Machine Learning Benchmarking under Data-Limited Conditions: A Case Study of Environmental Associations with *Loligo* spp. Catch Characteristics

Wasana Arkronrat[1], Chonlada Leearam[1], Rungtiwa Konsantad[1], Vutthichai Oniam[1,2*]

[1]Klongwan Fisheries Research Station, Faculty of Fisheries, Kasetsart University, Prachuap Khiri Khan 77000, Thailand
[2]Research and Academic Service Center for Andaman Region, Faculty of Fisheries, Kasetsart University, Krabi 81130, Thailand
* Corresponding author: Vutthichai Oniam, ffisvco@ku.ac.th

Abstract
The application of machine learning (ML) in small-scale fisheries is often constrained by limited data availability, posing challenges for model development and reliability. This study presents a ML benchmarking framework for data-limited artisanal fisheries, aiming to evaluate the predictive performance of supervised algorithms in assessing the effects of environmental variability on *Loligo* spp. catch size and quantity. The dataset comprised 278 fishing-trip observations collected from artisanal fishing vessels between October 2024 and September 2025. Catch records were integrated with environmental variables, including sea surface temperature (SST), sea surface salinity (SSS), dissolved oxygen (DO), and air temperature. Given the limited data typically available in small-scale fisheries, the study emphasizes model performance and interpretability as key considerations for practical fisheries applications. Three classification models—Decision Tree (DT), k-Nearest Neighbors (k-NN), and Logistic Regression (LR)—were comparatively evaluated to assess their predictive performance. Results identified SSS as an important environmental variable associated with squid catch characteristics. For catch size classification (≤10 cm vs. >10 cm mantle length), DT achieved the highest accuracy (89.04%) among the evaluated models. In contrast, catch-quantity classification across three categories showed lower performance overall, with DT achieving the best accuracy (57.14%). These findings demonstrate that, under data-limited conditions typical of artisanal fisheries, simpler and interpretable models such as DT can provide strong predictive performance. This study highlights the practical applicability of ML in small-scale fisheries and provides a methodological reference for low-data environments, supporting decision-making and adaptive fisheries management under environmental variability.

Keywords: Artisanal squid fishery, Classification performance, Environmental variability, Machine learning models, Sea surface salinity

1. Introduction

Machine learning (ML) has become an increasingly important tool for analyzing complex environmental and fisheries data, offering the ability to identify patterns and generate predictive insights beyond the capacity of traditional statistical approaches (Faisal et al., 2016; Gladju et al., 2022; Panda and Baral, 2023). In the fisheries sector, ML applications have demonstrated potential for improving catch prediction, optimizing fishing effort, and supporting sustainable resource management under changing environmental conditions (Jones et al., 2021; Kühn et al., 2025). However, most ML-based fisheries studies rely on large, high-resolution datasets, which are typically unavailable in small-scale or artisanal fisheries.

Artisanal fisheries, particularly in developing regions, are often characterized by limited data availability, fragmented records, and resource constraints in data collection (Arkronrat et al., 2017; Sukhsangchan et al., 2020). These limitations pose significant challenges for the application of advanced ML techniques, especially data-intensive models such as deep learning. As a result, there is a critical need to evaluate the suitability and performance of ML models under low-data conditions, where data scarcity may affect model reliability, stability, and generalizability. Despite the growing interest in ML applications in fisheries, studies explicitly addressing low-data scenarios remain limited.

Squid (*Loligo* spp.) represents an economically important fisheries resource, supporting both commercial and small-scale fishing operations across many regions (Nasution et al., 2020; Suryanto et al., 2021; FAO, 2025). In Thailand, particularly in Prachuap Khiri Khan province, artisanal squid fisheries play a key role in local livelihoods and coastal economies (Arkronrat et al., 2017; Konsantad and Arkrontat, 2025; Leearam et al., 2025). Environmental factors, including sea surface temperature (SST), salinity (SSS), and dissolved oxygen (DO), as well as atmospheric conditions, are known to influence squid distribution, growth, and catch variability (Rahman et al., 2022). However, the application of ML to assess these relationships in artisanal squid fisheries remains underexplored, especially under data-limited conditions.

Although ML applications in fisheries are increasingly common, most studies rely on relatively large datasets and focus on commercial-scale fisheries (Harford et al., 2019; Gladju et al., 2022; Song et al., 2023; Kühn et al., 2025). In contrast, artisanal fisheries often face data limitations that restrict the application of advanced predictive approaches. Therefore, the novelty of the present study lies not in the use of machine learning itself, but in benchmarking three complementary supervised learning algorithms under a small-dataset artisanal fishery context and evaluating the trade-off between predictive performance and model interpretability. This framework provides practical insights into the applicability of machine learning methods for data-limited fisheries where extensive monitoring data are unavailable. It was hypothesized that the three machine learning models would differ in predictive performance under data limited conditions, and that model performance would vary between squid catch-size and catch-quantity classification tasks. Therefore, this study presents a small-dataset ML benchmarking approach to evaluate the predictive performance of three supervised learning models in classifying squid catch size and quantity under environmental variability. Using field data collected from artisanal fishers in Prachuap Khiri Khan province, Thailand, three ML models—Decision Tree (DT), k-Nearest Neighbors (k-NN), and Logistic Regression (LR)—were comparatively assessed. The objectives of this study were to (1) evaluate the performance of three ML models under limited data conditions, (2) identify key environmental drivers influencing squid catch characteristics, and (3) provide a practical reference for applying ML techniques in small-scale fisheries where data availability is constrained.

2. Materials and methods

2.1 Fisheries and environmental data resources

This study utilized a structured field dataset collected from artisanal squid (*Loligo* spp.) fisheries operating with large cast nets in the coastal area of Khlong Wan subdistrict, Prachuap Khiri Khan province, Thailand (Fig. 1). Data collection was conducted over a 12-month period from October 2024 to September 2025, covering seasonal environmental variability.

Given the nature of small-scale fisheries, where data availability is typically limited and fragmented, this study was designed to reflect a low-data scenario representative of artisanal fisheries systems. Squid catch data were obtained from four local fishing vessels that were consistently monitored throughout the study period. Although the number of participating fishers was limited, high-frequency observations (daily or per fishing trip records) were collected, resulting in a temporally structured dataset suitable for machine learning analysis under data-constrained conditions.

The dataset included three main components. First, catch quantity (kg per fishing operation) was recorded for each fishing trip. Second, squid size (mantle length) was categorized into two classes: small (≤10 cm) and large (>10 cm), based on local market classifications, to define a binary classification task for catch-size prediction. Third, catch quantity was further grouped into three categories using K-means clustering: low (≤9.4 kg), medium (>9.4–<15.9 kg), and high (≥15.9 kg), to define a three-class classification task for catch-quantity prediction. The resulting two classification tasks were used as separate target variables for benchmarking the three machine learning models.

Environmental variables were selected based on their established relevance to squid fisheries (Arkronrat and Boutson, 2013; Arkronrat et al., 2017; Leearam et al., 2025), including sea surface temperature (SST, °C), sea surface salinity (SSS, ppt), and dissolved oxygen (DO, mg/L). These parameters were measured at four sampling stations (S1–S4) within the fishing area. SST and DO were recorded using an oxygen probe (Pro20i; YSI, USA), while SSS was measured using a refractometer (Master-S10 alpha; Atago, Japan). Water sampling was conducted biweekly, and monthly averages were calculated to align with catch data. In addition, air temperature data were obtained from the WeatherSpark geographic information system database (https://th.weatherspark.com), corresponding to the study area and period. All datasets were compiled, synchronized, and organized into a structured database for subsequent analysis.

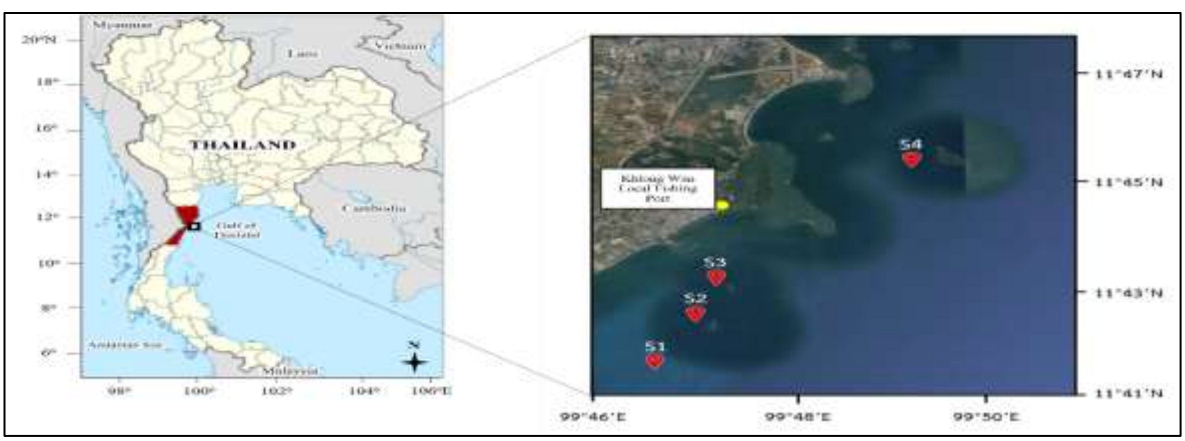


Fig. 1. Squid fisheries area based on large cast net gear (11°41'–11°48' N, 99°47'–99°50' E) in Prachuap Khiri Khan province, Thailand and water sampling points (S1–S4) for environmental data collection.

The dataset comprised 278 fishing-trip observations collected from four artisanal fishing vessels between October 2024 and September 2025. In the study area, squid fishing vessels typically operate 8–14 trips per month, with each trip consisting of a single overnight fishing operation. Accordingly, each observation represented one fishing trip as the unit of analysis, with a corresponding catch record and associated environmental conditions.

2.2 Data preprocessing

Data preprocessing was conducted to ensure data quality and suitability for machine learning under a low-data setting. Data quality checks were performed, and missing values were handled using mean imputation based on the training data or by excluding incomplete records where necessary. Catch-size and catch-quantity variables were defined as classification targets based on the predefined categories described above. Feature scaling was applied using standardization to ensure comparable feature ranges, particularly for k-NN and LR models. The final dataset (n = 278 observations) was randomly divided into training (80%, n = 222) and testing (20%, n = 56) subsets, and 10-fold cross-validation was implemented with the training set to improve robustness and reduce overfitting. Preprocessing procedures, including imputation and feature scaling, were performed using the training data within each cross-validation fold to minimize potential data leakage (Faisal et al., 2016).

2.3 Model building and training

Three supervised machine learning algorithms were evaluated, including Decision Tree (DT), k-Nearest Neighbors (k-NN), and Logistic Regression (LR). The selected models represent complementary learning strategies commonly applied to classification problems, including tree-based, instance-based, and regression-based approaches (Hou, 2018; Suyal and Goyal, 2022; Abebe, 2024).

Model development was conducted using Altair AI Studio (RapidMiner, Altair® AI Studio Trial 2025.1.0). Model hyperparameters were optimized using grid search within the training dataset, with 10-fold cross-validation used to identify the best-performing parameter configurations.

The DT model was constructed using the gain ratio splitting criterion, with the final optimized configuration consisting of a maximum tree depth of 10, minimum leaf size of 2, minimum split size of 4, and minimum gain threshold of 0.01. Both pre-pruning and post-pruning procedures were applied (confidence level = 0.1; number of pruning alternatives = 3) to improve model generalization. The k-NN model was configured with the optimized value of k = 5 nearest neighbors using a mixed Euclidean distance metric, with standardized input variables to account for differences in feature scales. LR was implemented using the default software settings with standardized input variables.

Given the relatively small dataset (n = 278 observations), hyperparameter optimization were conducted within the training set using 10-fold cross-validation to reduce the risk of overfitting. The independent hold-out test set (20%, n = 56) was retained separately from the training set and was not included in the cross-validation procedure. Model performance reported in this study was evaluated using pooled out-of-fold predictions from the 10-fold cross-validation.

2.4 Model evaluation

Following model training, model robustness and predictive performance were evaluated using 10-fold cross-validation on the training set to reduce overfitting. Predictions from the ten validation folds were pooled to generate out-of-fold predictions for performance evaluation. The models were evaluated for their ability to classify squid catch size and quantity based on environmental variables, and predicted outputs were compared with observed values.

Model performance was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score (Faisal et al., 2016). These metrics were calculated from the pooled out-of-fold predictions generated during the 10-fold cross-validation. Accuracy measures overall correctness, precision reflects the reliability of positive predictions, recall indicates the ability to correctly identify actual positive instances, and F1-score represents the harmonic mean of precision and recall, providing a balanced measure of model performance. These metrics were calculated using Equations (1)–(4):

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

$$\text{Recall} = TP / (TP + FN) \quad (3)$$

$$\text{F1-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

where TP (True Positive) denotes correctly predicted target-class instances, TN (True Negative) represents correctly predicted non-target instances, FP (False Positive) indicates non-target instances incorrectly predicted as target, and FN (False Negative) refers to target instances incorrectly predicted as non-target.

In addition, a confusion matrix was constructed from the pooled out-of-fold predictions to provide detailed insight into classification performance, including the distribution of correctly and incorrectly classified observations. The overall processing workflow and optimized model structures for classifying squid catch size and quantity are illustrated in Fig. 2.

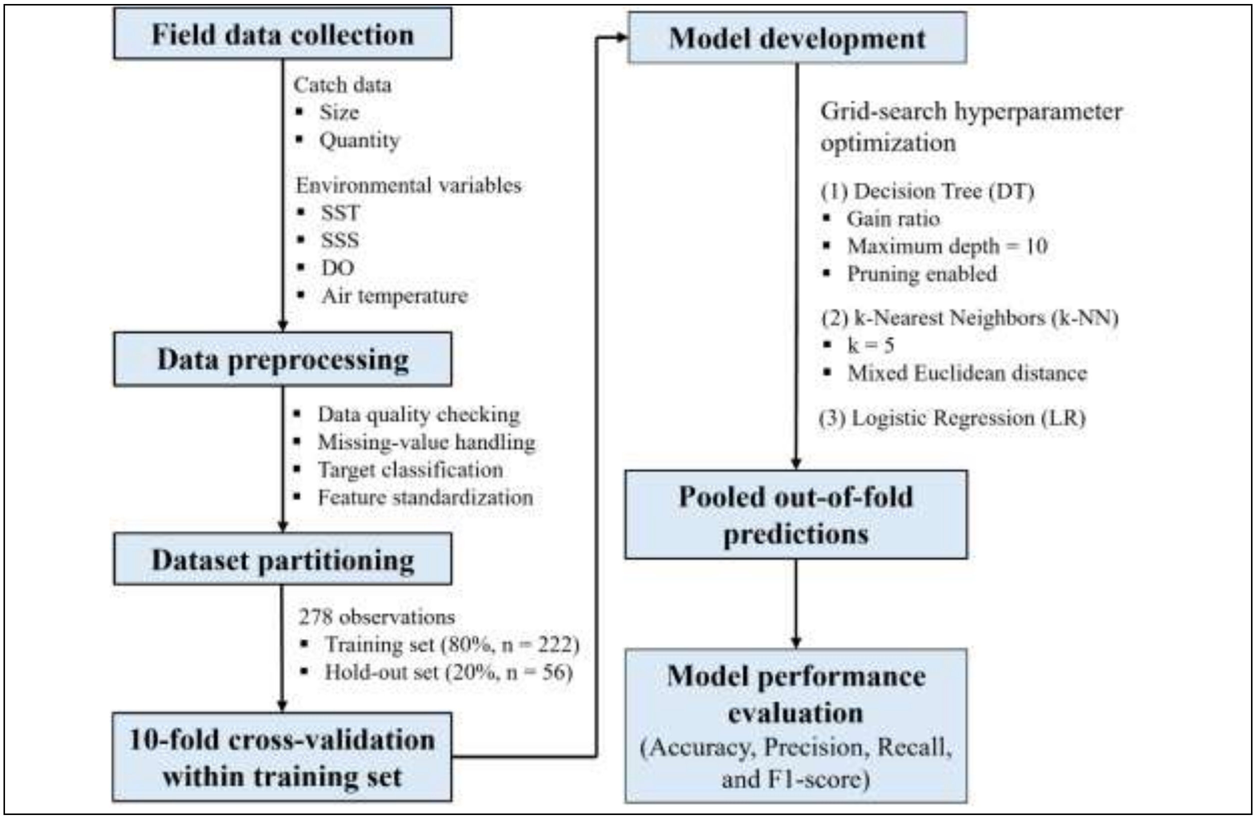


Fig. 2. Workflow of the machine learning framework for classifying *Loligo* spp. catch size and quantity under data-limited artisanal fishery conditions.

2.5 Data analysis

Altair AI Studio (RapidMiner, Altair® AI Studio Trial 2025.1.0) was used for K-means clustering and machine learning model development, including DT, k-NN, and LR. The IBM SPSS Statistics for Windows software (version 24.0; IBM Corp.; Armonk, NY, USA) was used for statistical analyses. Seasonal differences in air temperature, SST, SSS, DO, and squid catch variables among winter, summer, and rainy seasons were evaluated using one-way ANOVA. When significant differences were detected, Duncan's multiple range test was used for multiple comparisons at the 95% confidence level ($P<0.05$). Results are presented as mean ± standard deviation (SD).

3. Results

The average maximum air temperature in Prachuap Khiri Khan province from October 2024 to September 2025 ranged from 29.2 to 33.0 °C, while sea surface temperature (SST) in the fishing grounds ranged from 24.4 to 31.1 °C (Table 1). Seasonal analysis indicated that SST did not differ significantly between the summer and rainy seasons, but was significantly higher in both seasons than in winter ($P<0.05$). Significant seasonal variation was also detected in SSS ($P<0.05$), whereas DO remained relatively constant across seasons ($P > 0.05$) (Table 2).

Table 1 Average air temperature, sea surface temperature (SST), sea surface salinity (SSS), and dissolved oxygen (DO) in the *Loligo* spp. squid fishery area of local fishermen in Prachuap Khiri Khan province, Thailand.

Months	Average air temperature (°C)			SST (°C)	SSS (ppt)	DO (mg/L)
	Lowest monthly	Highest monthly	Mean monthly			
October 2024	25.1	31.1	28.1	26.1	33.0	3.8
November 2024	24.8	31.0	27.9	25.8	33.0	4.0
December 2024	24.2	29.2	26.7	24.4	34.0	4.1
January 2025	23.1	29.5	26.3	24.4	35.0	3.9
February 2025	23.1	30.8	26.9	27.3	34.0	3.9
March 2025	25.4	32.3	28.8	28.6	34.5	4.2
April 2025	26.1	32.8	29.4	31.1	35.0	4.1
May 2025	26.4	33.0	29.4	29.3	34.5	4.6
June 2025	26.1	33.0	29.6	28.9	33.2	4.5
July 2025	26.4	33.3	29.9	29.7	32.2	4.2

August 2025	26.4	32.8	29.6	29.4	32.0	3.9
September 2025	26.5	32.8	29.6	29.4	32.0	4.5

Note: Air temperature data sourced from the WeatherSpark geographic information system database (©WeatherSpark.com) for Prachuap Khiri Khan province.

Table 2 Seasonal mean values of selected water quality parameters at *Loligo* spp. squid fishing ground in Prachuap Khiri Khan province, Thailand.

Parameters	Winter season	Summer season	Rainy season	*P*-value
SST (°C)	25.4 ± 1.7^{b}	29.8 ± 1.3^{a}	29.7 ± 0.4^{a}	0.009
SSS (ppt)	34.3 ± 0.6^{a}	34.7 ± 0.3^{a}	32.5 ± 0.7^{d}	0.005
DO (mg/L)	4.9 ± 1.1^{a}	5.2 ± 2.1^{a}	5.1 ± 1.8^{a}	0.607

Note: Means with different lowercase superscripts (a, b) in same rows are significantly differences at $p < 0.05$ level (95% confidence).

Based on the model performance metrics summarized in Table 3, the DT model achieved the highest performance for squid catch size classification. For size classification (≤10 cm vs. >10 cm), the DT model achieved the highest performance, with an accuracy of 89.04% and the highest F1-score (87.20%), indicating a well-balanced trade-off between precision and recall. The LR model also showed relatively high performance, with an accuracy of 85.64% and an F1-score of 82.00%, whereas the k-NN model performed substantially worse, with the lowest accuracy (72.25%) and F1-score (71.80%).

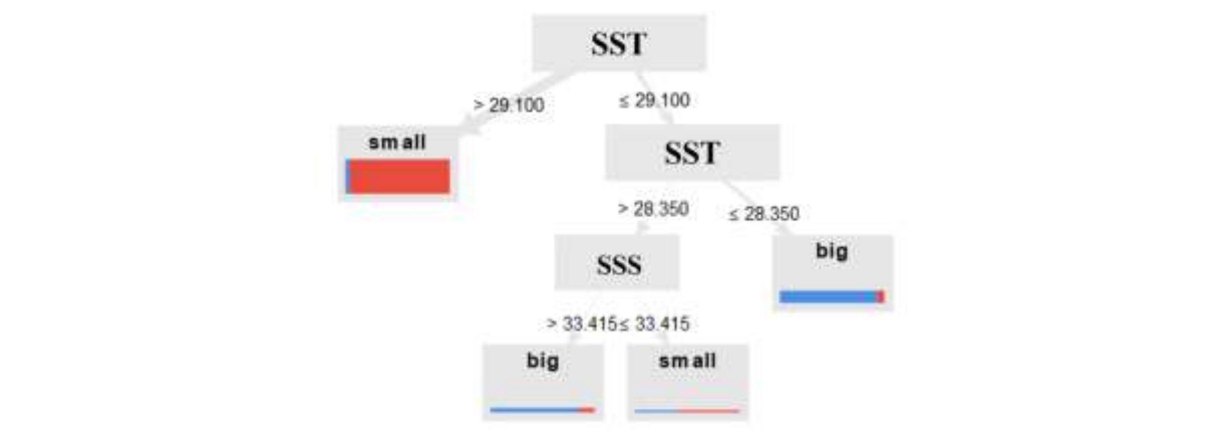
For catch quantity classification, overall model performance declined across all algorithms, reflecting the increased complexity of predicting multi-class outcomes. Nevertheless, the DT model remained the best-performing approach, achieving the highest accuracy (57.14%) and F1-score (53.80%). The k-NN and LR models showed lower performance, with accuracies of 51.87% and 45.66%, respectively, and F1-scores of 49.30% and 27.00%, respectively. Overall, the DT model performed best across both classification tasks, particularly for catch-size classification. Meanwhile, the independent hold-out test set (20%, n = 56) was excluded from the cross-validation performance metrics reported in Table 3.

Table 3. **Performance comparison of machine learning models for predicting *Loligo* spp. catch size and quantity under data-limited conditions. Metrics include accuracy, precision, recall, and F1-score (macro-averaged across classes).**

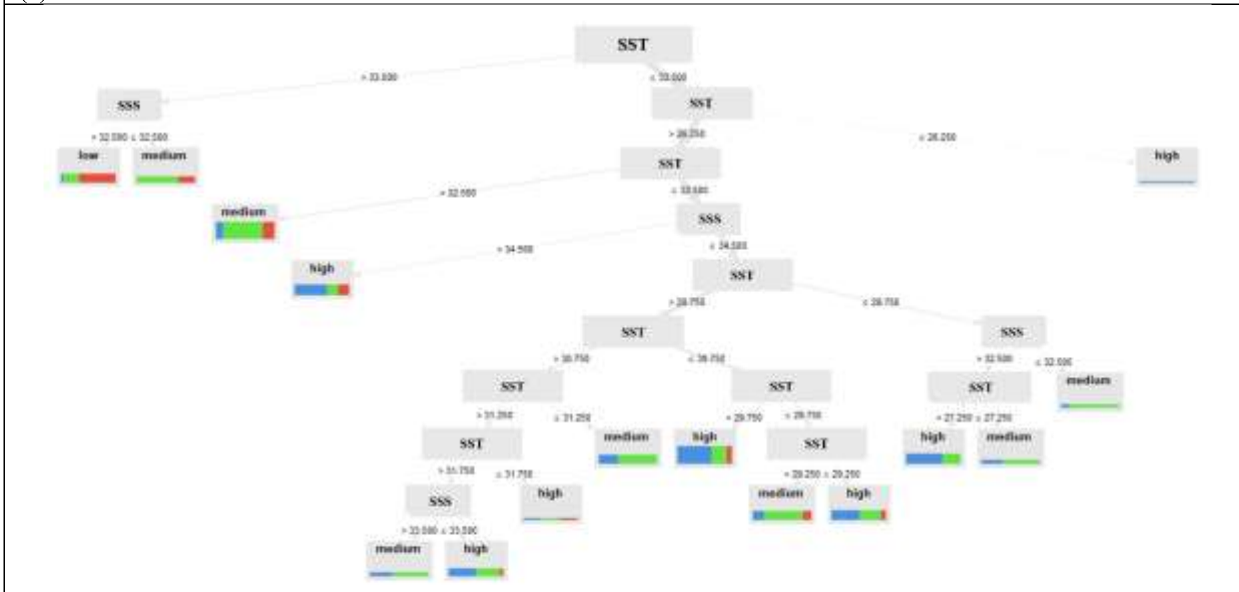
Parameters	Model	Model performance (%)			
		Accuracy	Precision	Recall	F1-score
Catch size	DT	89.04	88.03	86.51	87.20
	k-NN	72.25	73.87	77.18	71.80
	LR	85.64	86.69	79.80	82.00
Catch quantity	DT	57.14	59.46	52.82	53.80
	k-NN	51.87	49.32	50.06	49.30
	LR	45.66	32.85	34.04	27.00

Note: DT = Decision Tree, k-NN = k-Nearest Neighbors, and LR = Logistic Regression.

Decision Tree analysis further revealed interpretable decision patterns relating environmental conditions to squid catch size. When average temperature exceeded 29.10 °C, the model classified most catches as ≤10 cm. At intermediate temperatures (28.35–29.10 °C), SSS was an important splitting variable, with values >33.41 ppt with the >10 cm class, whereas lower salinity corresponded to the ≤10 cm class. When temperature was below 28.35 °C, the model classified most catches as >10 cm (Fig. 3a). In contrast, the DT model achieved a relatively low accuracy of 57.14% for catch-quantity classification (Fig. 3b). Thus, catch-quantity classification showed lower predictive performance than catch-size classification under the present data conditions.



(a)



(b)

Fig. 3. Decision Tree (DT) models for classifying *Loligo* spp. catch characteristics under environmental variability.

(a) DT model for catch-size classification showing the hierarchical decision rules based on sea surface temperature (SST) and sea surface salinity (SSS). (b) DT model for catch-quantity classification showing the decision structure generated from the environmental variables included in the analysis.

4. Discussion

Based on the results, the Decision Tree (DT) model was the most suitable approach for classifying *Loligo* spp. catch characteristics, particularly for squid size classification. However, for catch quantity prediction, although DT outperformed k-NN and LR, its accuracy remained relatively low (57.14%), indicating limited reliability under the current data conditions. Therefore, the DT model appears more suitable for exploratory classification of catch size than for reliable prediction of catch quantity under the present low-data conditions.

The application of machine learning (ML) in fisheries has demonstrated potential to improve productivity, reduce operational costs, and support sustainable resource management, while facilitating faster analysis and decision-making processes (Gladju et al., 2022; Panda and Baral, 2023; Kühn et al., 2025). In this study, the DT model consistently achieved the highest performance among the three tested models, indicating its potential suitability for small-scale, data-limited fisheries systems. These findings highlight the practical value of interpretable ML approaches for supporting decision-making among artisanal fishers under conditions of environmental variability.

The DT model also offers an advantage in interpretability over the k-NN and LR models because it provides transparent decision rules that facilitate the identification of environmental conditions associated with different catch outcomes (Hou, 2018). Previous studies have similarly demonstrated the effectiveness of DT-based approaches in fisheries applications, including fishing ground prediction under complex and incomplete data conditions (Zhu et al., 2017), adaptive fisheries management using multi-indicator frameworks (Harford et al., 2019), and species classification from image data when combined with feature extraction techniques (Rahma and Citra, 2025).

An important finding of this study was that SSS emerged as an important decision variable in the DT model for squid catch-size classification. Under intermediate SST conditions (28.35–29.10 °C), the DT model used an SSS threshold of 33.41 ppt to distinguish between catch-size classes, suggesting an association between salinity conditions and squid size classification. Although this result should be interpreted as an observed association rather than a causal relationship, it suggests that salinity variability may be an important environmental correlate of *Loligo* spp. catch characteristics in the study area. Salinity is recognized as a key factor influencing cephalopod habitat suitability, physiological processes, and prey availability, which may subsequently affect squid distribution and fishery performance. Similar observations have been reported in other squid fisheries, where SSS contributed to variations in catch distribution and abundance (Wang et al., 2021; Zhang et al., 2022). This finding is consistent with previous studies demonstrating that environmental variability can substantially affect squid distribution and fishery success. For example, increased SST has been associated with reduced squid catches and shifts in fishing grounds in the Pacific Northwest (Guerreiro et al., 2023). Likewise, the average maximum air temperature observed during the present study was approximately 0.6 °C higher than values reported a decade earlier for the same region (Arkronrat et al., 2017), suggesting that ongoing environmental change may contribute to shifts in squid distribution. Such changes may encourage adult squid to occupy cooler or deeper waters for growth and reproduction, thereby altering fishing-ground dynamics and potentially reducing catch efficiency, as previously reported by Ani and Robson (2021) and Hazri and Noor (2024).

Although temperature is a key driver, squid fisheries are influenced by multiple interacting factors, including lunar cycles, tides, and plankton availability, which affect habitat conditions and catch rates (Arkronrat and Boutson, 2013; Arkronrat et al., 2017). Previous studies have shown that environmental and operational factors jointly explain substantial variability in squid catches. For example, Wang et al. (2021) reported that 63.1% of catch variability in miter squid *Uroteuthis chinensis* was explained by SST, SSS, depth, fishing strategy, and season, while Zhang et al. (2022) and Fei et al. (2022) highlighted the importance of subsurface temperature and chlorophyll *a* concentration in determining squid abundance and distribution.

Although the dataset included 278 fishing-trip observations, these records were derived from repeated operations of four vessels within a single fishing area. Therefore, vessel-specific effects and temporal dependence cannot be excluded and may have reduced the effective sample size. The findings should consequently be interpreted as representative of the study fishery rather than broadly generalizable across all artisanal squid fisheries. In addition, catch-quantity classes were generated using K-means clustering and therefore represent data-driven categories rather than formally established fishery grading standards. Unequal class distributions may have introduced class imbalance, which may have contributed to the lower predictive performance of catch-quantity classification. Future studies should evaluate alternative classification approaches and larger multi-year datasets to improve model robustness.

Despite these limitations, the DT model provided the highest predictive performance among the evaluated algorithms and offers a useful, interpretable framework for exploring environmental associations with *Loligo* spp. fisheries under data-limited conditions. Nevertheless, the present model should be regarded as a preliminary exploratory model rather than an operational forecasting tool. Further validation using independent datasets from additional fishing grounds and years is required before broader application. Integration into user-friendly platforms, such as application programming interfaces (APIs) or web-based systems, could be considered in future development to support practical decision-making by local fishers.

5. Conclusion

This study applied three machine learning approaches—Decision Tree, k-Nearest Neighbors, and Logistic Regression—to evaluate the effects of temperature variability on *Loligo* spp. squid catch size and quantity in an artisanal fishery. Among the tested models, the Decision Tree (DT) demonstrated the best performance for size classification, achieving 89.04% accuracy, whereas performance for catch quantity classification remained moderate (57.14%). These results indicate that DT may be suitable for classification tasks under data-limited conditions, particularly when interpretability is required. The findings highlight the potential of machine learning as a practical tool for supporting small-scale fisheries, especially in contexts where data availability is limited and environmental variability is increasing. However, the relatively low accuracy for catch quantity prediction suggests that model outputs should be interpreted with caution. Overall, the DT model demonstrated the strongest performance among the evaluated algorithms and may serve as a preliminary prototype for low-data artisanal fishery applications. However, additional validation using larger and more diverse datasets is required before operational deployment.

Authorship contribution statement

Wasana Arkronrat: Project administration, Conceptualization, Methodology, Investigation, and Writing original draft. **Chonlada Leearam:** Software, Data curation, Formal analysis, and Investigation. **Rungtiwa Konsantad:** Data curation, Formal analysis, and Investigation. **Vutthichai Oniam:** Software, Validation, Writing - review and editing. All authors contributed to the interpretation of the results and approved the final manuscript.

Declaration of Generative AI and AI-assisted technologies in the writing process
During the preparation of this manuscript, the authors used Gemini (Google, model: Gemini 3.1 Pro) to improve language and readability. After using this tool, the authors carefully reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements
This study was a part of the Project for the Squids Resource Utilization to Local Economic Upgrading of Small-Scale Fishers' Klongwan Community by BCG Economy Model, supported by the Kasetsart University Research and Development Institute (KURDI), Bangkok, Thailand under research grant No. FF(KU)45.68. Gratitude is also extended to all esteemed committee members from KURDI for their invaluable guidance and constructive suggestions, which significantly contributed to the successful completion of this study.

References

1. Abebe, T.H., 2024. Regression analysis: a theoretical approach. J. Stat. Econom. Methods. 13 (1), 1–40. https://doi.org/10.47260/jsem/1311.
2. Ani, C.J., Robson, B., 2021. Responses of marine ecosystems to climate change impacts and their treatment in biogeochemical ecosystem models. Mar. Pollut. Bull. 166, 112223. https://doi.org/10.1016/j.marpolbul.2021.112223.
3. Arkronrat, W., Boutson, A., 2013. Squid falling net (large cast net) fishing and the catches in the coastal area of Klongwan, Prachuap Khiri Khan province, In: The 3rd International Fisheries Symposium (IFS 2013). Kasetsart University, Bangkok, p. 53.
4. Arkronrat, W., Boutson, A., Tunkijjanukij, S., 2017. Small-scale squid large cast-net fisheries during waxing and waning moon phases in the Klongwan coastal area, Prachuap Khiri Khan province, Thailand. J. Fish. Environ. 41 (2), 37–49.
5. Faisal, A., Jhanjhi, N.Z., Ashraf, H., Ray, S.K., Ashfaq, F., 2016. A comprehensive review of machine learning models: Principles, applications, and optimal model selection. IEEE Trans. J. 4, 1–29.
6. FAO. 2025. Fishery and Aquaculture Statistics– Yearbook 2023. FAO Yearbook of Fishery and Aquaculture Statistics. Rome
7. Fei, Y., Yang, S., Fan, W., Shi, H., Zhang, H., Yuan, S., 2022. Relationship between the spatial and temporal distribution of squid-jigging vessels operations and marine environment in the North Pacific Ocean. J. Mar. Sci. Eng. 10, 550. https://doi.org/10.3390/jmse10040550.
8. Gladju, J., Kamalam, B.S., Kanagaraj, A., 2022. Applications of data mining and machine learning framework in aquaculture and fisheries: A review. Smart Agric. Technol. 2 (1), 100061. https://doi.org/10.1016/j.atech.2022.100061.
9. Guerreiro, M.F., Borges, F.O., Santos, C.P., Xavier, J.C., Hoving, H.J., Rosa R., 2023. Impact of climate change on the distribution and habitat suitability of the world's main commercial squids. Mar. Biol. 170, 129. https://doi.org/10.1007/s00227-023-04261-w.
10. Harford, W.J., Dowling, N.A., Prince, J.D., Hurd, F., Bellquist, L., Likins, J., Wilson, J.R., Wilson, J.R., 2019. An indicator-based decision framework for the northern California red abalone fishery. Ecosphere. 10 (1), e02533. https://doi.org/10.1002/ecs2.2533.
11. Hazri, M.I.N.M., Noor, N.M., 2024. Fish diversity assessment in the Malaysian straits: Insights into climate change resilience. Aquac. Res. 2024, 3913221. https://doi.org/10.1155/2024/3913221.
12. Hou, Y., 2018. Decision tree algorithm for big data analysis. Adv. Intell. Syst. Res. 161, 270–274. https://doi.org/10.2991/tlicsc-18.2018.43.
13. Jones, M.D., Hutcheson, S., Camba, J.D., 2021. Past, present, and future barriers to digital transformation in manufacturing: A review. J. Manuf. Syst. 60 (9), 936–948. https://doi.org/10.1016/j.jmsy.2021.03.006.
14. Konsantad, R., Arkrontat, W., 2025. Structure and components of large cast net fishing gear: A case study of large cast net fishing practices by small-scale fishermen in Klongwan subdistrict, Prachuap Khiri Khan province, in: The 8th Marine Science Conference. Rajamangala University of Technology Srivijaya Trang Campus, Trang, pp. 647–648.
15. Kühn, B., Cayetano, A., Fincham, J.I., Moustahfid, H., Sokolova, M., Trifonova, N., Watson, J.T., Fernades-Salvador, J.A., Uusitalo, L., 2025. Machine learning applications for fisheries—at scales from genomics to ecosystems. Rev. Fish. Sci. Aquac. 33 (2), 334–357. https://doi.org/10.1080/23308249.2024.2423189.
16. Leearam, C., Konsantad, R., Arkronrat, W., 2025. Current Situation of Light Luring Squid Fishing with Large Cast Net by Small-Scale Fishers in Klongwan Subdistrict, Prachuap Khiri Khan Province, in: The 8th Marine Science Conference. Rajamangala University of Technology Srivijaya Trang Campus, Trang, pp. 645–646.
17. Nasution, U.H., Syamsurizal. Zahri, C., 2020. Analysis of supply chain management of squid sea fish in traditional fisherman in village of fisherman Indah Medan. Int. J. Educ. Hum. Soc. Sci. 1 (2), 121–129. https://doi.org/10.46576/english.v1i2.637.

18. Panda, P.K., Baral, D., 2023. Adoption of AI/ML in aquaculture: a study on pisciculture. J. Surv. Fish. Sci. 10 (1), 228–233.
19. Rahma, Y., Citra, P., 2025. Fish genus classification system using Viola-Jones feature extraction and boosting based on decision tree. J. Ilmu Komput. Apl. 8 (1), 18–26. https://doi.org/10.21009/j-koma.v8i1.03.
20. Rahman, M.A., Kanon, K.F., Islam, M.Z., Mojumdar, S., Ashik, A.-A., Molla, M.H.R., 2022. Impacts of climate change on aquaculture and fisheries: An integrated approach for adaptation and mitigation. J. Biol. Stud. 5 (2), 10–27.
21. Song, Y., Zhang, S., Tang, F., Shi, Y., Wu, Y., He, J., Chen, Y., Li, L. 2023. Behavior recognition of squid jigger based on deep learning. Fishes. 8, 502. https://doi.org/10.3390/ fishes8100502
22. Sukhsangchan, C., Phuynoi, S., Monthum, Y., Whanpetch, N., Kulanujaree, N., 2020. Catch composition and estimated economic impacts of ghost-fishing squid traps near Suan Son Beach, Rayong province, Thailand. Sci. Asia. 46, 87–92. https://doi.org/10.2306/scienceasia1513-1874.2020.014.
23. Suryanto, Oktaviani, D., Nugroho, D., Anggawangsa, R.F., 2021. Fleet diversity of squid fisheries in Indonesia Fisheries Management Area – 711. IOP Conf. Ser.: Earth Environ. Sci. 800, 012007. https://doi.org/10.1088/1755-1315/800/1/012007.
24. Suyal, M., Goyal, P., 2022. A review on analysis of K-Nearest Neighbor classification machine learning algorithms based on supervised learning. Int. J. Eng. Trends Technol. 70 (7), 43–48. https://doi.org/10.14445/22315381/IJETT-V70I7P205.
25. Wang, D., Yao, L., Yu, J., Chen, P., 2021. The role of environmental factors on the fishery catch of the squid *Uroteuthis chinensis* in the Pearl River Estuary, China. J. Mar. Sci. Eng. 9, 131. https://doi.org/10.3390/jmse9020131.
26. Zhang, Y., Yu, W., Chen, X., Zhou, M., Zhang, C., 2022. Evaluating the impacts of mesoscale eddies on abundance and distribution of neon flying squid in the Northwest Pacific Ocean. Front. Mar. Sci. 9, 862273. https://doi.org/10.3389/fmars.2022.862273.
27. Zhu, J., Xu, J., Zhang, C., Gao, Y., 2017. Marine fishing ground prediction based on Bayesian decision tree model, in: Proceeding of the 2017 International Conference on Management Engineering, Software Engineering and Service Science. Wuhan University, Wuhan, pp. 316–320.