**Bulletin of Electrical Engineering and Informatics**
Vol. 15, No. 2, April 2026, pp. 1337~1346
ISSN: 2302-9285, DOI: 10.11591/eei.v15i2.9935 

# Comparative analysis of ensemble learning algorithms in enhanced confidence-based assessments

**Nur Maisarah Nor Azharludin[1], Khyrina Airin Fariza Abu Samah[1], Mohamad Faiz Dzulkalnine[1], Ahmad Firdaus Ahmad Fadzil[1], Mohd Nor Hajar Hasrol Jono[1], Lala Septem Riza[2]**
[1]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Melaka Branch, Jasin Campus, Melaka, Malaysia
[2]Department of Computer Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

**Article Info**

***Article history:***

Received Jan 23, 2025
Revised Dec 29, 2025
Accepted Feb 22, 2026

***Keywords:***

Bias mitigation
Confidence-based assessment
Ensemble learning
Predictive performance
Student evaluation accuracy

**ABSTRACT**

This paper provides a comparative analysis of ensemble learning (EL) algorithms to enhance the confidence-based assessment (CBA) in evaluating student performance. Traditional CBA often suffers from misclassification caused by overconfidence and underconfidence, limiting its accuracy and fairness. To address these challenges, an enhanced CBA-EL model integrating bagging and boosting ensemble algorithms is proposed. Five bagging algorithms, which are random forest (RF), decision tree (DT), support vector machine (SVM), K-nearest neighbors (KNN), Naïve Bayes (NB), and four boosting algorithms, which are adaptive boosting (AdaBoost), eXtreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), and categorical boosting (CatBoost), were evaluated using a dataset of 276 responses collected from Pre- and Post-Quiz CBA in a discrete structures course. Algorithm performance was evaluated using accuracy, correlation, weighted mean precision (WMP), and weighted mean recall (WMR). RF achieved 73.19% accuracy, 0.725 correlation, 0.751 WMP, and 0.766 WMR, while CatBoost outperformed all with 86.23% accuracy and the highest correlation, WMP, and WMR values, with 0.842, 0.843, and 0.862, respectively. The findings indicate that integrating EL into CBA improves prediction accuracy and supports bias-aware student evaluation. This research advances reliable assessment practices and informs the development of adaptive learning systems.



***Corresponding Author:***

Khyrina Airin Fariza Abu Samah
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Melaka Branch
Jasin Campus
Merlimau, 77300, Melaka, Malaysia
Email: khyrina783@uitm.edu.my

## 1. INTRODUCTION

Confidence varies across individuals and is influenced by cognitive and emotional factors, which shape how learners judge their own performance [1], [2]. Yadav [3] stated that those with elevated self-esteem and self-efficacy are more likely to possess increased confidence in their capabilities. An alternative method for addressing the confidence inquiry involves employing a general knowledge questionnaire that utilizes ‘yes’ or ‘no’ responses to factual statements [4]. Confidence, shaped by cognitive and emotional factors, is a key determinant of student engagement and academic performance [5], [6]. Confident students are more likely to take on challenging tasks, persevere through difficulties, and achieve better academic results [7]. As such, confidence levels influence metacognitive awareness and may affect the accuracy of assessment outcomes when students misjudge their knowledge.

Traditional assessment methods often fail to capture discrepancies between students' perceived confidence and their actual understanding, leading to inaccurate evaluation of student performance. Confidence-based assessment (CBA), a two-dimensional evaluation approach, is employed to gauge students' confidence or anticipation of their answers, thereby assessing their actual knowledge [8]. One-dimensional assessment, or the traditional assessment approach, relies exclusively on the quantity of correctly answered questions to determine a student's test scores [9]. CBA provides deeper insight into students' cognitive understanding, but its accuracy depends on correctly interpreting the relationship between confidence and actual performance [10], [11]. It has been utilized across numerous domains for diverse objectives such as evaluating educational outcomes, enhancing pedagogical methods, increasing data accuracy, promoting self-regulated learning among students, enriching student learning experiences, delivering constructive feedback to educators, and invigorating student motivation [8], [11]. CBA evaluates learners' information accuracy through two CBA indicators: error analysis, or multiple-choice questions (MCQ), and confidence assessment, which is a question about the student's confidence in their responses. Upon completion, learners are classified into four categories of CBA based on the category knowledge quadrant (CKQ), including uninformed, misinformed, doubt, and mastery, to enhance learning performance corresponding to each category [12]. Multiple studies indicate that CBA offers a substantial primary benefit in enhancing the accuracy of evaluating the perceptions of evaluated participants [13]. Nonetheless, existing CBA models often produce biased or imprecise quadrant classifications because students may exhibit overconfidence or underconfidence in their responses [12]-[14]. These issues can obscure meaningful learning insights and impact academic interventions. Thus, enhancing the CBA model to address prediction bias and improve evaluation accuracy is essential to support more equitable and reliable assessments.

Helm *et al.* [15] stated that machine learning (ML) is an advancing field of computational algorithms that mimic human intelligence by learning from its environmental inputs. Recognized as one of the most prominent technologies of the Fourth Industrial Revolution, ML enables systems to learn and enhance their performance through experience without using detailed manual programming [16]. However, ML algorithms often encounter overfitting issues, particularly when applied to small educational datasets such as those used in CBA. While many single-model ML techniques have been applied to CBA, they are prone to generalization limitations. Ensemble learning (EL) is a preferred approach due to its ability to improve predictive performance and mitigate bias by combining the strengths of multiple learners [17]. While EL is a subset of ML, its methodologies help mitigate overfitting by combining multiple learners to improve generalization. By leveraging multiple models' diverse strengths and complementary capabilities, EL achieves superior overall performance compared to any single model [18]. This advantage positions EL as a more robust approach within ML, offering improved accuracy and generalization. Recent EL applications have shown strong performance in several domains, including educational analytics, where they improve prediction accuracy and support reliable student assessment [19]. There are three EL methods, including bagging, boosting, and stacking. However, this study focuses on bagging and boosting due to their suitability for small datasets and interpretability in educational contexts. In educational datasets, bagging methods such as random forest (RF) are typically more stable and effective at reducing variance when student response data contain noise [20]. In contrast, boosting algorithms like eXtreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), and categorical boosting (CatBoost) often achieve higher predictive accuracy because they iteratively correct misclassified learning patterns [21]. These differences highlight the importance of comparing bagging and boosting approaches specifically for improving CBA quadrant classification. Bagging and boosting algorithms are examined in this study because both methods offer complementary strengths that can enhance the classification accuracy of CBA. In bagging, the algorithms involved in this study are RF, decision tree (DT), support vector machine (SVM), K-nearest neighbors (KNN), and Naïve Bayes (NB), while algorithms for boosting are adaptive boosting (AdaBoost), XGBoost, LightGBM, and CatBoost.

This study proposes a comparative evaluation of EL algorithms to enhance the CBA model by improving predictive accuracy and reducing confidence-related misclassification. The enhancement of the CBA model in assessing student performance according to the specified quadrant using EL is called "enhanced CBA-EL". Thus, this study identifies the ensemble algorithms that best support accurate CKQ classification within the enhanced CBA model. Specifically, it aims to identify algorithms that address the limitations of confidence misclassification and support accurate, bias-aware student evaluation. The structure of this paper begins with a brief introduction in section 1. Section 2 presents the research methodology. Section 3 discusses the experimental results and validation by comparing the proposed method with related study algorithms. Finally, section 4 concludes the study and outlines possible future work.

## 2. METHOD

The methods used to conduct the comparative analysis, which identify the most suitable EL algorithms for the enhanced CBA-EL model, are crucial. Two parts of the methods are involved: i) data preparation and ii) EL algorithms performance in the enhanced CBA-EL. This section explains both methods to ensure we follow the correct steps in determining the suitable EL algorithms for the enhanced CBA-EL model. The procedure is designed to be replicable and logically sequenced, allowing researchers to reproduce the study with similar data and tools.

### 2.1. Data preparation

#### 2.1.1. Data collection

Data preparation involves collecting, integrating, organizing, and structuring data before employing analytical techniques predominantly used in data visualization. The data collected are related to the discrete structure (CSC510) subject. The training dataset is sourced from [22], which has 1,680 entries for each question and confidence level, with 10 questions and 10 corresponding confidence levels. This dataset was used for model pre-training to expose the ensemble algorithms to a broader range of confidence-knowledge patterns. Pre-training allowed the models to initialize with more stable parameter distributions before being fine-tuned on the primary dataset. After pre-training, all algorithms were trained and evaluated exclusively using the responses collected in this study, ensuring that the final performance outcomes reflect only the real CBA data. Testing data was gathered using the developed questionnaire.

The flow of the data collection is illustrated in Figure 1. The data gathered originates from five distinct classes instructed by two lecturers of CSC510. The students involved are Bachelor's Degree in Computer Science students from the College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM) Melaka Branch, Jasin Campus Melaka, Malaysia, who registered for the CSC510 subject. Five classes were involved: CS2513A, CS2303A, CS2303C, CS2533B, and CS2533C. The students answered two CBA questionnaires, Pre-Quiz and Post-Quiz. In total, the dataset consists of 276 student responses, comprising 138 Pre-Quiz entries and 138 Post-Quiz entries. There are 30 students' data for CS2513A, 39 for CS2303A, 23 for CS2303C, 29 for CS2533B, and 17 for CS2533C. To ensure generalizability, student samples were drawn from multiple classes and instructors, although limited to one subject.

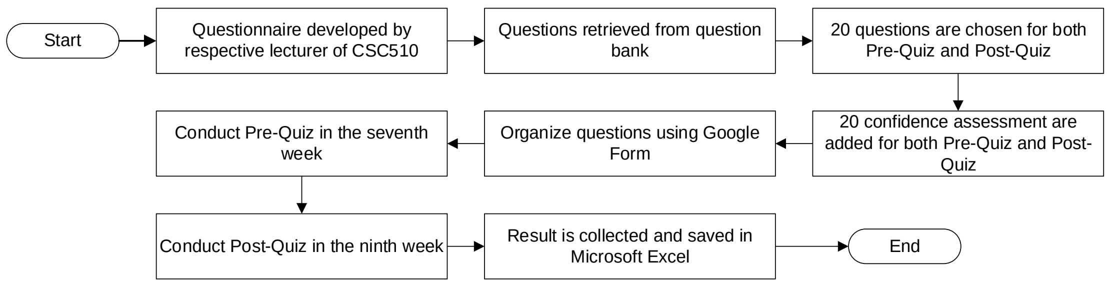


Figure 1. Flowchart of the process involved in data collection process

#### 2.1.2. Data pre-processing

Upon concluding data collection, the gathered data required cleaning and pre-processing to advance the model development. Data cleaning and pre-processing are crucial for ensuring the accuracy and dependability of the data for the development of the enhanced CBA-EL. The data pre-processing is conducted via RapidMiner software because it simplifies the data handling, making it accessible and efficient without extensive programming. Next, data cleaning and pre-processing entail selecting, renaming, mapping, and generating attributes that standardize formats to prepare the data for analysis. The error analysis and confidence assessment dataset comprises "0" and "1" after completing the data cleaning and pre-processing. Pre-processing also involved verifying missing values, validating attribute consistency, and using data normalization where necessary.

Upon data retrieval, the CKQ for each student can be assessed using standard CBA. Table 1 presents the classification of the standard CBA CKQ established for this study. A student is categorized as "Uninformed" in the CBA classification if their responses for both error analysis and confidence evaluation are "0". If a student's response for error analysis is "0" and the confidence assessment is "1", the student is categorized as "Misinformed". If a student's response for error analysis is "1" and their confidence evaluation is "0", the student is categorized as "Doubt". If the student's response for error analysis and confidence evaluation is "1", the student is categorized as "Mastery". The data analysis procedure encompassed data retrieval, attribute generation to ascertain the CKQ for each question, iteration through the attribute data, creation of attributes based on the aggregate data, and storing the final data in the software database. All CKQ classification procedures were implemented using attribute construction operators in RapidMiner Studio version 9.10, which was used for constructing rule-based models and analyzing confidence and error through binary indicators.

Table 1. Standard CBA classification of four CKQ

| Error analysis | Confidence assessment | CBA CKQ |
|---|---|---|
| 0 | 0 | Uninformed |
| 0 | 1 | Misinformed |
| 1 | 0 | Doubt |
| 1 | 1 | Mastery |

Nonetheless, the CBA CKQ of each student may be deceptive due to the potential for overconfidence, underconfidence, and overfitting issues. Consequently, the dataset comprising the indicators of CBA, error analysis, and confidence assessment for each student is provided as input parameters for the enhanced CBA-EL. The data supplied is processed in the enhanced CBA-EL to assess the student's quadrant utilizing the CKQ.

### 2.2. Ensemble learning algorithms performance in enhanced CBA-EL

The data undergo the process involved in the conceptual model, enhanced CBA-EL. It derived from the previous related research that included two CBA indicators, error analysis and confidence assessment, and the method used was EL. Five bagging algorithms and four boosting algorithms are involved in determining the algorithm used in the enhanced CBA-EL. These algorithms were selected because they are widely adopted in educational data mining, demonstrate robustness on small datasets, and are effective in mitigating variance and bias in classification tasks. The EL approach yielding the highest accuracy and performance is selected. This process improves the classification accuracy of student performance. All EL algorithms were implemented in RapidMiner using dedicated modelling operators. Model validation was performed using an 80/20 train–test split with fixed random seed values to ensure reproducibility of the results. Default hyperparameter settings were applied across all algorithms to maintain fair comparison among algorithms and to reduce the risk of overfitting due to the limited dataset size. The performance of the algorithms is compared in terms of four metrics, including accuracy, correlation, weighted mean precision (WMP), and weighted mean recall (WMR). Accuracy denotes the proportion of correctly classified data, commonly known as the percentage of correct predictions [23]. The formula for accuracy is shown in (1):

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

where $TP$ is the number of true positives, $TN$ is the number of true negatives, $FP$ is the number of false positives, and $FN$ is the number of false negatives. The second metric is correlation. Correlation quantifies the strength and direction of a linear relationship between two variables [24]. The formula for correlation is shown in (2):

$$Correlation = \frac{\Sigma(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})}{\sqrt{\Sigma(y_i-\bar{y})^2}\sqrt{\Sigma(\hat{y}_i-\bar{\hat{y}})^2}} \quad (2)$$

where $y_i$ is the actual values, $\hat{y}_i$ is the predicted values, $\bar{y}$ is the actual values' means, and $\bar{\hat{y}}$ is the predicted values' means. The third metric is WMP. WMP is designed to compute the weighted average precision, which evaluates the accuracy of positive predictions while accounting for the importance or weight of each class [25]. The formula for WMP is shown in (3):

$$WMP = \sum_{i=1}^{n} \frac{[(TP_i+FN_i) \cdot \frac{TP_i}{TP_i+FP_i}]}{TP_i+FN_i} \quad (3)$$

where $n$ is the total number of classes, $i$ is the index representing a specific class, $TP$ is the number of true positives, $FN$ is the number of false negatives, and $FP$ is the number of false positives. The last metric is

WMR. WMR is utilized to calculate the weighted average recall, which assesses the model's ability to correctly identify positive instances while considering the weights assigned to each class [25]. The formula for WMR is shown in (4):

$$WMR = \sum_{i=1}^{n} \frac{[(TP_i+FN_i)\cdot\frac{TP_i}{TP_i+FN_i}]}{TP_i+FN_i} \quad (4)$$

where $n$ is the total number of classes, $i$ is the index representing a specific class, $TP$ is the number of true positives, and $FN$ is the number of false negatives. WMP measures the precision of the model by evaluating how many of the predicted positives are correct, adjusted for class importance, while WMR evaluates how well the model captures actual positives while considering class distribution. These metrics are particularly valuable for imbalanced data scenarios.

## 3. RESULTS AND DISCUSSION

This section provides a detailed discussion of the results obtained from the comparison analysis of EL algorithms. The tasks included in this study involve validating by comparing algorithms for bagging and boosting. Multiple factors influence the performance of EL in an enhanced CBA-EL model. The primary metric is the accuracy of each bagging and boosting model. The other metrics involved in validating the performance of bagging and boosting algorithms are correlation, WMP, and WMR.

### 3.1. Performance ensemble learning bagging algorithm

The bagging performance is compared across five algorithms: RF, DT, SVM, KNN, and NB. The results indicate the quantity of true and predicted data utilized within the same dataset. The precision and recall percentages of the class indicate that the quadrant is accurately classified through bagging. Figure 2 compares bagging algorithm accuracy results, specifically RF, DT, SVM, KNN, and NB. The accuracy prediction of data using RF achieved an accuracy of 73.19%, DT with an accuracy of 68.12%, SVM was 31.16%, KNN had an accuracy of 63.77%, and NB yielded an accuracy of 72.46%. As the overall result for bagging accuracy, RF demonstrates the strongest overall accuracy.

The comparative bagging algorithms' performance of RF, DT, SVM, KNN, and NB of the enhanced CBA-EL model is shown in Figure 3. The efficacy of each method is assessed based on its correlation, WMP, and WMR. The correlation indicates that RF yields the greatest value at 0.725, succeeded by NB at 0.710, DT at 0.668, KNN at 0.597, and SVM at 0.000. The WMP results show that RF yields the highest value at 0.751, followed by NB at 0.724, DT at 0.669, KNN at 0.642, and SVM at 0.149, respectively. The last metric of the performance is WMR. NB achieves the highest outcome at 0.774, succeeded by RF, KNN, DT, and SVM, with results of 0.766, 0.691, 0.688, and 0.500, respectively. This indicates that RF provides more consistent CKQ predictions compared to the other bagging models.

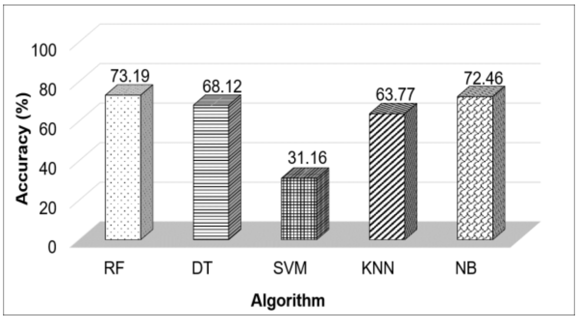


Figure 2. Accuracy comparison for five bagging algorithms

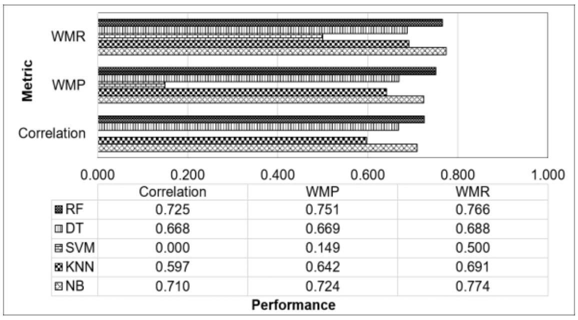


Figure 3. EL bagging algorithms' performance comparison in enhanced CBA-EL model

The results indicate that RF is the most effective EL bagging algorithm. RF has demonstrated good performance as the most suitable algorithm for bagging. RF's ensemble of decorrelated DT helps stabilize performance, making it robust against noise and variance in student responses. Although NB achieved a slightly higher WMR than RF, the overall algorithm performance is evaluated based on overall results, including accuracy, correlation, and WMP. These findings align with prior work [26], which highlights RF's robustness in educational settings due to its ability to reduce overfitting and handle imbalanced data. The following section will discuss the performance of the EL boosting algorithm.

### 3.2. Performance ensemble learning boosting algorithm

The boosting performance is compared between four algorithms: AdaBoost, XGBoost, LightGBM, and CatBoost. The accuracy prediction of data using AdaBoost was 63.04%, 84.06% using XGBoost, 81.16% using LightGBM, and 86.23% using CatBoost in the enhanced CBA-EL model. Among the boosting algorithms evaluated, CatBoost demonstrated the highest accuracy at 86.23%, surpassing AdaBoost, XGBoost, and LightGBM. CatBoost's superior performance is attributed to its ability to natively handle categorical features and prevent overfitting through ordered boosting. This is especially advantageous in educational datasets, which are often small and contain categorical confidence indicators. Therefore, CatBoost is the most appropriate algorithm for boosting. Figure 4 compares the accuracy among the AdaBoost, XGBoost, LightGBM, and CatBoost algorithms.

The comparative boosting algorithms' performance of AdaBoost, XGBoost, LightGBM, and CatBoost of the enhanced CBA-EL model is shown in Figure 5. The correlation indicates that CatBoost achieves the highest result at 0.842, followed by XGBoost at 0.809, LightGBM at 0.759, and AdaBoost at 0.613. The WMP results indicate that CatBoost achieved the highest value at 0.843, followed by XGBoost at 0.819, LightGBM at 0.776, and AdaBoost at 0.601. The final aspect of the performance is WMR. CatBoost achieved the highest result of 0.862, while XGBoost, LightGBM, and AdaBoost followed with results of 0.826, 0.810, and 0.603, respectively. The results indicate that CatBoost is the most effective algorithm for boosting within the enhanced CBA-EL model. This is because CatBoost shows the highest result for all the performance metrics. Compared to other boosting methods, CatBoost also minimizes prediction bias and enhances learning generalization, making it ideal for CBA-EL models. These findings support recent literature [27] indicating CatBoost's effectiveness in small-scale, real-world classification tasks.

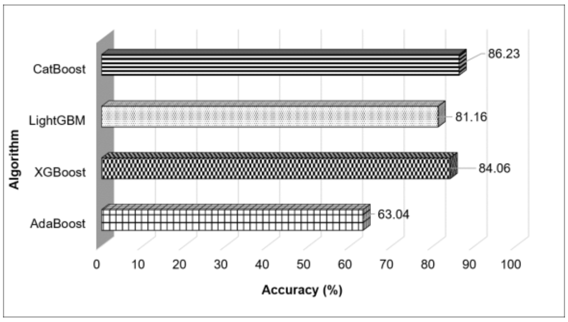


Figure 4. Accuracy comparison for four boosting algorithms

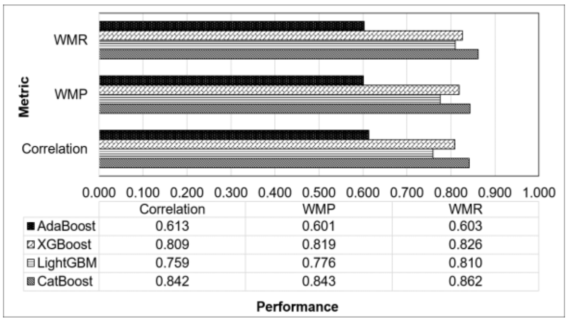


Figure 5. EL boosting algorithms' performance comparison in enhanced CBA-EL model

### 3.3. Discussion of findings and educational implications

The results confirm that both bagging (RF) and boosting (CatBoost) are highly suitable for enhancing CBA models due to their strong predictive performance and bias mitigation capabilities. These findings align with [19], where a stacking ensemble incorporating bagging and boosting variants achieved superior learning achievement predictions on large educational datasets through bias reduction, though the CatBoost model, with 86.23% accuracy, excelled on the smaller CBA dataset. Similarly, a study by [28] demonstrated stacked ensembles' robustness against noise in student performance prediction, supporting RF's stability, with 73.19% accuracy and 0.725 correlation, in bagging results for handling confidence variance. While SVM and AdaBoost underperformed, the other models showed moderate effectiveness. The enhanced CBA-EL model addresses a critical gap in existing CBA methods by offering a more accurate classification of student knowledge-confidence quadrants. These results extend prior EL applications in educational data mining, such as in the study [29] of RF's robustness, by adapting bagging and boosting to CBA-specific quadrants on limited data. Compared to larger-scale studies like [28], the enhanced CBA-EL demonstrates domain-specific gains, informing adaptive interventions. Therefore, educators can use the model's predictions to better tailor instructional interventions, especially for students who are overconfident or underconfident.

### 3.4. Limitations of the study

Despite the promising findings, this study has several limitations that should be acknowledged when interpreting the results. First, the dataset is limited to 276 CBA responses from a single course within one institution. While this dataset was sufficient to support comparative modelling and algorithmic evaluation, its restricted scope may limit the generalizability of the findings. Student confidence and knowledge calibration can vary substantially across disciplines, instructional designs, assessment formats, and institutions. Second, the study focuses solely on four boosting algorithms and five bagging algorithms. Although these algorithms were chosen due to their interpretability and suitability for small datasets, other ensemble methods, such as stacking or hybrid fusion approaches, were not evaluated. Third, the classification performance metrics used in this study may not fully capture long-term learning gains or behavioral changes. Finally, hyperparameters were primarily based on default configurations. Although this approach ensures reproducibility and avoids overfitting in small datasets, it may constrain the achievable performance of certain ensemble algorithms. Thus, future studies may explore optimization techniques to further improve performance.

## 4. CONCLUSION

CBA aims to capture the relationship between students' confidence in their responses and their actual knowledge, yet misalignment between the two can lead to biased and imprecise evaluation. There are three EL methods, which are bagging, boosting, and stacking. However, this study used only two ELs, bagging and boosting, as they are suitable for small datasets and are interpretable in educational settings. This study introduces an enhanced CBA utilizing EL methods to tackle these challenges, aiming to compare EL to determine an effective model for more dependable student performance assessment. There are five algorithms for bagging and four algorithms for boosting. The algorithms chosen are RF for bagging and CatBoost for boosting. The metrics used are accuracy, correlation, WMP, and WMR. The results demonstrate that EL effectively improves CKQ classification accuracy, with RF and CatBoost emerging as the most suitable algorithms for bagging and boosting, respectively, due to their consistent and robust performance across evaluation metrics. The findings indicate that an enhanced CBA-EL model offers a systematic method for educators to make better-informed decisions, revealing a lack of student comprehension that may not be apparent in conventional assessments. The results support the integration of RF and CatBoost into educational tools such as intelligent tutoring systems or adaptive learning platforms, allowing educators to make more reliable data-driven decisions. Although promising, these findings must be validated in larger, more diverse datasets to confirm the model's applicability across different subjects and academic levels. By addressing prediction bias and improving evaluation accuracy, this research contributes to more equitable and informed student assessment practices. Future research should evaluate the enhanced CBA-EL model using cross-institution and cross-disciplinary datasets to further validate its generalizability. Domain adaptation techniques may also be explored to enable the model to transfer effectively across diverse educational environments. In addition, more advanced ensemble techniques, such as stacking ensembles, as well as hybrid architectures that integrate deep learning models with EL, should be investigated to improve feature representation and classification robustness. These directions would support the development of more scalable and pedagogically adaptive CBA systems.

## ACKNOWLEDGMENTS

The authors would like to thank Universiti Teknologi MARA for supporting the research.

## FUNDING INFORMATION

Authors state no funding involved.

## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

| Name of Author | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nur Maisarah Nor Azharludin | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | | | ✓ | |
| Khyrina Airin Fariza Abu Samah | ✓ | ✓ | | ✓ | | ✓ | | | | ✓ | | ✓ | ✓ | ✓ |
| Mohamad Faiz Dzulkalnine | | ✓ | ✓ | | ✓ | | ✓ | | | ✓ | ✓ | | | |
| Ahmad Firdaus Ahmad Fadzil | | ✓ | ✓ | | ✓ | | | | | ✓ | ✓ | | | |
| Mohd Nor Hajar Hasrol Jono | | ✓ | | ✓ | | ✓ | ✓ | | | ✓ | | | | ✓ |
| Lala Septem Riza | | ✓ | | ✓ | | ✓ | | ✓ | | ✓ | | | | |

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

## CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author [KAFAS], upon reasonable request.

## REFERENCES


[1] A. J. Vance, K. Knafl, and D. H. Brandon, "Patterns of parenting confidence among infants with medical complexity," *Adv Neonatal Care*, vol. 21, no. 2, pp. 160–168, 2021, doi: 10.1097/ANC.0000000000000754.

[2] I. Cos, G. Senel, P. E. Maldonado, and R. Moreno-Bote, "The relationship between confidence and gaze-at-nothing oculomotor dynamics during decision-making," *PLoS One*, vol. 20, no. 7, pp. 1–22, 2025, doi: 10.1371/journal.pone.0310231.

[3] A. Yadav, "Investigating the role of self-efficacy in shaping academic and career trajectories of management," *Int. J. Manag.*, vol. 11, no. 10, pp. 226–236, 2020, doi: 10.34218/IJM.11.10.2020.022.

[4] K. Somatori and Y. Kunisato, "Metacognitive ability and the precision of confidence," *Front. Hum. Neurosci.*, vol. 16, pp. 1–16, 2022, doi: 10.3389/fnhum.2022.706538.

[5] E. Acosta-gonzaga, "The effects of self-esteem and academic engagement on university students' performance," *Behav. Sci. (Basel).*, vol. 13, pp. 1–12, 2023, doi: 10.3390/bs13040348.

[6] A. Malureanu, G. Panisoara, and I. Lazar, "The relationship between self-confidence, self-efficacy, grit, usefulness, and ease of use of elearning platforms in corporate training during the Covid-19 pandemic," *Sustain.*, vol. 13, no. 12, pp. 1–20, 2021, doi: 10.3390/su13126633.

[7] A. Adil, S. Ameer, and S. Ghayas, "Impact of academic psychological capital on academic achievement among university undergraduates: Roles of flow and self-handicapping behavior," *PsyCh J.*, vol. 9, no. 1, pp. 56–66, 2020, doi: 10.1002/pchj.318.

[8] N. M. N. Azharludin, K. A. F. A. Samah, M. F. Bin Dzulkalnine, and R. Hamzah, "Exploring enhancement in confidence-based assessment: A systematic literature review," *J. Theor. Appl. Inf. Technol.*, vol. 102, no. 23, pp. 8696–8713, 2024.

[9] R. Maqsood and P. Ceravolo, "Modeling behavioral dynamics in confidence-based assessment," in *IEEE 18th Int. Conf. Adv. Learn. Technol.*, IEEE, 2018, pp. 452–454, doi: 10.1109/ICALT.2018.00112.

[10] Q. Wu *et al.*, "Certainty-based marking on multiple-choice items: Psychometrics meets decision theory," *Psychometrika*, vol. 86, no. 2, pp. 518–543, 2021, doi: 10.1007/s11336-021-09759-0.

[11] S. Caspari-sadeghi, E. Mille, H. Epperlein, and B. Forster-heinlein, "Stimulating reflection through self-assessment: Certainty-based marking (CBM) in online mathematics learning," *Math. Teach. Res. J.*, vol. 14, no. 2, pp. 145–156, 2022.

[12] R. K. Ravikiran and K. R. A. Kumar, "Experimental performance analysis of confidence-based online assessment portal in e-learning using data mining," *Mater. Today Proc.*, vol. 47, no. 17, pp. 5912–5917, 2021, doi: 10.1016/j.matpr.2021.04.456.

[13] X. Zhao, "Evaluation of the modern value of imperial examination culture in the context of cultural confidence based on deep learning models," *Mob. Inf. Syst.*, pp. 1–11, 2022, doi: 10.1155/2022/9748146.

[14] J. D. Stanton, A. J. Sebesta, and J. Dunlosky, "Fostering metacognition to support student learning and performance," *CBE Life Sci. Educ.*, vol. 20, no. 2, pp. 1–7, 2021, doi: 10.1187/cbe.20-12-0289.

[15] J. M. Helm *et al.*, "Machine learning and artificial intelligence: Definitions, applications, and future directions," *Curr. Rev. Musculoskelet. Med.*, vol. 13, pp. 69–76, 2020, doi: 10.1007/s12178-020-09600-8.

[16] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Comput. Sci.*, vol. 2, no. 3, pp. 1–21, 2021, doi: 10.1007/s42979-021-00592-x.

[17] I. D. Mienye, Y. Sun, and S. Member, "A survey of ensemble learning: Concepts, algorithms, applications, and prospects," *IEEE Access*, vol. 10, pp. 99129–99149, 2022, doi: 10.1109/ACCESS.2022.3207287.

[18] R. Theisen, H. Kim, L. Hodgkinson, and M. W. Mahoney, "When are ensembles really effective ?," in *37th Conf. Neural Inf.*

*Process. Syst.*, 2023, pp. 15015–15026.
[19] T. Tong and Z. Li, "Predicting learning achievement using ensemble learning with result explanation," *PLoS One*, vol. 20, no. 1, pp. 1–25, 2025, doi: 10.1371/journal.pone.0312124.
[20] W. Xiao, P. Ji, and J. Hu, "A survey on educational data mining methods used for predicting students ' performance," *Eng. Reports*, vol. 4, no. 5, pp. 1–23, 2022, doi: 10.1002/eng2.12482.
[21] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: An interdisciplinary review," *J. Big Data*, vol. 7, no. 1, pp. 1–45, 2020, doi: 10.1186/s40537-020-00369-8.
[22] S. Recht, V. de Gardelle, and P. Mamassian, "Confidence database," Center for Open Science. [Online]. Available: https://osf.io/s46pr/. (Accessed: Dec. 01, 2023).
[23] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: An overview," *arXiv*, pp. 1–17, 2020, doi: 10.48550/arXiv.2008.05756.
[24] S. Pandey, "Principles of correlation and regression analysis," *J. Pract. Cardiovasc. Sci.*, vol. 6, no. 1, pp. 7–11, 2020, doi: 10.4103/jpcs.jpcs_2_20.
[25] Z. A. Malik, M. Siddique, A. Imran, A. U. Yasin, A. H. Butt, and Z. J. Paracha, "Performance evaluation of classification algorithms for intrusion detection on NSL-KDD using Rapid Miner," *Int. J. Innov. Sci. Technol.*, vol. 4, no. 1, pp. 135–146, 2022, doi: 10.33411/ijist/2022040110.
[26] J. Ajayi, Y. Xu, L. Li, and K. Wang, "Enhancing flight delay predictions using network centrality measures," *Information*, vol. 15, no. 9, pp. 1–11, 2024, doi: 10.3390/info15090559.
[27] T. Toharudin *et al.*, "Boosting algorithm to handle unbalanced classification of PM2.5 concentration levels by observing meteorological parameters in Jakarta-Indonesia using AdaBoost, XGBoost, CatBoost, and LightGBM," *IEEE Access*, vol. 11, pp. 35680–35696, 2023, doi: 10.1109/ACCESS.2023.3265019.
[28] J. Gu, "Predicting student academic achievement using stacked ensemble learning with deep neural networks and fuzzy-based feature selection," *Sci. Rep.*, vol. 15, pp. 1–20, 2025, doi: 10.1038/s41598-025-20779-z.
[29] M. Kumar, J. Wadhwa, and A. Qtaishat, "Utilizing Random Forest and XGBoost data mining algorithms for anticipating students ' academic performance," *I.J. Mod. Educ. Comput. Sci.*, vol. 16, no. 2, pp. 29–44, 2024, doi: 10.5815/ijmecs.2024.02.03.

## BIOGRAPHIES OF AUTHORS

**Nur Maisarah Nor Azharludin** is a postgraduate student at Universiti Teknologi MARA (UiTM). She has a Bachelor's degree and Master's degree in Computer Science. Her research interests are machine learning algorithms, data analysis, and visualization. She is actively involved in the research group at UiTM. She can be contacted at email: nurmaisarahnorazharludin@gmail.com.

**Khyrina Airin Fariza Abu Samah** is an Associate Professor at the Faculty of Computer and Mathematical Sciences in Universiti Teknologi MARA (UiTM) Melaka Branch, Jasin Campus. Before joining UiTM, she had 13 years of working experience in the semiconductor industry. She has a Diploma, Bachelor's Degree and Master's Degree in Computer Science and Ph.D. in Information Technology. Her research interests are in artificial intelligence, algorithm analysis, machine learning, data science, optimization, and evacuation algorithms. She can be contacted at email: khyrina783@uitm.edu.my.

**Mohamad Faiz Dzulkalnine** holds a Bachelor's degree in Computer Science (Bioinformatics) and both a Master's and Doctorate in research from Universiti Teknologi Malaysia (UTM), specializing in data imputation, optimization, machine learning, and fuzzy systems. He has served as a Senior Lecturer at Universiti Tun Hussein Onn Malaysia (UTHM). Currently, he contributes to academia as a Senior Lecturer at Universiti Teknologi MARA (UiTM), Jasin, Melaka. In addition to his research interests, he teaches Diploma in Computer Science courses, equipping students with foundational and advanced knowledge. He can be contacted at email: mohamadfaiz@uitm.edu.my.

**Ahmad Firdaus Ahmad Fadzil** is a senior lecturer with over 10 years of experience at Universiti Teknologi MARA (UiTM). A proud alumnus of UiTM, he completed his diploma, degree, master's, and Ph.D. at the institution. His research interests include artificial intelligence, image processing, and parallel processing, where he is dedicated to advancing innovative solutions in these fields. He can be contacted at email: firdausfadzil@uitm.edu.my.

**Mohd Nor Hajar Hasrol Jono** is a Senior Lecturer at the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM), Melaka Jasin Campus. Graduated with a Ph.D. in Multimedia Education, Universiti Pendidikan Sultan Idris. M.Sc. Information Technology, UiTM and Bachelor's Degree in Computer Science, Universiti Teknologi Malaysia. Once served as a software developer at a leading company in Malaysia. As an academic, research is one of the areas of interest for developing ideas, sharing skills, and applying the latest innovations according to the latest technology and education trends. He can be contacted at email: hasrol@uitm.edu.my.

**Lala Septem Riza** received Ph.D. in Computer Science from Universidad de Granada, Spain, in 2015. He works in the Department of Computer Science Education, Universitas Pendidikan Indonesia, Indonesia. He teaches machine learning, big data platforms, and statistical data science. His research interests are in machine learning, data science, and education. He can be contacted at email: lala.s.riza@upi.edu.