EXPLORING ENHANCEMENT IN CONFIDENCE-BASED ASSESSMENT: A SYSTEMATIC LITERATURE REVIEW

NUR MAISARAH NOR AZHARLUDIN[1], KHYRINA AIRIN FARIZA ABU SAMAH[2], MOHAMAD FAIZ BIN DZULKALNINE[3], RASEEDA HAMZAH[4]

[1,2,3,4] Universiti Teknologi MARA Melaka Branch, Jasin Campus Melaka, Malaysia

E-mail: [1]nurmaisarahnorazharludin@gmail.com, [2]khyrina783@uitm.edu.my*,
[3]mohamadfaiz@uitm.edu.my, [4]raseeda@uitm.edu.my

ABSTRACT

Confidence-based assessment (CBA) is a technique designed to evaluate an individual's degree of confidence or expectation regarding their response to discern their true level of knowledge. In this methodology, individuals assign confidence scores to their answers, indicating their level of certainty about the correctness of their choices. This approach enhances understanding an individual's abilities or comprehension by distinguishing between correct responses with high confidence and those with low confidence. Consequently, evaluators gain a more comprehensive understanding of an individual's competence by examining their cognitive processes and self-awareness. Despite its potential, there is a lack of systematic reviews focusing on enhancing CBA. This study addresses this gap by conducting a systematic literature review (SLR) on improving CBA methodologies. The present study follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, systematically analysing 11 articles published between 2019 and 2024. These articles were selected from three primary databases—Scopus, Web of Science, and ScienceDirect—and one supplementary database, Google Scholar. The review of these studies identified four major themes: sector, purpose, algorithm, and methods used. The findings of this SLR provide valuable insights into the current state of CBA research and suggest directions for future studies. In conclusion, this research offers significant benefits for scholars in the CBA field, providing a reference for enhancing the application and understanding of CBA.

Keywords: *Confidence, Confidence-Based Assessment, Confidence-Based Learning, Enhancement, Systematic Literature Review*

1. INTRODUCTION

Confidence-based assessment (CBA) is an innovative method of evaluating learners that involves measuring knowledge and confidence levels [1]. According to [2], determining an individual's knowledge is contingent upon their level of assurance over the material. Moreover, it should be noted that CBA plays a crucial role within the framework of confidence-based learning (CBL), as it contributes to the identification of learner deficiencies through the process of deficiency diagnosis (DD) [1]. The assessment evaluates the accuracy of learners' knowledge by using multiple-choice questions (MCQs). It gauges their level of confidence in their responses. Individuals must indicate their selection of the MCQ answer and assess their confidence level regarding the accuracy of their chosen solutions.

Despite a significant amount of literature pertaining to CBA, there has been a notable lack of systematic analysis, exploring the enhancement of CBA and generating potential themes on this topic. The review methods in the existing literature, involving identification, screening, eligibility, and data abstraction, have not been properly managed. Authors often choose publications that support their particular area of interest [3]. As expected, transparency and bias are significant issues in standard literature reviews.

Therefore, the study aims to analyze how to enhance CBA in current studies by applying SLR with an approach using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method. The authors' empirical findings can be validated by using this approach to identify the gaps and propose various possibilities for future research within this particular topic. The research

question guided the present article: *"what is the current implementation of CBA?"*

The primary objective of the ongoing SLR is to analyze the current implementation of CBA. Also, we plan to identify the sectors contributing to the enhancement, discover the purpose of CBA enhancement in the studies, and acknowledge the algorithms and methods used in the previous studies. The results of this SLR will give authors guidelines and additional research on the enhancement of CBA. The remaining content of the paper is structured as follows: Section 2 will present the Methods, Section 3 will summarize the Results of the Literature, Section 4 will provide extensive Discussions, and the final Section 5 presents the Conclusion along with identified gaps and recommendations for future work.

2. METHODS

The primary objective of the SLR is to systematically identify, examine, and integrate existing literature pertaining to prior studies or research. This process is conducted methodically and transparently, employing consistent and reproducible techniques at each stage. Systematic reviews are alternatively referred to as Meta-narrative reviews or mixed reviews. According to [4], a SLR is a rigorous and comprehensive approach to identifying, evaluating, and interpreting all existing research significant to a specific research question, topic area, or phenomenon of interest. In addition, the process of distinguishing between what is already known and what is yet to be discovered can be a laborious task. The predefined and explicit methodological procedures are crucial in conducting systematic literature reviews. The present study employed the PRISMA approach to conduct SLR on enhancing the CBA. The PRISMA methodology is widely recognized for its effectiveness in conducting SLRs across multiple research domains, including but not limited to social sciences [5], engineering [6], and business [7]. Figure 1 shows that the PRISMA framework encompasses four primary stages: identification, screening, eligibility, and data abstraction and analysis.
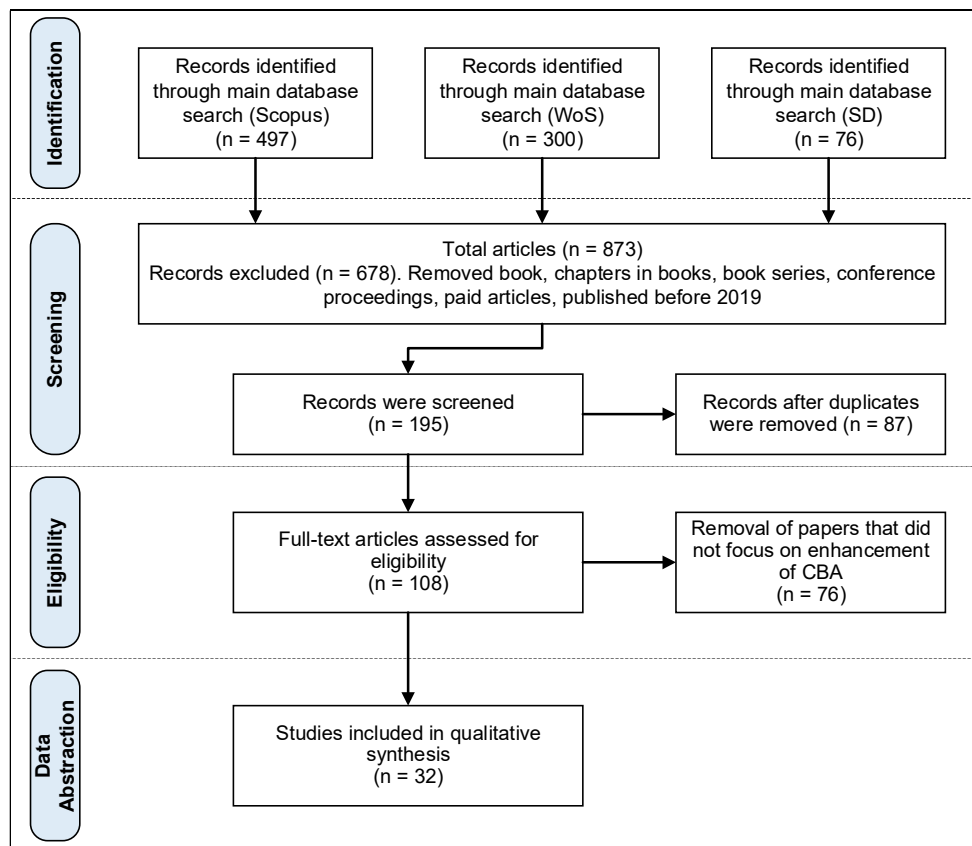

Figure 1: PRISMA Flow Diagram (Online Database's Identification of Studies)

2.1 Identification

Figure 1 shows the steps involved in the implementation of the PRISMA approach. The search of the literature process involved the evaluation of all 873 articles according to predetermined selection criteria. It accumulated 497 from Scopus, 300 from WoS, and 76 from SD for the first identification phase. This evaluation was conducted using the sorting function provided by the chosen databases. Uniform criteria were applied consistently across the chosen databases. In cases where sorting mechanisms were absent, articles were manually removed. Given the vast number of published publications, it is exceedingly challenging for researchers to evaluate all of them thoroughly. Therefore, [8] proposed that researchers should establish a specific timeframe to conduct their review.

The selection criteria included significant review articles and published research that utilized key terms such as *confidence-based assessment, confidence assessment, confident-based assessment, confident assessment, confidence estimation, confident estimation, confidence evaluation,* and *confident evaluation*. The study utilized a comprehensive literature search incorporating search parameters and key terms using Boolean operators such as "OR". The search engines utilized the main databases: Scopus, Web of Science (WoS) and ScienceDirect (SD). Google Scholar (GS) supports document searching if the documents from the main databases are locked. The search yielded over 873 research articles, of which 32 were deemed relevant based on the selection criteria for the current study.

2.2 Screening

Next, the process involved is screening. Through an examination of the search process conducted on the designated database, it has been observed that the quantity of studies pertaining to the augmentation of CBA has had a notable increase commencing in 2019 and continuing until January 2024, with the understanding that the current year has not yet concluded. Hence, the temporal interval spanning from 2019 to 2024 was designated as a criterion for inclusion. Moreover, in order to uphold the standard of the review, only publications containing empirical data and published in reputable journals are incorporated. In the screening phase, a total of 678 papers were eliminated from the analysis due to their failure to meet the predetermined inclusion criteria. Thus, the remaining 195 publications were deemed eligible for further examination.

The study utilized English-language research publications that had been published. Table 1 highlights the inclusion and exclusion criteria utilized in this study. The research articles that were discovered focused on the enhancement of CBA. The selection criteria excluded non-research papers such as review papers, conference proceedings, theses, chapters, and books from the SLR.

Table 1: The Criteria for Inclusion and Exclusion.

Criteria	Inclusion	Exclusion
Publication time	2019-2024	2018 and earlier
Document Type	Journal (research articles)	Other than journal
Language	English	Non-English
Focus of Research	Enhancement in CBA	Not related to enhancement in CBA
Country	Worldwide	(N/A)

2.3 Eligibility

The eligibility process represents the third phase, in which the authors manually monitor the recovered articles to ensure that all the remaining articles align with the predetermined requirements. The procedure involved reviewing the titles and abstracts of the articles. If the connection of the selected articles to the study remained unclear, the content of the articles was then assessed. A total of 76 articles were removed from the analysis due to various reasons. These included the presence of duplicate records across databases, the inclusion of non-English language publications, and the fact that the selected articles did not specifically focus on the enhancement of CBA. Additionally, some articles were excluded as they were in the form of review papers. In total, a selection of 108 articles was included in this process.

2.4 Data Abstraction and Analysis

The final phase, which is data abstraction, involves the process of data abstraction and analysis. The paper comprehensively analyses 32 selected articles, which were thoroughly researched and reviewed. The evaluations used particular studies aligned with the research inquiries and aims. The relevant subjects for the current study were identified by extracting the studies and thoroughly examining the publications' titles, abstracts, and full texts. A thematic analysis was conducted to uncover

themes pertaining to the improvement of CBA. The primary concerns, parallels, and distinctions clarified and depicted throughout the 32 papers were identified and classified.

In order to establish themes in this SLR, the researchers adhered to a six-step process for conducting thematic analysis. This approach was deemed appropriate for the qualitative analysis, as suggested by [9]. The process involves several steps. Firstly, it is essential to thoroughly understand and analyze the content of the 32 articles. Secondly, one must discern the commonalities and disparities among the topics addressed in these articles. Thirdly, it is necessary to generate or ascertain appropriate themes that can be constructed based on the identified similarities and differences found within the 32 articles. Lastly, it is crucial to ensure that the proposed themes align with the primary context of each article. In this study, a total of 32 publications were analyzed, identifying four distinct themes. Furthermore, a comprehensive report containing SLR findings was generated.

3. RESULTS

Table 2 shows the overall review of the collected 32 articles from the United States of America, China, United Kingdom, France, India, Netherlands, Australia, Austria, Belgium, Croatia, Italy, Poland, and Vietnam. Figure 2 illustrates the frequency of research papers collected from 2019 to 2024. The count of study papers has varied gradually with time, except from 2019 to 2020, which indicates a significant reduction (from seven to three articles), an increase from 2020 to 2021 (from three to six articles) and a decrease in studies from 2021 to 2022 (from six to three). However, there is a large increase in studies from 2022 to 2023 (from three to 10). The number of research publications in this scenario decreased from 10 articles in 2023 to three in 2024. This is the situation as the research only analyzed the prior studies until January 2024, and the year is not yet completed.
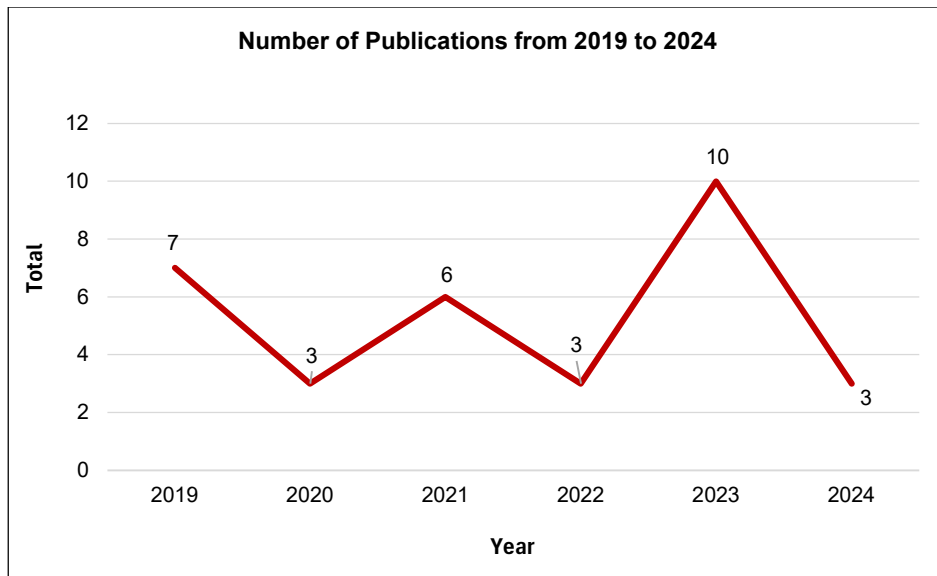

Figure 2: Number of Publications from 2019 to 2024 Related to Enhancement in CBA

Table 2: Matrix Table on SLR Research Study From 2019 to 2024

Authors		[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]	[29]	[30]	[31]	[32]	[33]	[34]	[35]	[36]	[37]	[38]	[39]	[40]	[41]
Country		France	Belgiu	USA	Croatia	Italy	USA	China	USA	USA	France	USA	China	USA	Austral	USA	USA	China	China	United	United	India	United Kingdom	India	Netherl	USA	Netherl	USA	USA	Vietna	Austria	USA	Poland
Sector Using CBA	HM		✓				✓		✓		✓			✓	✓		✓	✓	✓							✓							✓
	T				✓			✓																✓	✓						✓		
	SC	✓														✓					✓									✓			
	Ed								✓													✓	✓										
	Eg																				✓							✓				✓	
	Psy				✓							✓																					
	Ev																										✓		✓				
	Er			✓									✓																				
Purpose of Using CBA	AC								✓	✓	✓	✓	✓	✓	✓	✓																	
	PCP																✓	✓	✓	✓	✓	✓	✓										
	IAP																							✓	✓	✓	✓						
	IC																											✓	✓	✓	✓		
	IEQ																															✓	✓
	CAQQ	✓																															
	SR		✓																														
	AVSLR, FRP			✓																													
	IEP				✓																												
	SEC					✓																											
	ECKC						✓																										
	ETP, FU							✓																									
Algorithm Used in CBA	WSRT									✓				✓														✓	✓				
	FE								✓						✓																	✓	
	DST	✓									✓																						
	PCS								✓																								✓
	SRC											✓														✓							
	CNN																		✓								✓						
	BHC		✓																														
	I-LFOSLT			✓																													
	XGB, GCN-CNN				✓																												
	cDDM					✓																											
	WHO-ICRCBECC						✓																										
	PSF							✓																									
	K-W								✓																								

Methods Used in CBA:

Method	Checkmark column (approx.)
M-W	✓
St-T	✓
MEC-ST	✓
RFACGN	✓
CPCA	✓
SAC-DRL	✓
IR	✓
NUU	✓
HPP	✓
C-BEAM	✓
NB	✓
CETE, BA	✓
CK	✓
TIW, SP, SA	✓
CI	✓
VP	✓
S-B, G-B, N-B	✓
LS	✓ ✓ ✓ ✓ ✓ ✓
MCQ	✓ ✓
CF	✓ ✓
LE	✓
WoES	✓
LMHCA	✓
EICPF	✓
SSMF	✓
Pr-C, Po-C SA	✓
DKT	✓
SS	✓
SDn-NA	✓
CEI	✓
CBM	✓
SFM	✓
SAR	✓
DEA	✓
S-L	✓

PBE										✓						
ODQ											✓					
Ent												✓				
ECAPA-TDNN													✓			
FS-WA														✓		
P-W															✓	
GS																✓
CM																	✓

Theme 1: Sector Using CBA	Theme 2: Purpose of Using CBA	Theme 3: Algorithm Used in CBA	Theme 4: Methods Used in CBA
HM = Healthcare and Medical T = Technology SC = Safety-critical Ed = Education Eg = Engineering Psy = Psychology Ev = Environment Er = Energy	AC = Assess Confidence PCP = Predict Confidence or Performance IAP = Improve Accuracy and Performance IC = Improve Confidence IEQ = Improve Efficacy and Quality CAQQ = Compare Approach (Qualitative or Quantitative) SR = Study Relationship AVSLR, FRP = Ascertain the Veracity of Source Localization Result and Formulate Response Plan IEP = Improve Estimation of Prediction SEC = Sense the Evolution of Confidence ECKC = Evaluate the Change in Knowledge and Confidence ETP, FU = Enhance Target Positioning and Feature Update	WSRT = Wilcoxon Signed Rank Test FE = Fisher's Exact DST = Dempster-Shafer Theory PCS = Pearson Chi Square SRC = Spearman's Rank Correlation CNN = Convolutional Neural Network BHC = Bradford-Hill Criteria I-LFOSLT = Interconnection-Level FO Source Localization Tools XGB, GCN-CNN = XGBoost (Baseline Method) and Graph Convolutional Networks - Convolutional Neural Networks cDDM = Confidence-Driven Drift-Diffusion Model WHO-ICRC BECC = World Health Organization - International Committee of the Red Cross Basic Emergency Care course PSF = Point-spread function K-W = Kruskall-Wallis M-W = Mann-Whitney St-T = Student's t-Test MEC-ST = McNemar's Exact Chi-Squared Test RFACGN = Reliable Feature-Assisted Contrastive Generalization Net CPCA = Carcinogenic Potency Categorization Approach SAC-DRL = Soft Actor Critic-Deep Reinforcement Learning IR = Isotonic regression NUU = 'No Unknown Unknowns' HPP = Homogenous Poisson Process C-BEAM = Confidence-Based Evaluation Approach for MCQ NB = Naïve Bayes CETE, BA = Combination End-To-End and Backend Approaches CK = Cohen's Kappa TIW, SP, SA = Theories of Irrigation Works, Statistical Probability and System Analysis	LS = Likert-Scale MCQ = Multiple Choice Question CF = Confidence Function LE = Logical Expression WoES = WoE Scoring LMHCA = Low, Medium, High Confidence Assessment EICPF = Extend the Inducive Conformal Prediction Framework SSMF = Summarise Standard Mathematical Formulation Pr-C, Po-C SA = Pre-Confidence And Post-Confidence Self-Assessment DKT = Dual KCF Tracker SS = Sliding Scale 0-100 SDn-NA = Single, Double And N-Node Arguments CEI = Confidence Evaluation Index CBM = Confidence- Based Marking SFM = Structured Field Method SAR = Structure-Activity Relationships DEA = Deep Ensemble Approach S-L = Sample-Level PBE = Pivot Bootstrap Estimator ODQ = Operational Defined Questions Ent = Entropy ECAPA-TDNN = Emphasised Channel Attention, Propagation and Aggregation - Time Delay Neural Network FS-WA = First Step-Wise Approach P-W = Pixel-Wise GS = Grade Scoring CM = Condition Monitoring

		CI = Condition Indicator VP = Volcaniclastic Process S-B, G-B, N-B = Shannon-Based, Gini-Based, Norm-Based	

Figure 3: Themes and Subthemes Based on the Articles

The thematic analysis adaptation established four main themes and 77 subthemes. The content includes the sector using CBA identified as sector (nine subthemes), the purpose of using CBA identified as purpose (12 subthemes), the algorithm used in CBA designated as algorithm (30 subthemes), and the methods used in CBA identified as method (26 subthemes). The distribution is shown statistically, detailing the frequency of occurrence of the 77 subthemes in the articles.

Figure 3 illustrates the distribution of articles according to main themes and subthemes, as derived from the thematic analysis. The most reported values fall under the topics of sector using CBA, which are healthcare and medical, with 11 studies. Assess confidence has the highest number of studies for the purpose of using CBA, appearing in eight studies, followed by the Wilcoxon signed rank test for algorithm used in CBA in four studies, and the Likert scale for methods used in CBA in six studies.

Figure 4 illustrates the summary percentages of the four main themes among the 77 subthemes. The sector using CBA constituted 12% (9 of 77 articles' subthemes), the purpose of using CBA is 15% (12 of

77 articles' subthemes), the algorithm used in CBA represented 39% (30 of 77 articles' subthemes), and methods used in CBA is 34% (26 of 77 articles' subthemes). Consequently, it can be concluded that most studies on the enhancement of CBA focused on the algorithm used in CBA. The Wilcoxon signed rank test is the most used algorithm in the CBA, with four articles and 30 subthemes.
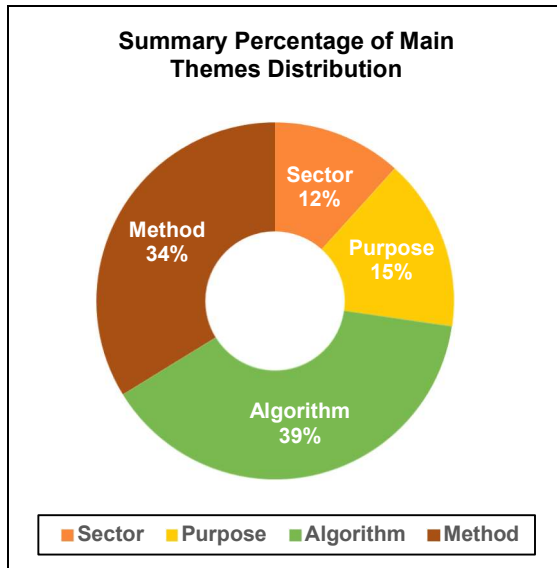
Figure 4: Summary Percentage of Themes Distribution

4. DISCUSSION

The aim of performing a systematic review is to analyze the factors involved in enhancing CBA. Confidence is important in learning as it supports a positive learning experience, helping students grow academically and personally. Based on the analysis, 32 papers identified are organized into five major themes: 1) sector using CBA, 2) purpose of using CBA, 3) algorithm used in CBA, 4) methods used in CBA, and 5) types of CBA used.

4.1 Sector Using CBA

The SLR reviewed all documented protocols from the previous research. Table 3 and Figure 5 show a set of sectors with the number of previous studies related to the enhancement of CBA based on the 32 articles that had been retrieved. The SLR identified eight sectors in all the articles. Based on the studies, the highest number of studies in the sector is healthcare and medical, with 11 studies (five from Scopus, two from WoS, and four from SD).

Next, the second highest sector was technology, with five studies (one study from Scopus, three

studies from WoS, and one study from SD), followed by the safety-critical sector with four studies (two studies from Scopus, no study from WoS, and two studies from SD). There are three sectors with the same number of studies, which are education (no study from Scopus, one study from WoS, and two studies from SD) and engineering (no study from Scopus, two studies from WoS, and one study from SD) with the total of three studies. The last three sectors are environment (no study from Scopus, one study from WoS, and one study from SD), energy (one study from Scopus, no study from WoS, and one study from SD) and psychology (two studies from Scopus, no study from WoS, and no study from SD), both with a total of two studies, respectively.

Based on the eight sectors mentioned, 13 countries were involved in the selected articles that focused on enhancement in CBA. Table 4 and Figure 6 illustrate the percentage distribution of the countries with the number of studies in each country. The highest number of studies are from the United States of America, with a total of 12 studies. Based on the studies from the United States of America, the percentage distribution from all the studies is 38%, which marks the majority of the studies, followed by China with 13%, contributing to four studies. The United Kingdom followed with 10%, contributing to three studies from the 32 articles. Furthermore, the three countries have the same two numbers of studies. The studies are in France, India, and the Netherlands, each contributing 6%. Lastly, the remaining seven countries, Australia, Austria, Belgium, Croatia, Italy, Poland, and Vietnam, contribute to one study each with 3% of the percentage, respectively.

Table 3: Number of Studies in Scopus, WoS, and SD Based on Sector

Sector	Scopus	WoS	SD	Total
Healthcare and Medical	5	2	4	11
Technology	1	3	1	5
Safety-critical	2	0	2	4
Education	0	1	2	3
Engineering	0	2	1	3
Psychology	2	0	0	2
Environment	0	1	1	2
Energy	1	0	1	2
Total	11	9	12	32
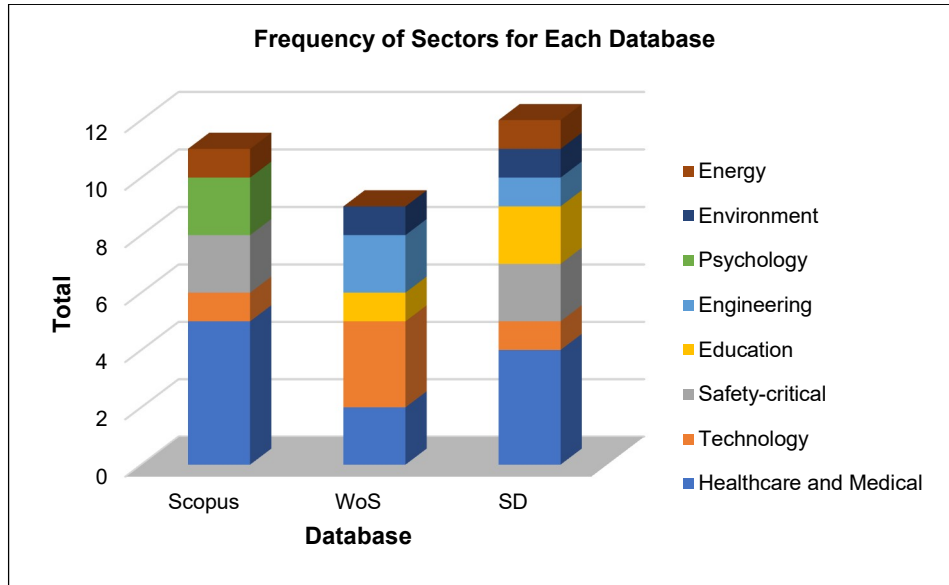
Figure 5: Frequency of Sectors for Each Database of Enhancement in CBA

Table 4: Number of Studies in Each Country in Previous Research

Country	Number of Studies
United States of America	12
China	4
United Kingdom	3
France	2
India	2
Netherlands	2
Australia	1
Austria	1
Belgium	1
Croatia	1
Italy	1
Poland	1
Vietnam	1

Figure 6: Percentage Country Based on Reviewed Studies of Enhancement in CBA
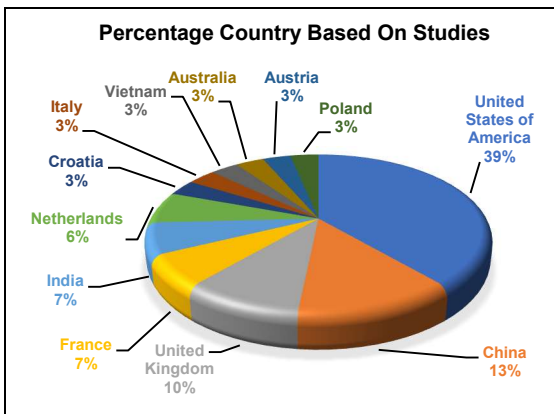
4.2 Purpose of Using CBA

Table 5 shows the list of the 32 previous studies and the purpose of each study used in this SLR. From the previous studies, the focus areas of those papers were: 1) assess confidence [17-24], 2) predict confidence or performance [25-31], 3) improve accuracy and performance [32-35], 4) improve confidence [36-39], and 5) improve efficacy and quality [40-41]. However, there are also purposes that specific only to the study that are: 1) compare approach (qualitative or quantitative) [10], 2) study relationship [11], 3) ascertain the veracity of source localization results and formulate response plan [12], 4) improve estimation of prediction [13], 5) sense the evolution of confidence [14], 6) evaluate the change in knowledge and confidence [15], and 7) enhance target positioning and feature update [16]. The collected papers covered the years from 2019 to January 2024.

Table 5: List of Previous Studies with Its Purpose

#	Studies	Purpose	Total
1	[17]		
2	[18]	Assess confidence	8
3	[19]		
4	[20]		

5	[21]		
6	[22]		
7	[23]		
8	[24]		
9	[25]	Predict confidence or performance	7
10	[26]		
11	[27]		
12	[28]		
13	[29]		
14	[30]		
15	[31]		
16	[32]	Improve accuracy and performance	4
17	[33]		
18	[34]		
19	[35]		
20	[36]	Improve confidence	4
21	[37]		
22	[38]		
23	[39]		
24	[40]	Improve efficacy and quality	2
25	[41]		
26	[10]	Compare approach (qualitative or quantitative)	1
27	[11]	Study relationship	1
28	[12]	Ascertain the veracity of source localization result	1

		and formulate response plan	
29	[13]	Improve estimation of prediction	1
30	[14]	Sense the evolution of confidence	1
31	[15]	Evaluate the change in knowledge and confidence	1
32	[16]	Enhance target positioning and feature update	1

Based on Figure 7, 12 purposes can be identified from the articles reviewed. The highest number of studies assesses confidence with knowledge, with eight studies displaying 25% of the overall percentage, followed by predicting confidence or performance with 22% and seven studies. Next, both improve accuracy or performance and confidence with four studies, with 13% of the percentage, respectively. The purpose of improving efficacy and quality was identified in two studies with a percentage of 6%. The smallest percentage, 3%, was given to the rest of the seven purposes, with only one study for each purpose.
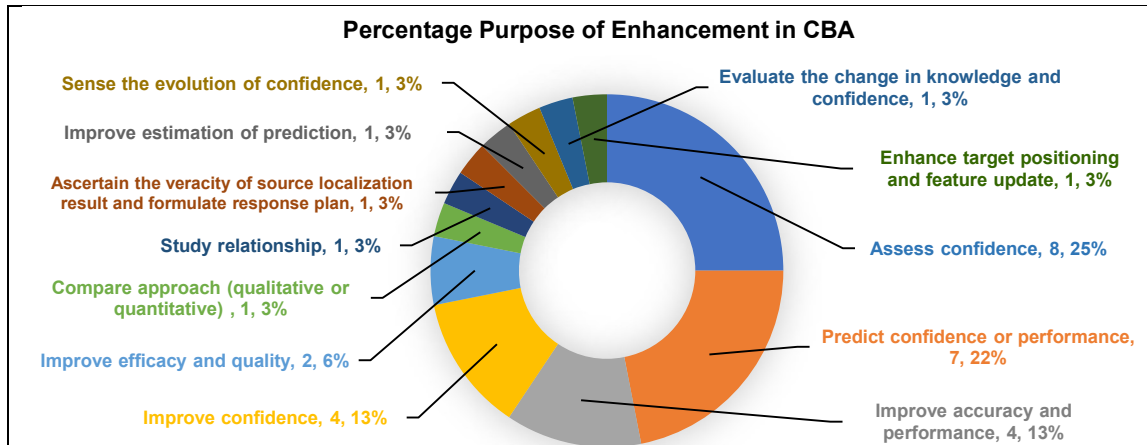
Figure 7: Percentage Purpose of Enhancement in CBA

4.3 Algorithm Used in CBA

As the studies are from different circumstances, the algorithms used in each study might be different. Even though the same algorithms were used, such as Dempster-Shafer theory, Spearman's rank correlation, Wilcoxon signed-rank test, and convolutional neural network (CNN), each of the studies for the algorithms used were relatively different. Table 6 displays all the algorithms used for the enhancement in CBA based on the 32 articles that had been retrieved.

Two studies used Dempster-Shafer's theory, both from the safety-critical sector. However, one study compares the approach used, either qualitative or quantitative [10], and the other assesses confidence [25]. A study by [18] used the Pearson Chi-Square and Fisher's Exact test in the education sector to improve the pre-and post-implementation degree's efficacy and quality. The Carcinogenic potency categorization approach (CPCA) algorithm in healthcare and medical studies of N-nitrosamines [25], the Soft Actor Critic-Deep Reinforcement Learning (SAC-DRL) algorithm [25] in the energy sector, and CNN in healthcare and medical sector

[27] are used to predict confidence or performance in the studies. The other studies that predict confidence or performance used Isotonic Regression [28], 'No unknown unknowns' and homogenous Poisson process [29], Confidence-based evaluation approach for MCQs (C-BEAM) [30], and Naïve Bayes [31].

Next, two studies used Spearman's rank correlation algorithm. One study used it to assess confidence in the healthcare and medical sector [20]. In contrast, the other study improves accuracy and performance in the environment sector [33]. The study's purpose to assess the confidence used reliable feature-assisted contrastive generalization net (RFACGN) in engineering, Wilcoxon signed rank test in education, student's t-test, McNemar's exact chi-squared test, Wilcoxon matched-pair signed-rank in psychology, Kruskall-Wallis, Mann-Whitney, chi-square and Fisher's Exact test in IBM SPSS in healthcare and medical, and Volcaniclastic process in environment as its algorithms. The Bradford-Hill criteria study relationships in the healthcare and medical sector. The related studies use the combination of end-to-end and backend approaches, feedforward neural networks (FNN) or CNN, and Cohen's Kappa algorithms to improve accuracy and performance.

Furthermore, the algorithm used to ascertain the veracity of source localization results and formulate a response plan is interconnection-level forced oscillations (FO) source localization tools. The Wilcoxon signed-rank test is used in several studies to improve confidence in healthcare and medical. The study aims to improve confidence by using theories of irrigation works, statistical probability and system analysis in safety-critical and condition indicator (CI) in engineering as its algorithm. The other algorithms used are XGBoost (baseline method) and Graph Convolutional Networks - Convolutional Neural Networks (GCN-CNN) in technology sector is used to improve estimation of prediction, confidence-driven drift-diffusion model (cDDM) in psychology to sense the evolution of confidence, WHO-International Committee of the Red Cross (WHO-ICRC) Basic Emergency Care (BEC) course in healthcare and medical to evaluate the change in knowledge and confidence, point-spread function (PSF) and Shannon-based, Gini-based, Norm-based in technology sector.

Table 6: List of Algorithms Used in Each Study of the Previous Works

Studies	Algorithm Used	Total
[18]	Wilcoxon signed rank test	4
[22]		
[36]		
[37]		
[17]	Fisher's Exact	3
[23]		
[40]		
[10]	Dempster-Shafer theory	2
[19]		
[17]	Pearson Chi Square	2
[40]		
[20]	Spearman's rank correlation	2
[33]		
[27]	CNN	2
[34]		
[11]	Bradford-Hill criteria	1
[12]	Interconnection-level FO source localization tools	1
[13]	XGBoost (baseline method) and GCN-CNN	1
[14]	cDDM	1
[15]	WHO-ICRC BEC course	1
[16]	PSF	1
[17]	Kruskall-Wallis	1
[17]	Mann-Whitney	1
[18]	Student's t-test	1
[18]	McNemar's exact chi-squared test	1
[21]	RFACGN	1
[25]	CPCA	1
[26]	SAC-DRL	1
[28]	Isotonic regression	1
[29]	'No unknown unknowns'	1
[29]	Homogenous Poisson process	1
[30]	C-BEAM	1
[31]	Naïve Bayes	1
[32]	Combination end-to-end and backend approaches	1
[35]	Cohen's Kappa	1
[38]	Theories of irrigation works, statistical probability and system analysis	1
[39]	CI	1
[24]	Volcaniclastic process	1
[41]	Shannon-based, Gini-based, Norm-based	1

4.4 Methods Used in CBA

Several methods are used in the articles reviewed in this SLR. The reviewed studies applied different methods for the analyses. The method that had the greatest number of studies is the Likert scale. The Likert scale is the most commonly used method to measure a participant's confidence level [42].

Table 7 shows a list of methods used in each study based on the 32 articles. There are six studies used the Likert scale that consists of either 10-, 5-, or 4-point; two studies used MCQ; two studies used confidence function; one study used the logical expression; one study used weight-of-evidence (WoE) scoring; one study used low, medium, high confidence assessment; one study used to extend the inducive conformal prediction framework; one study used summarise standard mathematical formulation of the basic drift-diffusion model (DDM); one study used pre-confidence and post-confidence self-assessment, and one study used dual kernelized correlation filter (KCF) tracker,

Table 7: List of Methods Used in Each Study

Studies	Methods Used	Total
[18]	Likert-scale	6
[20]		
[22]		
[36]		
[37]		
[40]		
[17]	MCQ	2
[30]		
[38]	Confidence function	2
[41]		
[10]	Logical expression	1
[11]	WoE scoring	1
[12]	Low, medium, high confidence assessment	1
[13]	Extend the inducive conformal prediction framework	1
[14]	Summarise standard mathematical formulation	1
[15]	Pre-confidence and post-confidence self-assessment	1
[16]	Dual KCF tracker	1
[17]	Sliding scale 0-100	1
[19]	Single, double and n-node arguments	1
[21]	Confidence evaluation index	1
[23]	CBM	1
[24]	Structured field method	1
[25]	SAR	1
[26]	Deep ensemble approach	1
[27]	Sample-level	1
[28]	Pivot bootstrap estimator	1
[29]	Operational defined questions	1
[31]	Entropy	1
[32]	ECAPA-TDNN	1
[33]	First step-wise approach	1
[34]	Pixel-wise	1
[35]	Grade scoring	1
[39]	Condition monitoring	1

In addition, other methods used in the studies are sliding scale 0-100 (one study), single, double and n-node arguments (one study), confidence evaluation index (one study), confidence-based marking (CBM) (one study), structured field method (one study), structure-activity relationships (SAR) (one study), deep ensemble approach (one study), sample-level (one study), pivot bootstrap estimator (one study), operationally defined questions (one study), entropy (one study), Emphasised Channel Attention, Propagation and Aggregation - Time Delay Neural Network (ECAPA-TDNN) (one study), first step-wise approach (one study), pixel-wise (one study), grade scoring (one study), and condition monitoring (one study). The previous works show different methods used in the studies as each article studies different topics, and CBA is widely used [43].

4.5 Outcome from the Systematic Review

As mentioned before, the main objective of this study is to conduct a systematic review and to understand the enhancement of CBA that the previous studies had done. This study focused on four factors: the sector related to the studies, the purposes of previous studies, the algorithms used, and the method used to enhance the CBA. The factors can be identified by the review of the 32 articles that had been finalized systematically. The articles are retrieved based on the inclusion criteria focused on the enhancement of CBA from five years back, from 2019 to January 2024. The main databases are Scopus, WoS, and SD, with a supporting database, GS.

The SLR recognized eight sectors involved in the review based on the articles. The sectors are safety-critical, education, healthcare and medical, energy, engineering, technology, environment, and psychology. From the studies' sectors, the countries of the articles can also be identified. There are 13 countries involved in the study of enhancement in CBA: the United States of America, China, Belgium, India, Netherlands, Croatia, Italy, Vietnam, United Kingdom, Australia, Austria, Poland, and France. The amount of study is still considered insufficient, as CBA is used not just in these eight sectors but as a well-known assessment to measure confidence. Thus, more than eight sectors may be involved in CBA.

This SLR focused on enhancing CBA based on 32 articles retrieved and discovered 12 distinct purposes of the studies. The purposes are to compare approaches (qualitative or quantitative), improve

efficacy and quality, predict confidence or performance, assess confidence or knowledge, study relationships, improve accuracy and performance, ascertain the veracity of source localization results and formulate a response plan, improve confidence, improve estimation of prediction, sense the evolution of confidence, evaluate the change in knowledge and confidence, and enhance target positioning and feature update. The purposes of the studies vary, following their respective primary objectives. Due to the study's cross-sectoral nature, the CBA indicators utilized in each study were distinct. Nevertheless, each study employed CBA. Subsequently, the improvement in CBA can be identified.

This SLR concentrated on the algorithms used in the enhancement studies in CBA, in which 34 different algorithms can be identified from the review. The algorithms are Dempster-Shafer theory, Pearson Chi-Square, Fisher's Exact, CPCA, SAC-DRL, CNN, Spearman's rank correlation, RFACGN, Bradford-Hill criteria, Wilcoxon signed rank, combination end-to-end and backend approaches, interconnection-level FO source localization tools, XGBoost as a baseline method and GCN-CNN, cDDM, Kruskall-Wallis, Mann-Whitney, WHO-ICRC BEC course, student's t-test, McNemar's exact chi-squared test, theories of irrigation works, statistical probability and system analysis, isotonic regression, 'no unknown unknowns' and homogenous Poisson process, PSF, C-BEAM, Naïve Bayes, CI, volcaniclastic process, Shannon-based, Gini-based, Norm-based, FNN, and Cohen's Kappa. It can be concluded that numerous algorithms are used to enhance CBA. Some studies did not just use one algorithm; some studies used several algorithms.

There were 26 methods also used in the studies. CBA is easy to apply in a study; thus, many methods can be used. A logical expression is used in one study from the safety-critical sector; six studies used a Likert scale of 10-, 5-, or 4-points. A 10-question Likert scale is used in the education sector; a 5-point Likert scale is used in healthcare and medical education; and a 4-point Likert scale is used in the healthcare and medical sectors to improve confidence. The study using SAR is from the healthcare and medical sectors; the deep ensemble approach is from the energy sector; sample-level methods are used in the healthcare and medical sectors; and single, double, and n-node arguments are used in the safety-critical sector to assess the confidence in the study.

Moreover, one study using the confidence evaluation index is from the engineering sector; one study using WoE scoring in healthcare and medical to study the most confident key event relationship with the adverse outcome; one study using ECAPA-TDNN in technology to improve accuracy and performance, and one study using the first step-wise approach in environment. In addition, other methods used in the studies are low, medium, and high confidence assessments used in one study in the energy sector, extending the induciue conformal prediction framework in the technology sector, summarizing standard mathematical formulation of the basic DDM, eight MCQs and a sliding scale of 0-100, 11-question pre-confidence and post-confidence self-assessment, and confidence function in the two studies in safety-critical and technology sector to improve confidence and improve efficacy and quality, respectively.

Besides, the pivot bootstrap estimator is used in healthcare and medical to predict confidence or performance, operational defined safety-critical questions, dual KCF tracker in technology, MCQ in education, entropy in technology, CBM in healthcare and medical, condition monitoring engineering, structured field method in the environment, pixel-wise engineering, and 8-grade scoring in healthcare and medical to improve accuracy and performance. Nevertheless, this shows that enhancement in CBA is developed widely in many sectors, with different purposes, and can be discovered using different algorithms and methods to improve CBA.

4.6 Gaps and Future Research Directions

Recognizing limitations guarantees the integrity and openness of the study. By clearly acknowledging any limitations or deficiencies in the technique or data, it aids in upholding the accuracy and dependability of the study. Therefore, it promotes further research into the contribution of knowledge to continuously advancing knowledge in a specific field or sector. Based on the research, several gaps were identified that resulted from this study. As CBA is important and frequently used in studies to test the participant's confidence in the subject matter, it is essential to ensure that the participants' confidence is precise.

However, not much research considers the time taken to answer the test as one of the most vital indicators in CBA for the accuracy of confidence in participants. Due to the limited access to databases in the university library system, only several databases can be accessed. In addition, only several

databases are used in this study: Scopus, WoS, and SD. Hence, the number of articles that could be retrieved was limited.

This study put forth several recommendations that should be considered for future research. Further research is required to examine the efficacy of CBA. The significance of CBA lies in its ability to evaluate both individuals' knowledge and their confidence in responding to the questions. The CBA evaluation is frequently utilized across many different sectors. The study's findings are also valuable for highlighting the improvements made to the broader community. Furthermore, expanding knowledge related to the present study on enhancing CBA can be achieved by utilizing diverse databases or search engines, such as IEEE Xplore and the Directory of Open Access (DOAJ). In addition, it is advisable to employ diverse search methodologies, including expert consultation, citation tracking, reference searching, and snowballing, in order to enhance existing methods utilized in the SLR of CBA.

4.7 Research Contribution

This SLR offers a thorough analysis of the current implementation of CBA, focusing on the sector using CBA, purpose of using CBA, algorithm used in CBA, and methods used in CBA. The review finds significant patterns by combining research from a variety of disciplines, including healthcare and medical, technology, safety critical, education, engineering, psychology, environment, and energy. The findings include the rationale behind the usage of CBA as well as the algorithms and techniques employed in CBA. In contrast to earlier evaluations that mostly concentrate on a single industry or set of methodologies, this study distinctively covers several areas, revealing shortcomings, including bias towards the particular area of concern. The SLR outcomes provide researchers, educators, and assessment designers with practical insights and suggestions to increase the scalability and dependability of CBA systems. The review substantially contributes to advancing CBA theory and practice across many industries by outlining a research plan that promotes solutions to existing problems.

5. CONCLUSION

The primary objective of this study is to conduct a systematic evaluation of the enhancement of CBA from 2018 to 2023. Using the PRISMA approach and thematic analysis, we systematically reviewed 32 selected publications from three main databases:

Scopus, WoS, and SD, and one supplementary database: GS. The study systematically explored four primary themes: sector using CBA, purpose of using CBA, algorithm used in CBA, and methods used in CBA. We identified 77 subthemes, which include the sector using CBA constituted 12% (9 of 77 articles' subthemes), the purpose of using CBA is 15% (12 of 77 articles' subthemes), the algorithm used in CBA represented 39% (30 of 77 articles' subthemes), and methods used in CBA is 34% (26 of 77 articles' subthemes). The SLR indicated that the main sector used CBA in the selected studies were healthcare and medical. The purpose of CBA utilized in most studies is to assess the studies' confidence. In addition, the algorithm most used related to CBA in the studies selected is the Wilcoxon Signed Rank Test, and the most used method in CBA is the Likert scale.

Thus, the outcomes of this SLR provide various noteworthy contributions to practical applications and the existing body of knowledge. The review provides an opportunity for interested parties, including educators, the general public, and researchers, to gain insight into the methodologies employed by past researchers in their efforts to enhance the accuracy of individual confidence through the improvement of CBA. The outcome yields a beneficial influence on the individual's result. Moreover, the findings provide valuable insights to the researchers into the particular domains and subject matter of research pertaining to CBA that warrant their attention. The review determined that multiple methods exist for enhancing the effectiveness of CBA, which can contribute to the strengthening of confidence in responding to questions. However, it is important to acknowledge that this study does have several inherent limitations. Due to constraints related to database accessibility, only three databases, specifically Scopus, WoS, and SD, were utilized. GS was employed as a supplementary database. Future research should consider a larger or smaller sample size of publications when considering the analysis assessment methods. The objective of a quality evaluation is not solely to identify flawless papers but rather to aid in identifying articles that align with the specific objectives of the review.

REFERENCES:

[1] R. Chatterjee and J. K. Mandal, "Two dimensional assessment technique for CBL," in *Proceedings of 2016 IEEE International Conference on Teaching, Assessment and Learning for Engineering, TALE 2016*, 2017, pp. 40–43, doi: 10.1109/TALE.2016.7851767.

[2] T. M. Adams and G. W. Ewen, "The importance of confidence in improving educational outcomes," in *25th Annual Conference on Distance Teaching & Learning*, 2009, pp. 1–5.

[3] H. A. M. Shaffril, S. F. Samsuddin, and A. A. Samah, "The ABC of systematic literature review: The basic methodological guidance for beginners," *Qual. Quant.*, vol. 55, no. 4, pp. 1319–1346, 2021, doi: 10.1007/s11135-020-01059-6.

[4] A. J. Sánchez-Garrido, I. J. Navarro, J. García, and V. Yepes, "A systematic literature review on modern methods of construction in building: An integrated approach using machine learning," *J. Build. Eng.*, vol. 73, pp. 1–29, 2023, doi: 10.1016/j.jobe.2023.106725.

[5] H. A. M. Shaffril, A. A. Samah, S. F. Samsuddin, and Z. Ali, "Mirror-mirror on the wall, what climate change adaptation strategies are practiced by the Asian's fishermen of all?," *J. Clean. Prod.*, vol. 232, pp. 104–117, 2019, doi: 10.1016/j.jclepro.2019.05.262.

[6] S. N. Ismail, A. Ramli, and H. A. Aziz, "Research trends in mining accidents study: A systematic literature review," *Saf. Sci.*, vol. 143, pp. 1–13, 2021, doi: 10.1016/j.ssci.2021.105438.

[7] R. G. Bilro, S. M. C. Loureiro, and P. Souto, "A systematic review of customer behavior in business-to-business markets and agenda for future research," *J. Bus. Ind. Mark.*, vol. 38, no. 13, pp. 122–142, 2023, doi: 10.1108/JBIM-07-2022-0313.

[8] C. Okoli, "A guide to conducting a standalone systematic literature review," *Commun. Assoc. Inf. Syst.*, vol. 37, no. 1, pp. 879–910, 2015, doi: 10.17705/1cais.03743.

[9] L. S. Nowell, J. M. Norris, D. E. White, and N. J. Moules, "Thematic analysis: Striving to meet the trustworthiness criteria," *Int. J. Qual. Methods*, vol. 16, no. 1, pp. 1–13, 2017, doi: 10.1177/1609406917733847.

[10] Y. Idmessaoud, D. Dubois, and J. Guiochet, "Confidence assessment in safety argument structure - Quantitative vs. qualitative approaches," *Int. J. Approx. Reason.*, vol. 165, pp. 1–26, 2024, doi:

10.1016/j.ijar.2023.109100.

[11] J. van Ertvelde *et al.*, "Optimization of an adverse outcome pathway network on chemical-induced cholestasis using an artificial intelligence-assisted data collection and confidence level quantification approach," *J. Biomed. Inform.*, vol. 145, pp. 1–9, 2023, doi: 10.1016/j.jbi.2023.104465.

[12] S. Biswas, J. Follum, and J. H. Eto, "Confidence Assessment for Regional Forced Oscillation Source Localization: Formulation and Field Validation," *IEEE Trans. Power Deliv.*, vol. 38, no. 6, pp. 3739–3748, 2023, doi: 10.1109/TPWRD.2023.3284418.

[13] D. Oršolić and T. Šmuc, "Dynamic applicability domain (dAD): compound-target binding affinity estimates with local conformal prediction," *Bioinformatics*, vol. 39, no. 8, pp. 1–12, 2023, doi: 10.1093/bioinformatics/btad465.

[14] D. G. Lee, J. Daunizeau, and G. Pezzulo, "Evidence or confidence: What is really monitored during a decision?," *Psychon. Bull. Rev.*, vol. 30, no. 4, pp. 1360–1379, 2023, doi: 10.3758/s13423-023-02255-9.

[15] S. M. Kivlehan *et al.*, "Evaluation of change in emergency care knowledge and skills among front-line healthcare providers in Ukraine with the Basic Emergency Care course: A pretest/post-test study," *BMJ Open*, vol. 12, no. 6, pp. 1–8, 2022, doi: 10.1136/bmjopen-2021-050871.

[16] C. Zhong, J. Ding, and Y. Zhang, "Video SAR Moving Target Tracking Using Joint Kernelized Correlation Filter," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 15, pp. 1481–1493, 2022, doi: 10.1109/JSTARS.2022.3146035.

[17] D. Owerko, K. Ryan, E. Cabacungan, K. Yan, and K. Saudek, "Neonatal hyperbilirubinemia: Assessing variation in knowledge and practice," *PLoS One*, vol. 18, pp. 1–9, 2023, doi: 10.1371/journal.pone.0282413.

[18] L. M. L. Agler, K. Noguchi, and L. K. Alfsen, "Personality traits as predictors of reading comprehension and metacomprehension accuracy," *Curr. Psychol.*, vol. 40, no. 10, pp. 5054–5063, 2021, doi: 10.1007/s12144-019-00439-y.

[19] R. Wang, J. Guiochet, G. Motet, and W. Schön, "Safety case confidence propagation based on Dempster-Shafer," *Int. J. Approx. Reason.*, vol. 107, pp. 46–64, 2019, doi: 10.1016/j.ijar.2019.02.002.

[20] L. Ehsan, C. Fealko, D. Maharaj, C. Huffman,

and N. Behrle, "Pediatric electrocardiogram interpretation improvement and confidence assessment for resident physicians via training skills workshop," *Prog. Pediatr. Cardiol.*, vol. 72, pp. 1–7, 2023, doi: 10.1016/j.ppedcard.2023.101700.

[21] Z. Shi, J. Chen, X. Zhang, Y. Zi, C. Li, and J. Chen, "A reliable feature-assisted contrastive generalization net for intelligent fault diagnosis under unseen machines and working conditions," *Mech. Syst. Signal Process.*, vol. 188, pp. 1–23, 2023, doi: 10.1016/j.ymssp.2022.110011.

[22] A. C. Lentz *et al.*, "Cadaveric laboratory simulation training of male stress urinary incontinence treatment improves trainee knowledge and confidence," *Urology*, vol. 143, pp. 48–54, 2020, doi: 10.1016/j.urology.2020.06.008.

[23] C. Scuderi and T. Pain, "Using confidence assessment to explore physiotherapists, dieticians, speech pathologists and occupational therapists knowledge of medicines," *J. Interprof. Care*, vol. 33, no. 5, pp. 593–597, 2019, doi: 10.1080/13561820.2018.1551860.

[24] J. Cowlyn, B. M. Kennedy, D. M. Gravley, and S. J. Cronin, "A confidence-based assessment method for distinguishing pyroclastic density current deposits from other volcaniclastic units," *Front. Earth Sci.*, vol. 8, pp. 1–14, 2020, doi: 10.3389/feart.2020.581195.

[25] S. Chakravarti, R. D. Saiakhov, and M. Girireddy, "Confidence score calculation for the carcinogenic potency categorization approach (CPCA) predictions for N-nitrosamines," *Comput. Toxicol.*, vol. 29, pp. 1–11, 2024, doi: 10.1016/j.comtox.2023.100298.

[26] J. Wu, C. Huang, H. He, and H. Huang, "Confidence-aware reinforcement learning for energy management of electrified vehicles," *Renew. Sustain. Energy Rev.*, vol. 191, pp. 1–12, 2024, doi: 10.1016/j.rser.2023.114154.

[27] W. Xue, T. Guo, and D. Ni, "Left ventricle quantification with sample-level confidence estimation via Bayesian neural network," *Comput. Med. Imaging Graph.*, vol. 84, pp. 1–7, 2020, doi: 10.1016/j.compmedimag.2020.101753.

[28] N. Tomašev *et al.*, "A clinically applicable approach to continuous prediction of future acute kidney injury," *Nature*, vol. 572, pp. 116–119, 2019, [Online]. Available: http://dx.doi.org/10.1038/s41586-019-1390-1.

[29] J. Rougier, "Confidence in risk assessments," *J. R. Stat. Soc. Ser. A Stat. Soc.*, vol. 182, no. 3, pp. 1081–1095, 2019, doi: 10.1111/rssa.12445.

[30] A. Aleem and M. M. Gore, "C-BEAM: A confidence-based evaluation of MCQs for providing feedback to instructors," *Comput. Appl. Eng. Educ.*, vol. 27, no. 1, pp. 112–127, 2019, doi: 10.1002/cae.22061.

[31] G. N. Tornetta, "Entropy methods for the confidence assessment of probabilistic classification models," *Statistica*, vol. 81, no. 4, pp. 383–398, 2021, doi: 10.6092/issn.1973-2201/11479.

[32] S. Dey and G. Saha, "Addressing the semi-open set dialect recognition problem under resource-efficient considerations," *Speech Commun.*, vol. 152, pp. 1–18, 2023, doi: 10.1016/j.specom.2023.102957.

[33] G. J. Piet, J. E. Tamis, J. Volwater, P. de Vries, J. T. van der Wal, and R. H. Jongbloed, "A roadmap towards quantitative cumulative impact assessments: Every step of the way," *Sci. Total Environ.*, vol. 784, pp. 1–19, 2021, doi: 10.1016/j.scitotenv.2021.146847.

[34] R. Liu, S. Cheng, L. Tian, and J. Yi, "Deep spectral learning for label-free optical imaging oximetry with uncertainty quantification," *Light Sci. Appl.*, vol. 8, no. 1, pp. 13–15, 2019, doi: 10.1038/s41377-019-0216-0.

[35] L. A. Hagens *et al.*, "Improvement of an interobserver agreement of ARDS diagnosis by adding additional imaging and a confidence scale," *Front. Med.*, vol. 9, pp. 1–11, 2022, doi: 10.3389/fmed.2022.950827.

[36] G. D. Haischer-Rollo, J. K. Aden, C. S. Percival, and C. M. Drumm, "A multimodal, resident-led newborn emergencies night curriculum," *Mil. Med.*, vol. 188, no. 9–10, pp. 3216–3220, 2023, doi: 10.1093/milmed/usad176.

[37] M. G. Drake *et al.*, "Development of a national academic boot camp to improve fellowship readiness," *ATS Sch.*, vol. 2, no. 1, pp. 49–65, 2021, doi: 10.34197/ats-scholar.2020-0091OC.

[38] L. H. Nguyen and Q. H. Nguyen, "Confidence-based assessment of safety of irrigation reservoir in Vietnam," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 10, pp. 2884–2893, 2019, doi: 10.35940/ijitee.J9616.0881019.

[39] M. Surányi, M. Leighton, C. Reinbrecht, and M. Bader, "Evaluation of a novel condition indicator with comparative application to hypoid gear failures," *Appl. Sci.*, vol. 11, no. 23, pp. 1–28, 2021, doi: 10.3390/app112311499.

[40] S. Dowd, J. A. Thompson, J. A. Pearson, S. Pinero, and V. C. Simmons, "Education for Registered Nurses Completing Preoperative

Anesthesia Interviews: A Quality Improvement Project," *J. Perianesthesia Nurs.*, vol. 38, no. 3, pp. 382–393, 2023, doi: 10.1016/j.jopan.2022.06.003.

[41] M. Bukowski, J. Kurek, I. Antoniuk, and A. Jegorowa, "Decision confidence assessment in multi-class classification," *Sensors*, vol. 21, no. 11, pp. 1–15, 2021, doi: 10.3390/s21113834.

[42] A. T. Jebb, V. Ng, and L. Tay, "A review of key Likert scale development advances: 1995–2019," *Front. Psychol.*, vol. 12, pp. 1–14, 2021, doi: 10.3389/fpsyg.2021.637547.

[43] P. Walsh, P. Owen, and N. Mustafa, "The creation of a confidence scale: The confidence in managing challenging situations scale," *J. Res. Nurs.*, vol. 26, no. 6, pp. 483–496, 2021, doi: 10.1177/1744987120979272.