross-validation was applied in the classification experiment to produce consistent results for the top five (SMO, NB, MLP, IBK, and RF) according to the literature on fisheries. The dataset was divided into 10 folds in WEKA. Ten times, the model is trained and assessed, with the remaining nine folds being used for training and a separate fold serving as the test set. The study conducts a detailed comparison of five classification algorithms (SMO, NB, MLP, IBK, and RF) at the single classification level. This ensures a comprehensive understanding of the strengths and weaknesses of each algorithm in the context of fish landing. While previous studies may focus on individual algorithms, this approach systematically evaluates multiple classifiers, providing a more robust foundation for selecting the most suitable algorithm.

### 3.5Multi-classification

Multiple classifiers were combined (termed “fusion”) to improve the accuracy of a single classification. The top two classifiers in terms of accuracy are selected, followed by the rest of the algorithms until the accuracy declines, at this point, the task will stop. First, the two models with the highest accuracy in the single classification were chosen. Then, these classifiers were combined to create the second fusion. The aim of selecting two classifiers with the highest accuracy is to maximize the performance of the model. This approach may raise the strengths of multiple models by improving the overall accuracy. The fusion that achieves the highest accuracy was combined with another classifier derived from successive fusions. Fusion derivation was halted when there was no more possible combination. The dynamic fusion of classifiers is a distinctive feature, acknowledging that no single classifier may be universally optimal.

### 3.6Accuracy assessment

Both single and multi-classification performance was assessed based on the accuracy and error rate percentage. The single classifier with the highest accuracy is used as a base for multi-classification fusion. The base classifier is combined with the other single classifier to build the second fusion. The second fusion with the highest accuracy is then used as a base for the next fusion. The performance of the multi-classification system is meticulously measured based on accuracy and error rate. This thorough evaluation ensures a quantitative assessment of the algorithm's effectiveness. The accuracy and error rate are complements of each other, they are the most frequently used metrics for assess performance in classification [48]. The classification with the highest accuracy will gain the lowest error rate.

In summary, the proposed fish landing classification algorithm stands out in its holistic integration of data acquisition and pre-processing, incorporation of class conditional probability, comprehensive comparison of single classification algorithms, dynamic multi-classification fusion, and thorough performance measurement. These novel aspects collectively contribute to the advancement of fish classification methodologies, addressing existing challenges and providing a robust framework for informed decision-making in the fisheries sector.

# 4.Results

The proposed work is using the single and multi-classifier methods for the fish landing dataset. In this section, we will interpret and discuss the results presented in the previous section.

## 4.1Single classification task

*Figure 3* shows the accuracy performance for five classifiers (SMO, NB, MLP, IBK, and RF) with and without feature selection. RF demonstrated the highest accuracy (73.80%) among all classifiers without feature selection. Nonetheless, RF and SMO achieved the highest accuracy with the same percentage (76.19%) after applying the feature selection, indicating its effectiveness in handling the selected features. Meanwhile, NB and MLP recorded lower accuracy at 71.42% each. It often performs well with certain types of data, especially when the independence assumption holds. Therefore, these classifiers were superior compared to IBK.

## 4.2Multi-classifiers fusion classification task

*Figure 4* shows the result of RF and SMO combined with other classifiers. Both RF and SMO were selected to be tested as the base because both achieved the highest accuracy in the single classification. The RF+SMO and SMO+RF fusions performed better than other fusions by charting the highest accuracy (76.19%), followed by RF+NB, RF+MLP, RF+IBK, SMO+NB, SMO+MLP, and SMO+IBK with the same percentage (71.43%). Both RF and SMO individually demonstrated high accuracy in single classifications. Combining these classifiers might have leveraged their respective strengths, such as RF's ability to handle complex relationships and SMO's versatility in classifying linear and non-linear data. The success of RF+SMO and SMO+RF fusions in achieving the highest accuracy suggests a synergy between these two classifiers. The equal accuracy among other fusion combinations implies a balanced performance, influenced by the compatibility of individual classifier characteristics. *Figure 5* demonstrates the findings of the three-classifier fusion. The highest classification accuracy was obtained by combining RF+SMO+NB and SMO+RF+NB (76.19%), while the RF+SMO+MLP, RF+SMO+IBK, SMO+RF+MLP, and SMO+RF+IBK fusions achieved similar percentages (73.81%). The fusion (RF+SMO+NB and SMO+RF+NB) achieved the highest accuracy. It leverages the strengths of RF, SMO, and NB, potentially benefitting from RF's ability to handle complexity, SMO's capability with non-linear data, and NB's simplicity and efficiency.

*Figure 6* shows the output for the fusion of four classifiers, where RF+SMO+NB+MLP and SMO+RF+NB+MLP (80.95%) performed better than RF+SMO+NB+IBK and SMO+RF+NB+IBK (73.81%). The fusion of RF+SMO+NB+MLP and SMO+RF+NB+MLP outperformed others due to the complementary strengths of the individual classifiers. RF and SMO offer robustness and versatility, NB provides simplicity, and MLP contributes the ability to capture complex relationships.

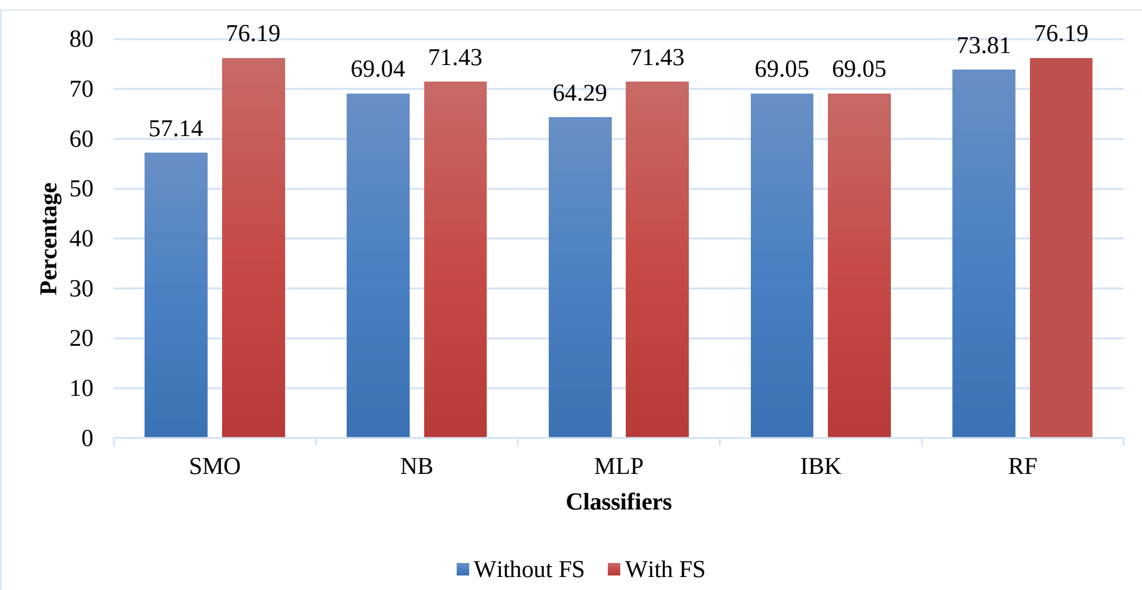


**Figure 3** Single classifier with and without feature selection (FS)

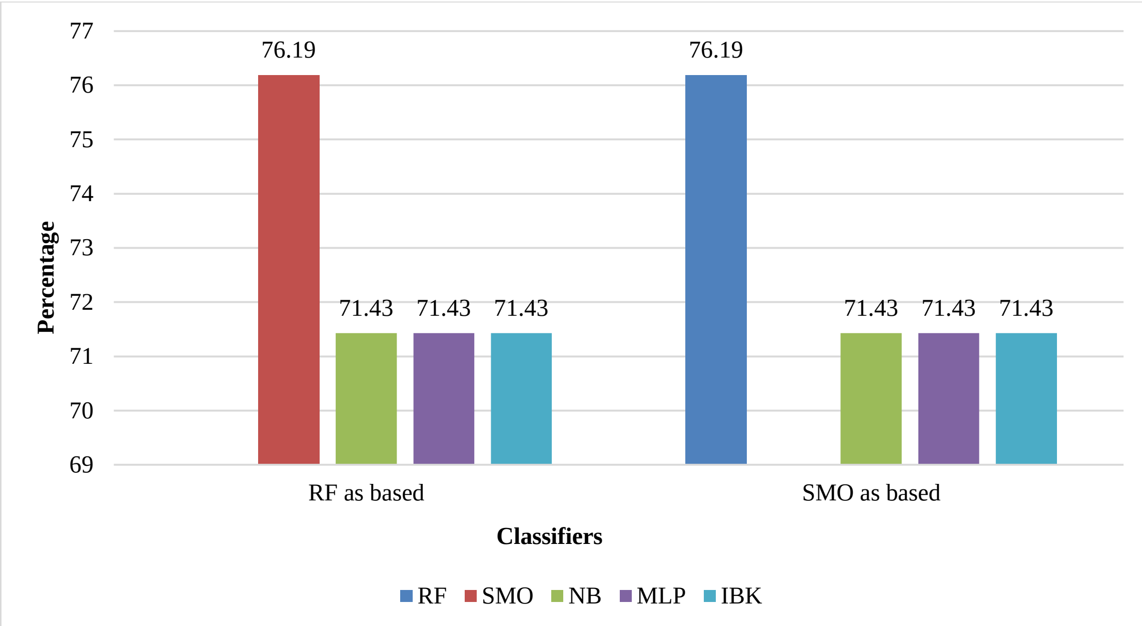


**Figure 4** Fusion of two classifiers

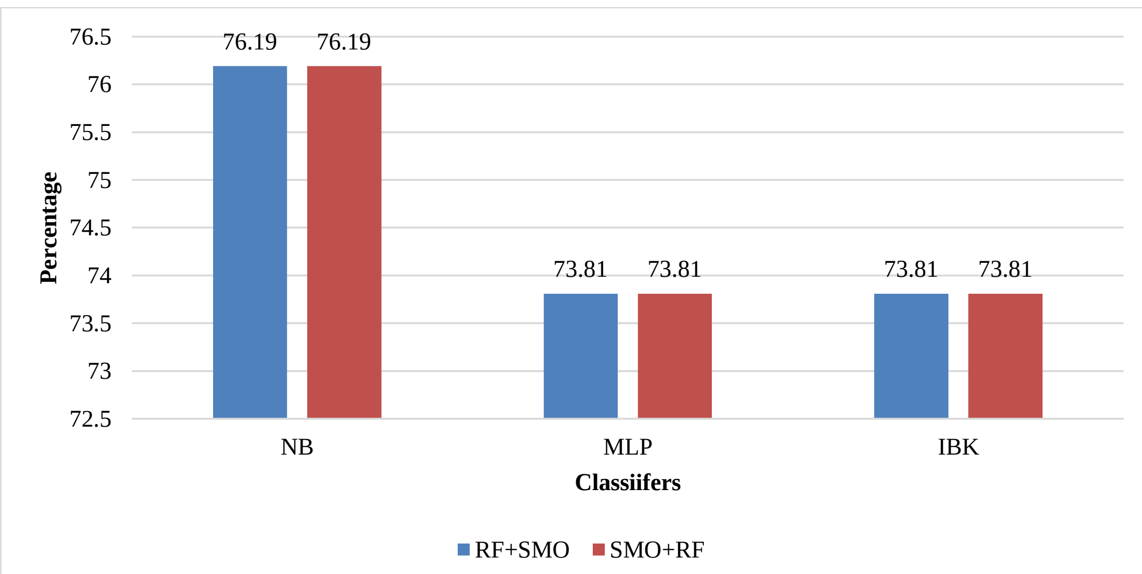


**Figure 5** Fusion of three classifiers

*Table 3* and *Table 4* show the confusion matrices for the top two fusion of four classifiers, RF+SMO+NB and SMO+RF+NB. The both fusion RF+SMO+NB and SMO+RF+NB are making the same predictions across different classes.

**Table 3** Confusion matrix for RF+SMO+NB

| | | Predicted | | |
|---|---|---|---|---|
| | | **K** | **S** | **BN** |
| Actual | K | 6 | 7 | 1 |
| | S | 2 | 12 | 0 |
| | BN | 0 | 0 | 14 |

**Table 4** Confusion matrix for SMO+RF+NB

| | | Predicted | | |
|---|---|---|---|---|
| | | **K** | **S** | **BN** |
| Actual | K | 6 | 7 | 1 |
| | S | 2 | 12 | 0 |
| | BN | 0 | 0 | 14 |

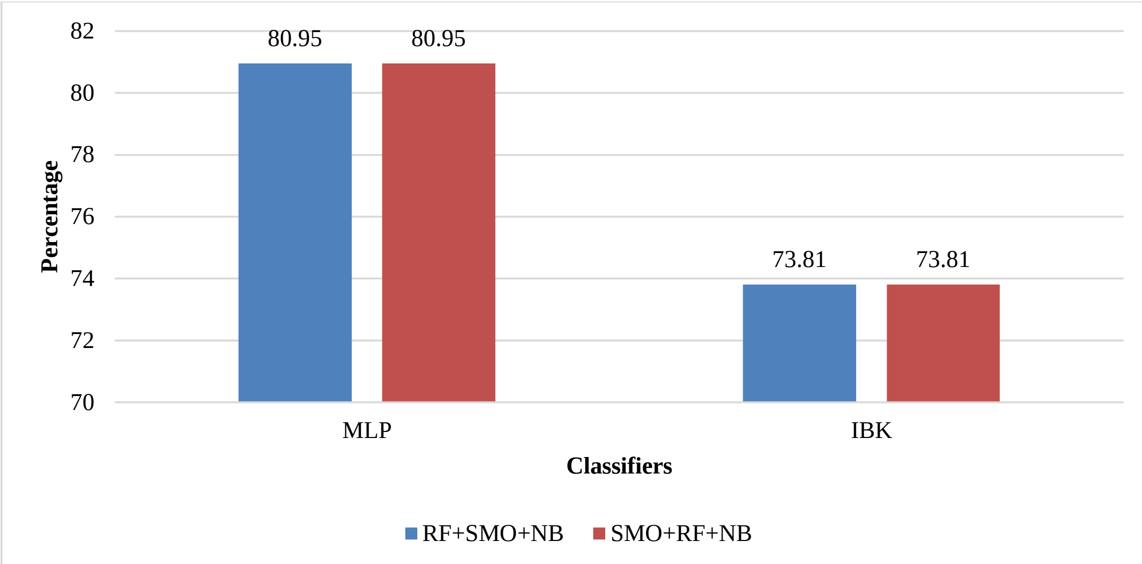


**Figure 6** Fusion of four classifiers

*Table 5* compares the accuracy of single classification and multi-classification. The best accuracy for single classification was achieved by RF and SMO. RF and SMO achieved the highest accuracy (76.19%) individually, showcasing their effectiveness as standalone classifiers. Nevertheless, the accuracy of the 4th fusion, RF+SMO+NB+MLP, and SMO+RF+NB+MLP, was superior to the single classification, indicating that classification accuracy can be improved by combining the classifiers. These fusion outperformed single classifiers, highlighting the potential benefits of combining classifiers. These fusions achieved the highest accuracy (80.95%). Overall, the results highlight the potential of classifier fusion for enhancing accuracy.

**Table 5** Comparison of the accuracy

| | | **RF** | **SMO** | **NB** | **MLP** | **IBK** |
|---|---|---|---|---|---|---|
| Single Classifier | | 76.19 | 76.19 | 71.43 | 71.43 | 69.05 |
| 2nd Fusion | (RF as base) | NA | 76.19 | 71.43 | 71.43 | 71.43 |
| | (SMO as base) | 76.19 | NA | 71.43 | 71.43 | 71.43 |
| 3rd Fusion | (RF+SMO as base) | NA | NA | 76.19 | 73.81 | 73.81 |
| | (SMO+RF as base) | NA | NA | 76.19 | 73.81 | 73.81 |
| 4th Fusion | (RF+SMO+NB as base) | NA | NA | NA | 80.95 | 73.81 |
| | (SMO+RF+NB as base) | NA | NA | NA | 80.95 | 73.81 |

## 4.3Error rate

Apart from accuracy, the classification algorithm is also evaluated based on the lowest error rate of the dataset. *Table 6* presents the error rates for various combinations of classifiers, highlighting the effectiveness of different fusion levels in minimizing errors. At the 2nd fusion level, various combinations of two classifiers were evaluated. The error rates ranged from 23.81% to 28.57%. Notably, RF+SMO and SMO+RF achieved the lowest error rates at this fusion level. The 3rd fusion level involved combinations of three classifiers. Notably, RF+SMO+NB and SMO+RF+NB achieved the lowest error rates at 23.81%, suggesting that the inclusion of specific classifiers contributed to error rate reduction. At the 4th fusion level, combinations of four classifiers were considered. Impressively, RF+SMO+NB+MLP and SMO+RF+NB+MLP achieved the lowest error rates at 19.05%, indicating that the fusion of these specific classifiers led to significant improvement in classification accuracy. The results suggest that the combination of RF, SMO, NB, and MLP classifiers in different orders consistently outperformed other combinations in terms of error rates. Notably, the 4th fusion level with RF+SMO+NB+MLP and SMO+RF+NB+MLP achieved the lowest error rates, highlighting the effectiveness of these specific classifier combinations.

**Table 6** Error rate

| Fusion level | Classifiers | Error rate |
|---|---|---|
| 2nd | RF+SMO | 23.81 |
| | RF+NB | 28.57 |
| | RF+MLP | 28.57 |
| | RF+IBK | 28.57 |
| | SMO+RF | 23.81 |
| | SMO+NB | 28.57 |
| | SMO+MLP | 28.57 |
| | SMO+IBK | 28.57 |
| 3rd | RF+SMO+NB | 23.81 |
| | RF+SMO+MLP | 26.19 |
| | RF+SMO+IBK | 26.19 |
| | SMO+RF+NB | 23.81 |
| | SMO+NB+IBK | 26.19 |
| 4th | RF+SMO+NB+MLP | 19.05 |
| | RF+SMO+NB+IBK | 26.19 |
| | SMO+RF+NB+MLP | 19.05 |
| | SMO+RF+NB+IBK | 26.19 |

## 5.Discussion

Five classifiers; SMO, NB, MLP, IBK, and RF were applied to the fish landing dataset. The accuracy of each classifier was recorded. Since the performance of the single classification is unsatisfactory, the multi-classification method was implemented on the dataset. The classifiers were gradually combined using majority voting in WEKA. The classifiers with the highest accuracy are selected as the base of the fusion. The findings of this study can be summarized as follows:

i) The class conditional probability was selected as a feature selection. The performance of single classifications was improved using this feature selector. The utilization of class conditional probability as a feature selection technique is a novel aspect of our approach. This method demonstrated improvements in single classification performance, showcasing its potential to enhance the relevance and interpretability of features in fish dataset classification.

ii) RF and SMO were selected as the base for the second fusion during multi-classification since their accuracy is the highest on single classification. The observation that there is no significant difference in accuracy when using either RF or SMO as a base because both have the same accuracy and plotted the same confusion matrix in the single classification. This provides flexibility in choosing classifiers based on factors such as computational efficiency or interpretability.

iii) The performance of the multi-classification is increased when the number of classifiers combined is increased. The accuracy of the fourth fusion (RF+SMO+NB+MLP/SMO+RF+NB+MLP) is better than the third fusion (RF+SMO+NB/SMO+RF+NB). It was improved from 76.19 to 80.95%. The finding that the accuracy of multi-classification improves as the number of classifiers combined increases highlights the potential benefits of incorporating diverse classifiers.

iv) The accuracy and error rate are complements of each other, the higher the accuracy, the lower and the error rate. For example, the accuracy of RF+SMO+NB+MLP fusion is 80.95% while its error rate is 19.05%. The recognition of the complementarity between accuracy and error rate is an essential observation. This understanding can guide decision-making in real-world applications by considering both metrics simultaneously.

The refined classification algorithm, especially the multi-classification approach with feature selection, can find applications in fisheries management systems to assist in species identification and landing predictions. The insights into the interchangeability of certain classifiers and the impact of increasing the number of classifiers in fusion provide practical guidance for implementing robust and adaptable classification systems in real-world scenarios.

It can be concluded that the practical implications of these findings extend to improved feature selection, optimized classifier selection for fusion, and considerations for system adaptability and performance evaluation in real-world applications, particularly in fisheries and related domains.

### 5.1Limitations

The findings are based on the fish landing dataset used in this study. The effectiveness of the proposed algorithm may be influenced by the specific characteristics of this dataset. It could be considered to testing the algorithm on diverse fish datasets to assess its generalizability and robustness across different contexts. While the algorithm shows promise in fisheries management, its applicability to other domains may vary. Different datasets with distinct characteristics may require further customization and validation. Another limitation that can be highlighted is use of class conditional probability as a feature selection technique assumes independence among features, which may not hold in all scenarios. Explore alternative feature selection methods that account for potential dependencies among features to further improve the model's accuracy could also be considered.

A complete list of abbreviations is summarised in *Appendix I.*

## 6.Conclusion and future work

Accurate prediction of the fish landing dataset is pivotal for effective fisheries management planning, and leveraging machine learning classifier algorithms can provide valuable insights. This study's findings underscore the exceptional performance of the RF+SMO+NB+MLP and SMO+RF+NB+MLP fusion classifiers, surpassing both individual classifiers and alternative fusion combinations with a remarkable accuracy of 80.95% and an impressively low error rate of 19.05%. The results affirm the potential of these fusion classifiers as promising tools for precise classification in fish landing datasets. Future work could explore the development of deep learning methods, aiming to further enhance the classification performance in fisheries management applications.


### Acknowledgment

We thank Universiti Malaysia Terengganu for providing funding support for this project (UMT/TAPE-RG 2021/Vot 55327).


### Conflicts of interest

The authors have no conflicts of interest to declare.

### Data availability

The dataset used in this study is not available to the public. This is due to restrictions related to the inclusion of information that could potentially compromise the privacy and confidentiality of the individuals or entities involved. Consequently, the dataset cannot be shared openly, in alignment with ethical standards and privacy regulations that govern the collection and use of sensitive data.

### Author’s contribution statement

All authors contributed equally to the study's conception and design, data collection, analysis, interpretation of results, and manuscript preparation.

### References


[1] Fathoni S, Rachman MA, Arasy AK. Analysis determinant supply & demand fisheries. In IOP conference series: earth and environmental science 2019 (pp. 1-7). IOP Publishing.

[2] Alhatali A, Soosaimanickam A. A comparative study of the efficient data mining algorithm to find the most influenced factor on price variation in Oman fish markets. Sakarya University Journal of Computer and Information Sciences. 2018; 1(2):1-6.

[3] Islam MM, Chuenpagdee R. Towards a classification of vulnerability of small-scale fisheries. Environmental Science & Policy. 2022; 134:1-2.

[4] Andrews N, Bennett NJ, Le BP, Green SJ, Cisneros-montemayor AM, Amongin S, et al. Oil, fisheries and coastal communities: a review of impacts on the environment, livelihoods, space and governance. Energy Research & Social Science. 2021; 75:102009.

[5] Sheth V, Tripathi U, Sharma A. A comparative analysis of machine learning algorithms for classification purpose. Procedia Computer Science. 2022; 215:422-31.

[6] Phyu TN. Survey of classification techniques in data mining. In proceedings of the international multiconference of engineers and computer scientists 2009 (pp. 727-31).

[7] Seliya N, Abdollah ZA, Khoshgoftaar TM. A literature review on one-class classification and its potential applications in big data. Journal of Big Data. 2021; 8(1):1-31.

[8] Chen P, Fan R, Lin C. A study on SMO-type decomposition methods for support vector machines. IEEE Transactions on Neural Networks. 2006; 17(4): 893-908.

[9] https://aiml.com/what-is-a-multilayer-perceptron-mlp/. Accessed 19 December 2023.

[10] David SK, Rafiullah M, Siddiqui K. Comparison of different machine learning techniques to predict diabetic kidney disease. Journal of Healthcare Engineering. 2022; 2022:1-9.

[11] Rodriguez-galiano VF, Ghimire B, Rogan J, Chica-olmo M, Rigol-sanchez JP. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing. 2012; 67:93-104.

[12] Riccio V, Jahangirova G, Stocco A, Humbatova N, Weiss M, Tonella P. Testing machine learning based systems: a systematic mapping. Empirical Software Engineering. 2020; 25:5193-254.

[13] Ribeiro VH, Reynoso-meza G. Ensemble learning by means of a multi-objective optimization design approach for dealing with imbalanced data sets. Expert Systems with Applications. 2020; 147:113232.

[14] Rosly R, Makhtar M, Awang MK, Awang MI, Rahman MN, Mahdin H. Comprehensive study on ensemble classification for medical applications. International Journal of Engineering & Technology. 2018; 7(2.14):186-90.

[15] Mahfouz A, Abuhussein A, Venugopal D, Shiva S. Ensemble classifiers for network intrusion detection using a novel network attack dataset. Future Internet. 2020; 12(11):1-19.

[16] Saputra S, Yudhana A, Umar R. Implementation of naïve bayes for fish freshness identification based on image processing. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi). 2022; 6(3):412-20.

[17] Rohani A, Taki M, Bahrami G. Application of artificial intelligence for separation of live and dead

rainbow trout fish eggs. Artificial Intelligence in Agriculture. 2019; 1:27-34.
[18] Maniriho P, Mahoro LJ, Niyigaba E, Bizimana Z, Ahmad T. Detecting intrusions in computer network traffic with machine learning approaches. International Journal of Intelligent Engineering & Systems. 2020; 13(3):433-45.
[19] Snapir B, Waine TW, Biermann L. Maritime vessel classification to monitor fisheries with SAR: demonstration in the North Sea. Remote Sensing. 2019; 11(3):1-16.
[20] Lastanto JA, Djatna T. A predictive analytics of production planning model for yield performance in organic broiler industry using ensemble learning. In IOP conference series: earth and environmental science 2022 (pp. 1-9). IOP Publishing.
[21] Jany Arman R, Hossain M, Hossain S. Fish classification using saliency detection depending on shape and texture. Computación y Sistemas. 2022; 26(1):303-10.
[22] Ogunlana SO, Olabode O, Oluwadare SA, Iwasokun GB. Fish classification using support vector machine. African Journal of Computing & ICT. 2015; 8(2):75-82.
[23] Deep BV, Dash R. Underwater fish species recognition using deep learning techniques. In 6th international conference on signal processing and integrated networks 2019 (pp. 665-9). IEEE.
[24] Rahman LF, Marufuzzaman M, Alam L, Bari MA, Sumaila UR, Sidek LM. Developing an ensembled machine learning prediction model for marine fish and aquaculture production. Sustainability. 2021; 13(16):1-14.
[25] Amini F, Hu G. A two-layer feature selection method using genetic algorithm and elastic net. Expert Systems with Applications. 2021; 166:114072.
[26] Hou D. Determine the impact on fisheries by predicting the migration of fish near Scotland. In IOP conference series: earth and environmental science 2021 (pp. 1-7). IOP Publishing.
[27] Agarwal AK, Tiwari RG, Khullar V, Kaushal RK. Transfer learning inspired fish species classification. In 8th international conference on signal processing and integrated networks 2021 (pp. 1154-9). IEEE.
[28] Little LR, Kuikka S, Punt AE, Pantus F, Davies CR, Mapstone BD. Information flow among fishing vessels modelled using a Bayesian network. Environmental Modelling & Software. 2004; 19(1):27-34.
[29] Fitrianah D, Fahmi H, Hidayanto AN, Arymurthy AM. A data mining based approach for determining the potential fishing zones. International Journal of Information and Education Technology. 2016; 6(3):187-91.
[30] Boettiger C, Mangel M, Munch S. Avoiding tipping points in fisheries management through gaussian process dynamic programming. Proceedings of the Royal Society B: Biological Sciences. 2015; 282(1801):1-9.
[31] Kamakshi V, Krishnan NC. Explainable image classification: the journey so far and the road ahead. AI. 2023; 4(3):620-51.
[32] Badger JJ, Large SI, Thorson JT. Spatio-temporal species distribution models reveal dynamic indicators for ecosystem-based fisheries management. ICES Journal of Marine Science. 2023; 80(7):1949-62.
[33] Zhang H, Chen Q, Zhou A, Wang Y. Beyond deep learning: an evolutionary feature engineering approach to tabular data classification. ICLR conference 2023 (pp.1-23).
[34] Li D, Yang Y, Song YZ, Hospedales T. Learning to generalize: meta-learning for domain generalization. In proceedings of the AAAI conference on artificial intelligence 2018 (pp.3490-7).
[35] Chhabra HS, Srivastava AK, Nijhawan R. A hybrid deep learning approach for automatic fish classification. In proceedings of ICETIT: emerging trends in information technology 2020 (pp. 427-36). Springer International Publishing.
[36] Alfatinah A, Chu HJ, Tatas, Patra SR. Fishing area prediction using scene-based ensemble models. Journal of Marine Science and Engineering. 2023; 11(7):1-16.
[37] Saleh A, Sheaves M, Jerry D, Azghadi MR. Transformer-based self-supervised fish segmentation in underwater videos. arXiv preprint arXiv:2206.05390. 2022**.**
[38] Zainuddin M, Safruddin S, Selamat MB, Farhum A, Hidayat S. Prediction of potential fishing zones for skipjack tuna during the northwest monsoon using remotely sensed satellite data. Marine Science. 2017; 22(2):59-66.
[39] Mugo R, Saitoh SI. Ensemble modelling of Skipjack tuna (Katsuwonus pelamis) habitats in the western North Pacific using satellite remotely sensed data; a comparative analysis using machine-learning models. Remote Sensing. 2020; 12(16):1-15.
[40] Alaba S, Shah C, Nabi MM, Ball J, Moorhead R, Han D, et al. Semi-supervised learning for fish species recognition. In ocean sensing and monitoring XV 2023 (pp. 248-55). SPIE.
[41] Saleh A, Sheaves M, Jerry D, Azghadi MR. Applications of deep learning in fish habitat monitoring: a tutorial and survey. Expert Systems with Applications. 2023:121841.
[42] Malik H, Naeem A, Hassan S, Ali F, Naqvi RA, Yon DK. Multi-classification deep neural networks for identification of fish species using camera captured images. Plos One. 2023; 18(4):1-32.
[43] Alkhulaifi A, Alsahli F, Ahmad I. Knowledge distillation in deep learning and its applications. PeerJ Computer Science. 2021; 14:1-24.
[44] Masuda H, Jukei T, Hasegawa T. Fish species identification using a CNN-based multimodal learning method. In proceedings of the 2nd international conference on image, video and signal processing 2020 (pp. 15-9). ACM.
[45] Edmondson E, Fanning L. Implementing adaptive management within a fisheries management context: a

systematic literature review revealing gaps, challenges, and ways forward. Sustainability. 2022; 14(12):1-16.
[46] Palaniappan S, Hameed NA, Mustapha A, Samsudin NA. Classification of alcohol consumption among secondary school students. JOIV: International Journal on Informatics Visualization. 2017; 1(4-2):224-6.
[47] Amarnath B, Balamurugan SAA. Conditional probability based feature selector for effective data classification. Revista Tecnica De La Facultad De Ingenieria Universidad Del Zulia. 2016; 39(7): 1-8.
[48] https://machinelearningmastery.com/failure-of-accuracy-for-imbalanced-class-distributions/. Accessed 14 June 2023.

**Rosaida Rosly** is a senior lecturer at the Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu (UMT). She received her PhD and MSc in Computer Science from Universiti Sultan Zainal Abidin (UniSZA). Her research interests include Data Mining, Machine Learning, Deep Learning and Classification.
Email: rosaida@umt.edu.my

**Mustafa Man** is a Associate Professor at the Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu (UMT). Heobtained his PhD from Universiti Teknologi Malaysia (UTM) and MSc from Universiti Putra Malaysia (UPM). His research interests encompass Database, Internet of Things (IOT), image processing and data mining.
Email: mustafaman@umt.edu.my

**Amir Ngah** is a Associate Professor at the Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu (UMT). He received his PhD from University of Durham and MSc from UPM. His research interests include Software Engineering, Software Maintenance and Software Testing.
Email: amirnma@umt.edu.my

**Nor Saidah Abd Manan** is a research assistant at the Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu (UMT). She received her MSc in Computer Science from Universiti Sultan Zainal Abidin (UniSZA). Her research interests include Data Mining and Classification.
Email: norsaidahabdmanan.91@gmail.com

## Appendix I

| S. No. | Abbreviation | Description |
|---|---|---|
| 1 | AI | Artificial Intelligence |
| 2 | ANN | Artificial Neural Networks |
| 3 | BN | Biji Nangka |
| 4 | CNNs | Convolutional Neural Networks |
| 5 | DOFM | Department of Fisheries Malaysia |
| 6 | IBK | Instance-Based for K-Nearest Neighbor |
| 7 | K | Kebasi |
| 8 | k-NN | k-Nearest Neighbor |
| 9 | LIBSVM | Library for Support Vector Machine |
| 10 | MLP | Multi-Layer Perception |
| 11 | NB | Naïve Bayes |
| 12 | RF | Random Forest |
| 13 | S | Siakap |
| 14 | SMO | Sequential Minimal Optimization |
| 15 | SVM | Support Vector Machine |
| 16 | WEKA | Waikato Environment for Knowledge Analysis |