

Review

A Systematic Review of Electroencephalography-Based Emotion Recognition of Confusion Using Artificial Intelligence

Dasuni Ganepola ¹, Madduma Wellalage Pasan Maduranga ², Valmik Tilwari ^{3,*} and Indika Karunaratne ¹¹ Department of Information Technology, University of Moratuwa, Moratuwa 10400, Sri Lanka² Department of Computer Engineering, Faculty of Computing, General Sir John Kotelawala Defence University, Rathmalana 10390, Sri Lanka³ Department of Electrical Engineering, Korea University, Seoul 02841, Republic of Korea

* Correspondence: valmik@iitg.ac.in or valmik@korea.ac.kr

Abstract: Confusion emotion in a learning environment can motivate the learner, but prolonged confusion hinders the learning process. Recognizing confused learners is possible; nevertheless, finding them requires a lot of time and effort. Due to certain restrictions imposed by the settings of an online learning environment, the recognition of confused students is a big challenge for educators. Therefore, novel technologies are necessary to handle such crucial difficulties. Lately, Electroencephalography (EEG)-based emotion recognition systems have been rising in popularity in the domain of Education Technology. Such systems have been utilized to recognize the confusion emotion of learners. Numerous studies have been conducted to recognize confusion emotion through this system since 2013, and because of this, a systematic review of the methodologies, feature sets, and utilized classifiers is a timely necessity. This article presents the findings of the review conducted to achieve this requirement. We summarized the published literature in terms of the utilized datasets, feature preprocessing, feature types for model training, and deployed classifiers in terms of shallow machine learning and deep learning-based algorithms. Moreover, the article presents a comparison of the prediction accuracies of the classifiers and illustrates the existing research gaps in confusion emotion recognition systems. Future study directions for potential research are also suggested to overcome existing gaps.



Citation: Ganepola, D.; Maduranga, M.W.P.; Tilwari, V.; Karunaratne, I. A Systematic Review of

Electroencephalography-Based Emotion Recognition of Confusion Using Artificial Intelligence. *Signals* **2024**, *5*, 244–263. <https://doi.org/10.3390/signals5020013>

Received: 22 February 2024

Revised: 17 March 2024

Accepted: 9 April 2024

Published: 25 April 2024

Keywords: EEG signals; emotion recognition; confusion; learning activities; machine learning; deep learning

1. Introduction

The COVID-19 pandemic, which led to educational institutes' longest closing in history and an impending recession, brought abrupt and consequential changes to every education system in the world. The normal instruction mode of delivery was disrupted, forcing them to transpose into the online mode. Virtual classrooms and MOOCs are significant online systems where usage dramatically increased during this time [1]. The transition was challenging to both the educator and the learner; however, many learners in the higher education systems now favor online learning as it provides flexibility to achieve a satisfactory work–study balance for them [1–3]. Yet, challenges remain still for the educator, such as the incapacity to assess learner emotions via online mode.

Education psychologists refer to the emotions stimulated during learning as “academic emotions”. Academic emotions can be of three types: Positive, Negative, and Mixed (i.e., both positive and negative). Positive emotions are pleasant feelings for learning like joy, zeal, and motivation while negative emotions denote unpleasant feelings that make a learner demotivated [2]. Confusion is an emotion that can be experienced by anyone during the learning process. According to educational psychologists, confusion is felt when a learner finds it difficult to align new knowledge with the existing knowledge stored in the



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

brain. This emotion is often utilized in the learning environment as a motivator; however, prolonged confusion levels among learners lead to frustration and boredom [1–3]. Due to this, educators need to thoroughly monitor learners’ confusion levels during their learning process. Various technologies that detect confusion emotion within a digital learning environment, such as learner analytics and sentiment analysis, are presently being utilized [3]. Bio-physiological signal approaches such as using Electroencephalography (EEG), Electrocardiogram (ECG), skin conductance, etc., are now becoming a popular technology over the aforementioned ones, as the latter can offer more accurate and reliable emotion dynamics over the former, which relies on indirect cues. As emotions are considered the source of most physiological responses, the validity and reliability of emotion recognition through these signals are enhanced [4–7]. Moreover, physiological signals can provide real-time emotional detection, which allows for immediate feedback in human–computer applications. Hence, many researchers in the emotion recognition domain prefer this approach. Out of the physiological signals, many researchers prefer EEG as it captures emotions from the source of generation, that is, the brain [2–4].

Confusion emotion detection among learners using EEG is still an emerging research domain, where the first publication was reported in 2013 [4]. However, when observing the trend of publications in this domain (Figure 1), this domain did not show intense popularity when compared to research conducted on other emotion recognition systems (Figure 2). This lesser popularity of the confusion emotion recognition domain was the problem that drove the authors to conduct this review with the intention of finding the root causes of this unwavering popularity. Hence, this systematic review was conducted to (i) have a state-of-the-art understanding of the current context; (ii) identify present research gaps and their plausible root causes; and (iii) explore future research directions to encourage potential research.

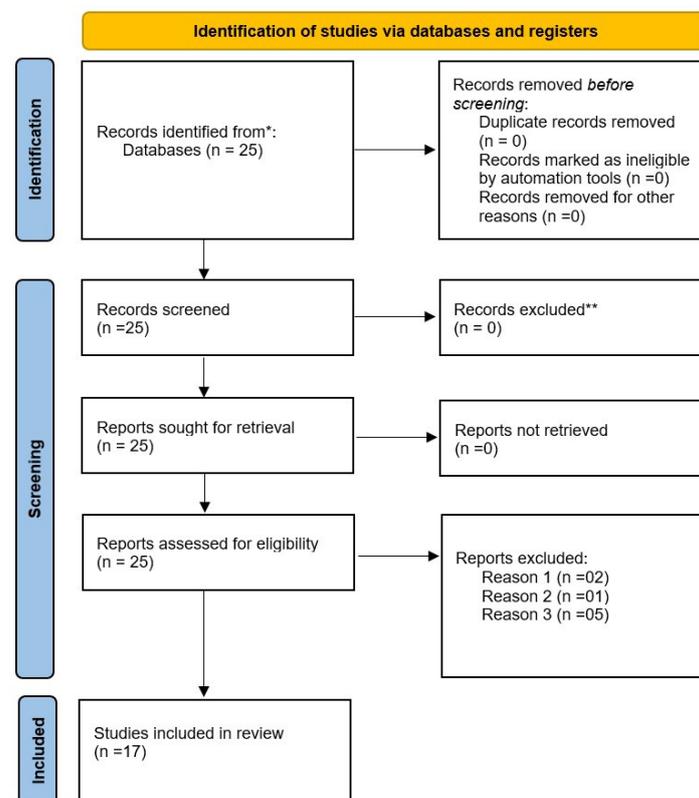


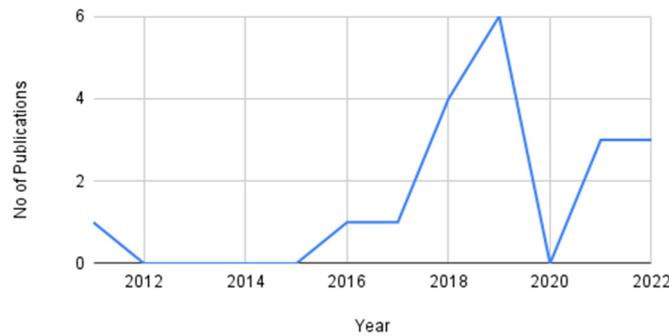
Figure 1. PRISMA workflow diagram of this study. * Indicates that the manuscript reports the number of records identified from each database which is listed in Table 1. ** refers to the number of articles excluded through title/abstract screening.

Table 1. Breakdown of obtained articles from the databases.

Research Database	No. of Extracted Literature
IEEE Xplore	06
ACM Digital Library	10
arXiv	4
Springer	3
Other	2



(a)



(b)

Figure 2. (a) Trend of Research publications in EEG-based emotion recognition (Source [8]). (b) Trend of Research publications in EEG-based confusion emotion recognition. (Source: Author).

The following Sections of the review comprise (i) a description of the methodology followed to conduct the review; (ii) a detailed synopsis of the published literature within the research domain concerning utilized datasets, feature preprocessing, feature types for model training, and deployed machine learning and deep learning classifier algorithms; (iii) a comparison of the prediction accuracies of the confusion emotion classifiers and illustration of the existing research gaps in the confusion emotion recognition systems—the algorithm comparison was made based on two categories, machine learning and deep learning; and (iv) providing conclusions that the recommendations would be ideal for future researchers to develop efficient systems for recognizing confused emotions in practical learning environments.

2. Methodology

The review was conducted by defining a review protocol that described the article selection and search strategy, article screening, data extraction, and critical evaluation.

In order to conduct the critical evaluation, three (03) Research Questions (RQ) were formulated as follows: RQ 01: What is the present context of EEG-based confusion emotion

recognition systems? RQ 02: What existing major research gaps are yet to be addressed and what are their associated research problems? RQ 03: Can the identified issues be resolved? If so, what could be the recommendations that can pave the way for future research?

The selection and screening of articles were performed as follows:

(a) Selection of Articles:

To minimize bias, all literature was selected from peer-reviewed data sources that were extracted using the Google Scholar web search engine from the following research databases. Table 1 illustrates the breakdown of obtained articles from the databases.

Accordingly, twenty-five (25) literature articles were identified. The authors did not use any automation tool for the selection. The following search strings were used to search literature from the above databases for selection: “Confusion emotion”, “recognition”, “detection”, “learning”, “online education”, “EEG”, “Machine learning”, “Deep Learning”, and “Artificial Intelligence”. The terms were even combined using Boolean operators such as ‘AND’ and ‘OR’ to broaden the search space and identify as many eligible articles as possible.

(b) Screening of Articles:

The selected articles were then screened using inclusion criteria: (i) articles written in the English Language; (ii) articles that exclusively presented research on confusion emotion recognition using EEG signals; (iii) Articles that presented work on confusion emotion recognition within learning/educational environments. Articles were then excluded due to the following reasons: Reason 01: Articles with only extended abstracts.

Reason 02: Articles with incomplete work, e.g., did not report/could not find results/findings of the study; Reason 03: Articles that had conducted research on confusion emotion recognition but not using (a) EEG signals, (b) EEG frequency domain features, or (c) were not about learners’ confusion emotion.

Data extraction was performed using the seventeen (17) screened articles. Data regarding the utilized datasets, feature preprocessing, feature types for model training, and deployed ML/DL algorithms and their performances were extracted. All data were recorded in evidence tables, which are presented in Tables 2–10. The authors then performed a critical evaluation of the data extracted, which is presented in the Discussion Section. A detailed description of the articles considered for the systematic review is listed in Appendix A.

Table 2. Timeline of the published literature in chronological order.

Year	Publication
2011	Wang et al. [4]
2016	Yang et al. [8]
2017	Nie et al. [9]
2018	Zhou et al. [10], Ibtehaz et al. [11], Tahmassebi et al. [12]
2019	Zhou et al. [13], Kumar et al. [14], Erwianda et al. [15], Wang et al. [16], Reñosa et al. [17],
2021	Dakoure et al. [18], Benlamine et al. [19], He et al. [20]
2022	Daghiri et al. [21], Abu-gellban et al. [22], Men et al. [23]

Table 3. Published datasets in ascending order (2011–2021).

Author	Year	Nature of Subjects	Stimuli	EEG Headset	No of/Type of Electrodes Used	EEG Sampling Rate	No. of Data Samples	Access
Wang et al. [4]	2011	10 college Students	MOOC videos	Single channel Neurosky Mindset	Single channel (frontal lobe Fp1 location)	512 Hz	12,811	Free and public use
Zhou et al. [10]	2018	16 college students	Raven’s tests	Emotiv Epoc+	16 channels 14 channels and 2 references	Not reported	Not reported	Unavailable for public use
Zhou et al. [13]	2019	28 23 college students and 5 Masters level students	Raven Test and Sokoban Cognitive Game	OpenBCI	10 channels 08 channels and 2 references	250 Hz	Not reported	Unavailable for public use
Dakoure et al. [18]	2021	10 CS * undergraduates	Cognitive ability tests	Emotiv Epoc	16 channels 14 channels and 2 references	128 samples/s	128	Unavailable for public use
Benlamine et al. [19]	2021	20 CS * undergraduates	3D adventure serious game	Emotiv Epoc	16 channels 14 channels and 2 references	128 samples/s	128	Unavailable for public use

* CS—Computer Science.

Table 4. EEG preprocessing techniques.

Technique	Overview
Down-sampling	This technique reduces the data in the EEG signal. For example, signals recorded at 512 Hz can be down-sampled into 128 Hz. This technique mostly opts for wireless transmission [9,10].
Filtering for Artifact removal	Artifacts are noises/disturbances recorded in the EEG signal. Artifacts can be internal such as eye blinks or external such as electrode displacement in the EEG headset. Independent Component Analysis and Regression are commonly used approaches for artifact removal [8–10].
Re-referencing	When obtaining EEG recordings, researchers place a reference electrode mostly at one Mastoid Cz. This will record voltages relative to another electrode. This technique can be performed by changing the reference electrode to another position. This is a good option when initial data recorded have not been collected without proper reference [10].

Table 5. The outline of Preprocessing Techniques used in Publications is listed in Table 3.

Publication	Preprocessing Technique
Zhou et al. [7]	Z-score standardization was used for EEG normalization to reduce individual differences in the signals.
Dakoure et al. [8]	Performed Artifact Removal using high pass filtering at 0.5 Hz and Low pass filtering at 43 Hz—the authors stated that the first filter was used as they wanted to deploy ICA. The second filter was used because ICA was sensitive to low frequencies and the headset used for EEG acquisition did not record signals above 43 Hz. ICA deployment. This algorithm breaks down the EEG signal into independent signals coming from particular sources.

Table 6. Domain Types of EEG Features.

	Feature Types	Significance to EEG Data Analysis
Frequency Domain (FD).	Statistical measures extracted from Power Spectral Density (PSD): energy, intensity weighted mean frequency, intensity weighted bandwidth, spectral edge frequency, spectral entropy, peak frequency, the bandwidth of the dominant frequency, power ratio [18]	PSD serves as the base calculation of this domain. The EEG series' power distribution over frequency is represented by PSD values. Computing PSD values are advantageous as neuroscientists believe they directly illustrate the neural activity of the human brain. [4,8,19,20].
	Relative PSD	The most frequently used frequency domain features in EEG signal analysis [16]. Relative PSD is the ratio of the PSD values of the frequency band to be analyzed to the total frequency band. This measure reduces the inter-individual deviation associated with absolute power. However, the accuracy for analyzing brain changes based on the non-stationary EEG signal is limited [18–21].
	Differential Entropy	It is the fundamental concept that quantifies the uncertainty/randomness of a continuous signal. It measures the amount of information consisting of a signal per unit time [22]. Differential entropy is the most used for emotion classification nowadays. However, this measure only considers relative uncertainty but does not calculate absolute uncertainty [22–25].
Time Domain (TD)	Statistical measures such as minimum and maximum values to quantify the range of data or the magnitude of signal baseline, mean, mode, variance, skewness, and kurtosis [15]	The time domain decomposes the raw EEG signal about time. It is assumed that statistical distributions can identify EEG seizure activities from normal activities [22–25].
	Hjorth Parameters	These are a set of statistical descriptors introduced by Bengt Hjorth in 1970. The descriptors describe temporal domain characteristics of mostly EEG and ECG signals. The parameters provide information about the mobility, activity, and complexity of the above signals [23–26].
Time-Frequency Domain (TFD)	Statistical measures such as mean, variance, standard deviation, absolute mean, absolute median	These features are estimated to differentiate EEG signal variations through statistical properties in each designated frequency sub-band for a specific time domain [18–20].
	Energy, Root Mean Square (RMS), and Average Power	The signal amplitudes that correlate to frequency sub-bands for a specific time domain are examined [18–20].
	Short-Time Fourier Transform (STFT)	The time is frequently modified by a fixed window function using the short-time Fourier transform (STFT). A number of brief-duration stationary signals are superimposed to form the non-stationary process. However, the low-frequency subdivision and the high-frequency temporal subdivision criteria cannot be satisfied [17].
	Wavelet Transform (WT)	The STFT local analysis capability is inherited by the Wavelet Transform (WT) method. WT has a higher resolution to investigate time-varying and non-stationary signals [17].

Table 7. Comparison of ML Architectures.

Reference	ML Algorithm/s	Type of Features	Prediction Performance	Significance
Wang et al. [4]	Naïve Bayes—Gaussian	Power Spectral Density (PSD) values	51–56%	The classifier was chosen as it works well with sparse and noisy training data. They stated that the difficulties in interpreting EEG data and their large dimensionality were found to have a negative impact on the accuracy of the classification, which was not significantly different from the applied educational researchers’ method of direct observation.
He et al. [26]	Naïve Bayes—Gaussian	PSD values	58.4%	They found that Random Forest and XG Boost are approximately 10% higher when compared with Naïve Bayes and KNN. Concluded that Naïve Bayes is not suitable for this dataset as the classifier makes an incorrect assumption that each value in the multi-dimensional sample has an independent impact on categorization.
	Naïve Bayes—Bernoulli		52.4%	
	k Nearest Neighbor (kNN)		56.5%	
	Random Forest (RF)		66.1%	
	XG Boost		68.4%	
Dakoure et al. [8]	Support Vector Machine (SVM)	PSD values	68.0%	They classified confusion levels into three (03) levels, which had not been previously done. They concluded that SVM is better suited for EEG classification than KNN with justifications stating that KNN does not abstract and learn patterns in data like SVM. It merely computed distances and formed clusters, and the instances of the same cluster are close in the feature space.
	kNN		65.2% (k = 20)	
	kNN		54.86% (n = 1)	
Kumar et al. [27]	Logistic Regression CV	PSD values	53.88%	The authors employed 32 supervised learning algorithms with different parameter settings. (Algorithms with the highest accuracies are only displayed here.) They concluded that Random Forest with Bagging had the highest accuracy. However, they did not provide justifications for their conclusions. They recorded universal-based models to improve the accuracy.
	Linear Discriminant Analysis		53.38%	
	Ridge Classifier		53.38%	
	SVM		55.18%	
	RF with Bagging		61.89%	
	XG Boost		59.21%	
	XG Boost		82%	
Erwianda et al. [28] 2019	XG Boost-Recursive Feature Elimination (RFE)	PSD values	83%	RFE was utilized as the feature selection technique and TPE as the hyperparameter optimization technique. The study revealed that the best features were Theta, Delta, and Gamma-2 for confusion detection. However, an explanation for this fact was not provided.
	XG Boost-RFE—Tree-Structured Parzan Estimator (TPE)		87%	
Daghri et al. [29]	Gradient Boosting (GB)	PSD values	100%	A novel feature engineering approach was proposed to produce the feature vector. Class probabilities from RF and GB were utilized to develop the feature vector. Results indicated that 100% accuracy was obtained via this approach. Further, it stated that the DL models that were trained did not perform well when compared with ML. The reason stated was the dataset was too small for better performance.
	RF			
	Support Vector Classifier (SVC)			
	Logistic Regression (LR)			
Yang et al. [30]	SedMid Model	EEG time series data and data extracted from audio and video features of the lectures	87.76%	The authors introduced a new model named Sequence Data-based Mind-Detecting (SedMid). First of a kind that detected confusion levels by combining other sources. The Sedmid model developed mixed time series EEG signals with audio–visual features.

Table 7. Cont.

Reference	ML Algorithm/s	Type of Features	Prediction Performance	Significance
Benlamine et al. [9]	kNN	EEG time series	92.08% ($k = 5 \approx \log$ (number of samples))	A novel approach where the researchers attempted to recognize confusion emotion in learners in a multi-class classification approach rather than a binary classification. At fixed time frames, facial data showing the confusion levels of each participant were recorded. The corresponding EEG signals were recorded. Finally, separate EEG recordings were obtained for high, medium, and low levels of confusion. The Model was trained using SVM and KNN; however, they preferred SVM due to (a) Popular in BCI research; (b) Robustness against nonlinear data; (c) Efficient in high dimensional space. The authors did not review why KNN produced higher results, although they stated that KNN is not suitable for EEG classification.
	SVM		92.08%	

Table 8. Comparison of prediction accuracies using ML and DL algorithms.

Reference	Deployed Algorithm		Prediction Accuracy
Benlamine et al. [19]	ML	kNN	92.08%
	DL	LSTM *	94.8%
He et al. [20]	ML	XGBoost	68.4%
	DL	LSTM *	78.1%

* LSTM—Long Short-Term Memory.

Table 9. Comparison of DL Architectures.

Reference	DL Algorithm/s Deployed	Prediction Performance	Significance
Zhou et al. [13]	Convolutional Neural Network (CNN)	71.36%	This study selected CNN as it accepts raw data out of several channels as direct inputs, reducing the need to transform the EEG raw data into the standard frequency bands and the process subsequent feature extraction method.
Ni et al. [9]	CNN	64.0%	This study was conducted to improve the accuracy of the dataset [4]. Concluded that CNNs and DBNs are not a good choice for this dataset as there is a high chance of overfitting. They recommended LSTMs, although different time steps (i.e., a feature for every 0.5 s) supplied to the LSTM share the same weights in the neural network, the forget gate can learn to utilize previous hidden states in the LSTM. Bi-directional LSTMs employ sequential data from both directions to learn a representation in both directions including context data, which enables a more reliable and accurate model.
	Deep Belief Network (DBN)	52.7%	
	LSTM	69.0%	
	Bi-LSTM	73.3%	
Wang et al. [16]	Confounder Bi-LSTM (CF-Bi-LSTM)	75%	This study was conducted to improve the accuracy of the dataset of [4] by the same authors. They introduced the Confounder Filtering method that reduces confounders and improves the generalizability of the deep neural network and concluded that this approach improves the performance of bioinformatics-related predictive models.
Abugellban et al. [22]	CNN with Rectified Linear Unit (ReLU)	98%	This study addressed the performance issues of CDL 14, which had ignored the demographic information of the students. Concluded that demographic information is naturally influenced by confusion emotion detection.
Reñosa et al. [17]	Artificial Neural Network (ANN)	99.78%	The research attempted to classify confusion levels of students as a percentage using combined averaged power spectra of all frequency bands and the standard deviation of each frequency band as inputs for the ANN.
Zhou et al. [13]	05 layered CNN Adaptive Moment Estimation (ADAM) for optimization	91.04%	The authors provided a detailed description of how they conducted the human experiment to collect EEG recordings by inducing confusion through game-based learning. They utilized Adam Optimization to speed the gradient descent process and remove excessive swings during the CNN model training process. They highlighted that the binary classification of the confusion emotion state is preliminary since confusion is a complex emotion state. They recommended multi-class classification as a good alternative.

Table 9. Cont.

Reference	DL Algorithm/s Deployed	Prediction Performance	Significance
Men et al. [23]	LSTM	77.3%	This paper addresses the issue of an overfitting problem with LSTM and Bi-LSTM algorithms by introducing an attention layer. This is a technique in neural networks that attempts to resemble cognitive attention. The authors stated that the attention layer achieved good accuracy in textual and image classification models. They utilized two DL algorithms: LSTM and Bi-LSTM but were not satisfied with either model as they had not met their expectations. They concluded that LSTM had heavy overfitting and Bi-LSTM did not learn features from the training data.
	LSTM + Attention Layer	81.7%	
	Bi-LSTM	67.9%	
	Bi-LSTM + Attention Layer	69.7%	
Ibtehaz et al. [11]	CNN + Logistic Regression	81.88%	The authors proposed a new algorithm where spectral features of EEG signals are fused with the temporal features. Spectral features were extracted using CNN and classification was performed through ML algorithms. Thus, this study implemented a hybrid ML and DL model. The features for the classifier algorithms were the activation map received from the Global Average Pooling layer. The authors concluded that their approach yielded higher accuracies since CNN extracts better features for classification rather than manually picking features for model training.
	CNN + Naïve Bayes—Gaussian	77.98%	
	CNN + Support Vector Machine	78.075%	
	CNN + Decision Trees	73.922%	
	CNN + Random Forest	79.88%	
	CNN + k Nearest Neighbor (kNN)	82.92%	
Tahmassebi et al. [12]	CNN + AdaBoost	80.79%	This study aims to classify confused learners from non-confused learners using Genetic Programming. They developed models from the NSGA-II algorithm and multi-objective genetic programming approach. Based on the fitness and complexity measures, the authors defined three models of which the third—knee model—produced the highest accuracy. They concluded that this model can be a good substitute for traditional algorithms as it has an 80% shorter computational run time.
	NSGA II—knee model	73.96%	

3. Results

The authors observed that in comparison to the growing popularity of emotion recognition systems using EEG, confusion emotion recognition has less popularity. The publication frequency of this research domain shows an upward trend from 2017 and a steep downward trend during the year 2020 while increasing again in 2021. It can be speculated that the global pandemic situation would have had an impact on this research domain in 2020.

Of the articles published between 2013 and 2023 (Table 2), only five of the researchers were found to have produced their own datasets by conducting human tests. Others focused on achieving higher prediction accuracies using shallow learning and deep learning techniques and compared different algorithms. Table 2 shows the timeline of the published literature.

3.1. Datasets

Only five EEG databases or datasets linked to confusion emotion recognition in a learning environment have been published up until this point (Table 3). The datasets mainly took three approaches to induce confusion among their learners: (i) Wang and his group of researchers [4] used an approach where the students were presented with new information via brief videos. They made their participants self-learn new concepts. Some videos had content removed in between to increase confusion; (ii) Zhou et al. [6,7] and Dakoure et al. [8] made their participants solve logical reasoning problems to induce confusion. They used famous cognitive ability tests (Raven, and Sokoban tests) as a source of a confusion emotion-inducing stimulus [6–8]; and (iii) Benlamine et al. [9] induced confusion in their participants by making them play a serious adventure computer game, yet the authors had a minimal explanation of how they induced confusion via the game environment in cognitive terms.

The below table provides a detailed summary of those databases/datasets.

It was noted that only [4] is publicly accessible and can be downloaded from [5]. It is the only dataset that encourages further research in confusion emotion recognition within an educational setting, which is the rationale behind the extensive citations of this dataset in numerous studies that did not gather EEG data. This led to the emergence of a research

community where numerous studies evaluated and compared the machine learning and deep learning predictions made using this dataset.

3.2. Approaches for EEG Preprocessing

Preprocessing in the context of EEG data typically refers to reducing noise from the signals to identify the relevant signals. Preprocessing of EEG data is crucial for many important reasons: (i) since spatial information is lost when EEG is acquired from the device that is attached to the scalp, the EEG signals may not accurately reflect the signals coming from the brain; (ii) weaker EEG signals can be hidden by the noise that EEG data frequently contain, e.g., blinking or muscle movement artifacts can taint the data and distort the image; and (iii) pertinent EEG signals can be separated from random EEG signals [10–12].

There are no standard EEG preprocessing techniques since EEG preprocessing is still an active area of research [12]. However, many researchers working with EEG data commonly use down-sampling, re-referencing, and artifact removal as mandatory preprocessing techniques. A short summary of these techniques is provided in Table 3.

The preprocessing techniques that were followed by the researchers who created the EEG datasets related to learner confusion emotion are as follows:

It was observed that only two publications discussed their preprocessing method; it was not mentioned even in the widely used, publicly available dataset [4]. The accuracy of the EEG data in the dataset is, hence, questionable as to whether the artifacts had been removed and whether the dataset contains only the relevant signals related to confusion emotion.

3.3. Types of EEG Features

A feature is a unique or distinguishing measurement, transformation, or component that is taken from a pattern's segment. During the feature extraction process, a feature vector is created from a regular vector [13]. An efficient feature extraction pipeline needs to be followed during emotion recognition from EEG due to the weak, nonlinear, and time-varying nature of the signals [14]. Features from the EEG signals are classified generally into three categories: frequency domain (FD), time domain (TD), and time–frequency domain (TFD). The FD represents the amplitudes of EEG signals in their relevant frequency domain. This provides valuable insights into the underlying characteristics such as dominant frequencies, harmonics, and other spectral features of the signal. However, transforming a signal into its FD results in a loss of time information. The drawback is that this domain assumes the signal is stationary, meaning that the statistical properties of the signal remain constant over time. Real-world signals are mostly non-stationary. This leads to misinterpretations about the signal [15].

The time domain describes how EEG signals vary with time. TD analysis has the capacity to simulate all nonlinearities; however, the computation can be time-consuming [16].

Features in the time–frequency domain (TFD) have now widely been used in EEG signal processing as TFD accurately analyzes non-stationary waves. TFD combines information from the time and frequency domains and simultaneously allows for localized analysis in the time–frequency domain. This way, the signal's temporal information is not lost [17].

The feature categories commonly used from each domain are summarized in Table 6.

It was observed that the majority of the researchers utilized PSD values in the frequency domain for implementing the models. Time domain features were rarely used, and time–frequency domain features have never yet been used for confusion emotion recognition. A detailed summary of the features utilized by each researcher is illustrated in Table 7.

Wang et al.'s [4] dataset, which many of the researchers used to conduct their own research, consists of EEG data in the form of PSD values. They stated that when training models use features in the time series domain, the models tend to overfit. They assumed that this was caused by there not being enough EEG data samples in the dataset.

Yang et al. [8] and Benlamine et al. [19] used time domain features for model training. Ref. [8] introduced a new algorithm model that analyzes time series features together with data from audio and video features of lectures to detect learners' confusion states. Ref. [19] trained existing ML algorithms.

3.4. Machine Learning (ML) Architectures and Performance Comparison

ML algorithms have been the most popular among researchers in this domain. The most common algorithm type deployed was supervised classification such as Support Vector Machines, Naïve Bayes, Random Forest, k Nearest Neighbor, and Gradient Boosting. Unsupervised learning and ensemble learning were found to be infrequently used in the studied literature Table 7 summarizes all the ML algorithms that had been utilized.

3.5. The Need for Deep Learning

The literature says that deep learning (DL) networks can extract more distinct and interpretable features from EEG signals [23,31]. EEG signals tend to have a poor signal-to-noise ratio, which might make it difficult to distinguish important signals from other noisy features like artifacts. Also, they are complex, meaning that the neural oscillations behave nonlinearly and are very dynamic. DL algorithms are designed in such a way that they progressively extract high-level features from EEG signals by using multiple processing layers to learn representations of the data [31–34]. This indicates that DL models also work effectively with inputs not used for training the model. Due to this, these models can efficiently work with inputs from real-world data samples. Traditional ML models do not have this ability, and thus, tend to overfit [35–37]. This was observed even in the publications in this domain, where many authors stated that the ML models tended to overfit.

Table 8 shows research that compared ML and DL algorithms in their studies. Benlamine et al. [9] and He et al. [26] compared the prediction accuracy of ML and DL classifiers and concluded that DL algorithms are much more accurate (Table 8). Yet, neither offered a convincing justification as to the reason why DL algorithms offered better accuracy than ML algorithms.

It was noted that many studies relied heavily on Convolutional Neural Networks (CNNs). This algorithm, even though it is a classifier, is deployed in the domain of emotion recognition to extract features from raw signals. The convolutional layers' activation maps before the final fully connected layer are used as the features for training the model [37–40]. Long Short-Term Memory (LSTM) is the second most used DL algorithm by the researchers. This is an extended form of a Recurrent Neural Network (RNN), which has the ability to learn long-term dependencies in time series data. LSTM was used mostly to determine the correlations in the EEG feature vector [39–43]. Bi-directional Long Short-Term Memory (Bi-LSTM), which consists of two LSTM algorithms that take inputs in forward and backward directions, can learn feature representations within a shorter time span. Researchers also state that Bi-LSTMs are better at learning emotional information than LSTMs [43–47]. The next Section provides a detailed synopsis of the DL algorithms used so far in this domain.

Table 10. Summary of the identified research gaps and the suggested recommendations.

Research Gap	Associated Research Problem	Suggested Research Works for Future
Bias–Variance Dilemma in existing Confusion Emotion Models	Model having hindered performance. The model’s prediction performance is unreliable for new and unseen data.	Metadata such as demographics play an important role in human emotional states [48]. Hence, they can be considered as inputs for model training. Use advanced EEG feature preprocessing techniques to remove noise in EEG data. Use advanced feature selection techniques to identify the optimal EEG features that can predict confusion emotion. Construct models having bias–variance tradeoff.
Limited research on the development of cost-sensitive confusion emotion models	Good-performing models are unaffordable for end users.	Research on constructing cost-sensitive models having a tradeoff between cost and model performance.
Limited datasets	Unable to test developed models on independent datasets for their prediction reliability on new and unseen data.	Conduct human experiments to collect EEG recordings and create datasets.
Low Generalizability in existing confusion emotion models	Models would not perform well if new data having different levels of complexities were provided.	Collect EEG recordings among a broader demographic population of learners. Test the models on EEG recordings from real-world populations rather than recordings obtained from a laboratory experiment setup.

3.6. Deep Learning Architectures and Performance Comparison

The literature shows that the trend for research on confusion emotional detection is now shifting from ML to DL, and research publications during the last two years show the higher prediction accuracies of DL that ML failed to achieve. We also observed that much of the research focuses on finding how to improve DL algorithms to achieve better accuracies. Table 9 reviews the current research that deploys DL algorithms for confusion emotional state detection.

4. Discussion and Future Directions

EEG-based confusion emotion recognition of learners is an emerging research domain. The authors reviewed seventeen articles that were published in this research domain from its first publication [5] to the year 2023 and were able to extract trends and highlight major issues that the present domain experienced, mainly in the sections of datasets, widely used EEG features, and the ML and DL algorithms utilized for confusion emotion recognition.

The research questions were answered as follows:

4.1. Answering Research Question One

“What is the present context of EEG-based confusion emotion recognition systems?”

It was observed that this research domain is still limited and growing at a slow pace due to the smaller number of publications (only seventeen) that were identified till the present. Only five research works [5,16,19,24,25] were dedicated to creating datasets and of the five, only one dataset [5] is still available publicly. Other publications comprised research works that built confusion emotion classifiers using Wang et al.’s [5] dataset (Figures 3 and 4).

A crucial question lies concerning the slow pace of this research domain. Although deriving a specific and conclusive answer cannot be reached, it was assumed that this is occurring due to the following reasons:

- (a) Limited creation of datasets that collect EEG recordings by conducting human experiments. This is due to the following reasons: (i) difficulty in stimulus design for inducing confusion within an educational setting; (ii) unavailability of standardized psychological procedures for deliberate induction of confusion through learning activities; and (iii) practical problems and challenges relating to EEG acquisition remain, such as the need for the necessary technological infrastructure, legal approvals, time, and financial commitment for the acquisition of a significant number of EEG recordings from participants [46].

- (b) Limited publicly accessible datasets: datasets that are publicly accessible help overcome obstacles related to EEG acquisition. The unavailability of datasets may demotivate those researchers who are interested in conducting research in this domain yet cannot overcome the above obstacles. The availability of a wide range of datasets offers diversified opportunities to analyze EEG data and extract uncovered information that would promote more research. Hence, in any study field, the increased availability of datasets is just as crucial as developing the datasets themselves.

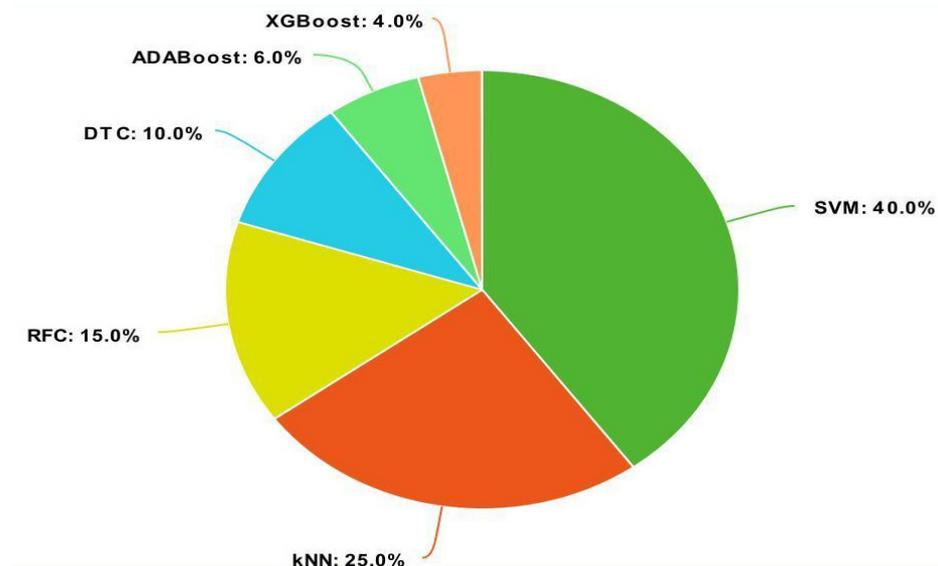


Figure 3. Most popular machine learning algorithms used in existing literature.

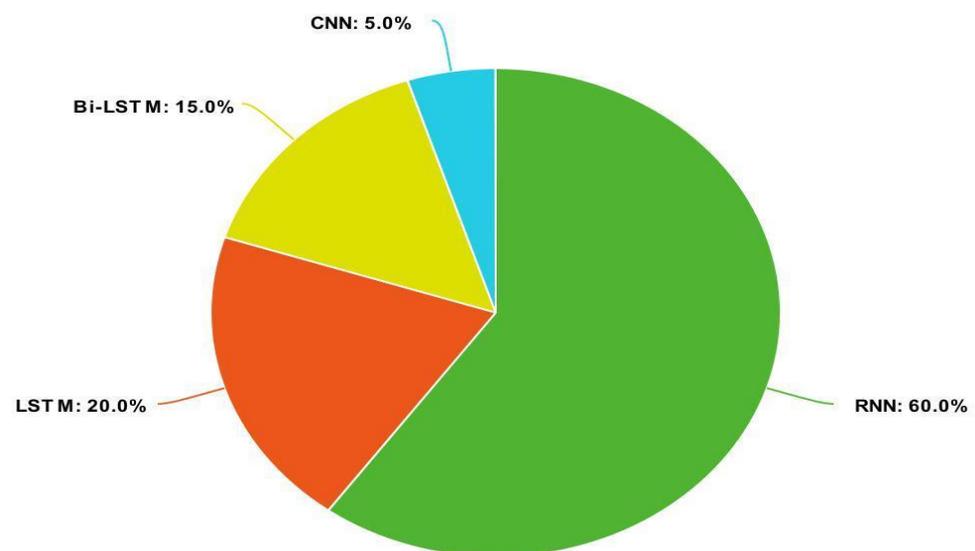


Figure 4. Most popular deep learning algorithms used in existing literature.

The following can be summarized regarding the present context of the research domain: (i) EEG signals for the analysis were captured either from the frontal side of the brain using a single-channel EEG headset or from different parts of the brain via multi-channel EEG headsets; (ii) after acquisition, the EEG signals were preprocessed mainly using the Fast Fourier transformation (FFT) method; and (iii) the preprocessed signals were then classified using supervised ML or DL classification algorithms. The most commonly deployed ML and DL algorithms are shown with their usage percentages in Figures 5 and 6, respectively.

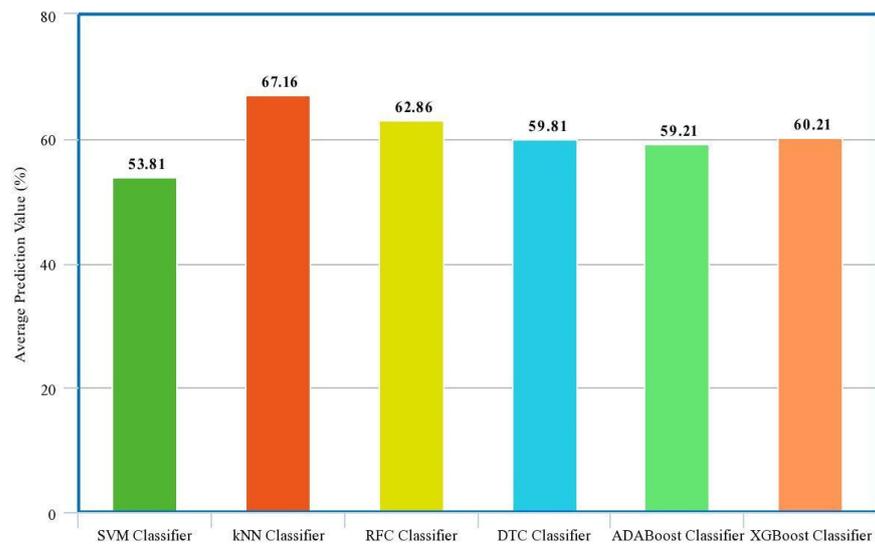


Figure 5. Averaged prediction accuracies of popular machine learning algorithms used in previous literature.

All works utilized supervised binary classification ML/DL algorithms to identify confused learners from non-confused learners. The average prediction accuracies for the most common machine learning and deep learning algorithms that were trained using Wang et al.’s dataset [5] are illustrated in Figures 3 and 4, respectively.

The tendency to use DL algorithms has now increased more than ML. The literature published closer to the present date mainly studies applying different DL algorithms to improve the accuracy of detection, but the cost of ML and DL algorithms were not compared. When considering the nature of the input features used for model training, the features of EEG signals were independently extracted and combined afterward into a single feature vector. The input features were either Power Spectral Density (PSD)/EEG frequency ranges or statistical measures of the EEG signals.

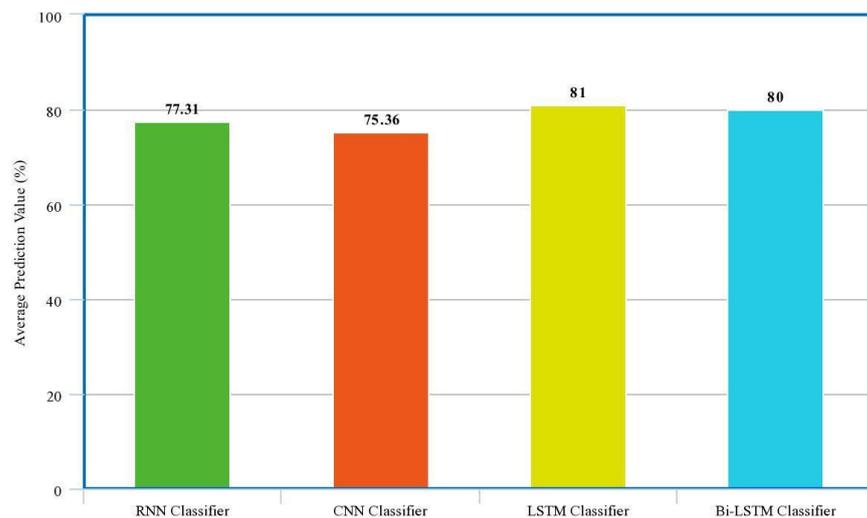


Figure 6. Averaged prediction accuracies of popular deep learning algorithms used in previous literature.

4.2. Answering Research Question Two

“What existing major research gaps are yet to be addressed and what are their associated research problems”?

The following major research gaps were revealed that had not yet been addressed.

Research gap 01—Bias–Variance Dilemma in existing confusion emotion models: The bias–variance dilemma is a conflict between model bias and variance where model variance

increases when bias is being reduced, and vice versa [49,50]. This conflict prevents the model's ability to learn beyond the training dataset, thus reducing the reliability and prediction accuracy of the model [51]. This dilemma causes the model either to overfit (provides accurate predictions for training data but not for new data) or underfit (generates erroneous predictions for both new and unseen data).

The authors observed that works by [5,20–22,26] developed models that had a prediction accuracy within the range of 50% and [20,24,27] had models that had prediction accuracies between 90% and 100%. The high variation in the reported ML prediction accuracies concerned the author as to why it might happen. It was assumed that this could be due to the bias–variance dilemma. The assumption was supported by the concluding remarks of [27], which stated that achieving 100% accuracy levels might be due to overfitting. Another solid support for the above assumption was provided by [21]. Their initial model developed using XGBoost had an accuracy level of 52%. In order to boost performance, they optimized the model using Tree Parzan Estimator (TPE) and boosted the performance to 58%. However, the authors did not mention the bias–variance dilemma. However, the theory on the bias–variance dilemma states that model optimization using TPE or hyperparameter tuning algorithms reduces this dilemma resulting in a boost in the prediction accuracy of the model if underfitted [50].

A critical root cause for the bias–variance dilemma could be the quality of the dataset [47]. If the dataset consists of noise and is not preprocessed sufficiently, this would lead to the creation of this dilemma [48]. The above reason could be the reason for the dilemma experienced by the models in [52]. Both studies utilized the dataset of Wang et al. [5]. It did not describe feature preprocessing in a detailed manner. Since EEG recordings are highly susceptible to noises and artifacts, a proper preprocessing pipeline is required for better accuracy of the EEG recordings. Hence, the quality of the EEG data samples in the dataset is questionable, and it is suspected that this lower quality led to the creation of the bias–variance dilemma.

An associated research problem for this could be that by not including an examination of the bias–variance dilemma in the machine learning development pipeline, the developed models would become affected by the dilemma, therefore causing the model to either overfit or underfit, and thus making the predictions for new and unseen data unreliable.

The authors only focused on the bias–variance dilemma in machine learning algorithms. Deep learning algorithms were given less focus as the authors intended to bridge the research gaps that were in machine learning algorithms in order to achieve the aim of the study.

Research gap 02—Limited research on the development of cost-sensitive confusion emotion models: It was revealed that models developed with DL algorithms had better accuracy than ML algorithms (Tables 7 and 9). However, DL models are much more expensive than ML models as they are computationally expensive [53] and heavily dependent on computing power, requiring powerful hardware for processing. Although the performance of DL models is higher, their operational costs mean that it is unaffordable for end users to utilize DL models in their day-to-day activities [54]. This is a problem when designing models that fit real-world use. Since the end users of the models developed in this study are educators, it is crucial to develop cost-effective models that perform well. The above type of models can be designed utilizing the theories of cost-sensitive machine learning [50]; however, these have not yet been utilized in existing works till the present day, to the best of the authors' knowledge.

Research gap 03—Limited diversified datasets comprising EEG recordings of learners engaged in diversified learning activities: It was revealed that till the present, there are only four datasets that consist of EEG data samples related to learners' confusion levels. Of the four datasets, the only dataset publicly available is Wang et al.'s [5] dataset. This dataset comprises 12,811 data samples that are utilized for model training and testing in existing works that have built models using various ML and DL algorithms.

This dataset has its own issues that are identified as follows: (i) It comprises data extracted from a well-defined sample: a group of ten college students whose subject majors

were not specified. They were asked to learn from MOOC videos with topics; however, they failed to specifically mention the subjects covered in the videos. They stated that their audience would be confused; however, they did not justify the confusion induction stimulus. When a well-defined population is used to extract training data, ML and DL algorithms are prone to overfit, and the generalizability of the model is influenced by the EEG recordings of the human subjects who participated in the experiment [49]; (ii) Metadata on demographics of the participants such as age group, gender, ethnicity, and psychological factors during the time of the experiment are not reported. Metadata provides complete and detailed information about the population on which the algorithms were trained, tested, and validated. This is crucial for assuming that algorithm performance can be extrapolated to different populations [49]. Also, integrating metadata into algorithm development would have implications for increased classification accuracy as it would closely reflect the practice of an educator who analyzes students not only through cognitive perspectives but also through demographics [49].

The unavailability of another dataset similar to that of [5] leads to a problem, whereby the developed models cannot be tested for their levels of generalizability, which requires new and unseen data.

Research gap 04—Existing confusion emotion models were not tested on new, previously unseen data: It was revealed that none of the research works reported evaluating the trained model performance within a real-world population. This could be due to: (i) unavailability of datasets other than the dataset used for model development or (ii) resource unavailability for the authors to test the developed system on new and unseen data. The trained models were evaluated against the data from the dataset itself. The majority of the evaluation utilized either a 5-fold or 10-fold evaluation technique to predict the model's ability to predict given new instances. Although this approach tests the model's prediction performance, which was trained on a particular complexity, it does not test the performance of the model if it is given data having different complexities than the one trained on (e.g., data consisting of similar patterns but not the same as those of the training dataset) [9]. Due to this, it is questionable whether the trained models would have good performance if given independent datasets.

4.3. Answering Research Question Three

“Can the identified research gaps be filled? If so, what could be the recommendations that can pave ways for future research?”

Yes, the identified gaps can be filled through technological solutions. Table 10 presents a summary of the identified research gaps, their associated problems, and the suggested works to bridge the gaps as future work in this research domain.

5. Conclusions

Confusion is an emotion state experienced frequently during the learning process. Educational psychologists believe that this emotion stimulates when a learner finds it difficult to align new knowledge with the existing knowledge stored in the brain [2]. This review paper focused on the works conducted in the research domain of electroencephalography (EEG)-based confusion emotion recognition using artificial intelligence among learners. It was identified that this domain is still in its infancy stage. The first research work conducted was in 2013 by Wang et al. [4]. Of the articles published between 2013 and the year this review was conducted (2023), there were only 17 research works published. This review paper focused on all the 17 studies on confusion emotion recognition of learners that were conducted. This review provided a detailed summary of these research works with reference to utilizing the EEG features, features extraction methods, system performance, and algorithms for confusion emotion recognition in learners. Algorithm comparison was performed based on two categories: ML and DL. A comparison table of the models built, performance graphs, and information about publicly available datasets is also provided. Major research gaps in this domain were identified and are explained elaborately in this

review concerning their possible root causes. The authors provided recommendations to bridge the research gaps and believe that the recommendations would be an ideal way for future researchers to develop efficient systems for the recognition of confusion emotion in practical learning environments.

Author Contributions: Conceptualization, D.G., I.K. and M.W.P.M.; methodology, D.G.; software, D.G.; validation, M.W.P.M., I.K. and V.T.; formal analysis, D.G.; investigation, D.G.; resources, D.G. and I.K.; data curation, D.G. and M.W.P.M.; writing—original draft preparation, D.G.; writing—review and editing, M.W.P.M., I.K. and V.T.; visualization, M.W.P.M., I.K. and V.T.; supervision, M.W.P.M., I.K. and V.T.; project administration, V.T.; funding acquisition, M.W.P.M. and V.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

Appendix A

Table A1. Publications selected for Systematic Review.

No.	Year of Publication	Authors	Name of Journal/Conference Proceedings	Title	Citation
1	2013	Wang, Haohan, Yiwei Li, Xiaobo Hu, Yucong Yang, Zhu Meng, and Kai-min Chang	AIED Workshops	Using EEG to Improve Massive Open Online Courses Feedback Interaction	[5]
2	2016	Yang, Jing Kang, Haohan Wang, Jun Zhu, and Eric P. Xing	arXiv preprint arXiv:1611.10252	Sedmid for Confusion Detection: Uncovering Mind State from Time Series Brain Wave Data	[8]
3	2017	Ni, Zhaoheng, Ahmet Cem Yuksel, Xiuyan Ni, Michael I. Mandel, and Lei Xie	8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics	Confused or not confused? Disentangling brain activity from EEG data using bi-directional LSTM recurrent neural networks.	[9]
4	2018	Yun Zhou; Tao Xu; Shiqian Li; Shaoqi Li	2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)	Confusion State Induction and EEG-based Detection in Learning	[10]
5	2021	Nabil Ibtehaz, Mahmuda Naznin	NSysS'21: Proceedings of the 8th International Conference on Networking, Systems and Security	Determining confused brain activity from EEG sensor signals	[11]
6	2018	A. Tahmassebi, A. H. Gandomi and A. Meyer-Baese	IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, Brazil	An Evolutionary Online Framework for MOOC Performance Using EEG Data	[12]
7	2019	Zhou, Yun, Tao Xu, Shaoqi Li, and Ruifeng Shi	Universal Access in the Information Society	Beyond engagement: an EEG-based methodology for assessing user's confusion in an educational game	[13]
8	2019	Bikram Kumar, Deepak Gupta, Rajat Subhra Goswami	International Journal of Innovative Technology and Exploring Engineering	Classification of Student's Confusion Level in E-Learning using Machine Learning	[14]
9	2019	Erwianda, Maximillian Sheldy Ferdinand, Sri Suning Kusumawardani, Paulus Insap Santosa, and Meizar Raka Rimadana.	International Seminar on Research of Information Technology and Intelligent Systems	Improving confusion-state classifier model using xgboost and tree-structured parzen estimator	[15]
10	2019	Wang, Yingying, Zijian Zhu, Biqing Chen, and Fang Fang.	Cognition and Emotion volume 33, no. 4.	Perceptual learning and recognition confusion reveal the underlying relationships among the six basic emotions.	[16]
11	2019	Claire Receli M. Reñosa; Argel A. Bandala; Ryan Rhay P. Vicerra	2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)	Classification of Confusion Level Using EEG Data and Artificial Neural Networks	[17]
12	2021	Dakoure, Caroline, Mohamed Sahbi Benlamine, and Claude Frasson	International FLAIRS Conference Proceedings, vol. 34	Confusion detection using cognitive ability tests	[18]

Table A1. Cont.

No.	Year of Publication	Authors	Name of Journal/Conference Proceedings	Title	Citation
13	2021	Benlamine, Mohamed Sahbi, and Claude Frasson	In Intelligent Tutoring Systems: 17th International Conference, Proceedings. Springer International Publishing	Confusion Detection Within a 3D Adventure Game	[19]
14	2021	He, Shuwei, Yanran Xu, and Lanyi Zhong	IEEE 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE)	EEG-based Confusion Recognition Using Different Machine Learning Methods	[20]
15	2022	Daghriri, Talal, Furqan Rustam, Wajdi Aljedaani, Abdullateef H. Bashiri, and Imran Ashraf	IEEE Electronics volume 11, no. 18	Electroencephalogram Signals for Detecting Confused Students in Online Education Platforms with Probability-Based Features	[21]
16	2022	Hashim Abu-gellban; Yu Zhuang; Long Nguyen; Zhenkai Zhang; Essa Imhmed	2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)	CSDL EEG: identifying confused students based on EEG using multi-view deep learning	[22]
17	2022	Xiuping Men, Xia Li	International Journal of Education and Humanities	Detecting the confusion of students in massive open online courses using EEG	[23]
18	2022	Na Li, John D. Kelleher, Robert Ross	arXiv preprint:2206.02436	Detecting interlocutor confusion in situated human-avatar dialogue: a pilot study	REJECTED (Reason 03 of Exclusion Criteria)
19	2022	Na Li, Robert Ross	arXiv preprint:2206.01493	Transferring studies across embodiments: a case study in confusion detection	REJECTED (Reason 03 of Exclusion Criteria)
20	2022	Na Li, Robert Ross	arXiv preprint:2208.09367	Dialogue policies for confusion mitigation in situated HRI	REJECTED (Reason 03 of Exclusion Criteria)
21	2022	Anala, Venkata Ajay Surendra Manikanta, and Ganesh Bhumireddy.	A thesis is submitted to the Faculty of Computing at Blekinge Institute of Technology	Comparison of machine learning algorithms on detecting the confusion of students while watching MOOCS	REJECTED (Reason 03 of Exclusion Criteria)
22	2023	Li, Na, and Robert Ross.	Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction, pp. 142–151. 2023	Hmm, You Seem Confused! Tracking Interlocutor Confusion for Situated Task-Oriented HRI	REJECTED (Reason 03 of Exclusion Criteria)
23	2023	Rashmi Gupta; Jeetendra Kumar	5th IEEE Biennial International Conference on Nascent Technologies in Engineering (ICNTE)	Uncovering of learner's confusion using data during online learning	REJECTED (Reason 03 of Exclusion Criteria)
24	2023	Tao Xu, Jiabao Wang, Gaotian Zhang, Ling Zhang, and Yun Zhou	Journal of Neural Engineering	Confused or not: decoding brain activity and recognizing confusion in reasoning learning using EEG	REJECTED (Reason 02 of Exclusion Criteria)
25	2019	Borges, Niklas, Ludvig Lindblom, Ben Clarke, Anna Gander, and Robert Lowe	2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)	Classifying confusion: autodetection of communicative misunderstandings using facial action units	REJECTED (Reason 01 of Exclusion Criteria)
26	2020	F Ibrahim, S Mutashar and B Hamed	IOP conference series: materials science and engineering, volume 1105	A Review of an Invasive and Non-invasive Automatic Confusion Detection Techniques	REJECTED (Reason 02 of Exclusion Criteria)

References

- Bhattacharya, S.; Singh, A.; Hossain, M. Health system strengthening through massive open online courses (moocs) during the COVID-19 pandemic: An analysis from the available evidence. *J. Educ. Health Promot.* **2020**, *9*, 195. [CrossRef] [PubMed]
- Ganepola, D. Assessment of Learner Emotions in Online Learning via Educational Process Mining. In Proceedings of the 2022 IEEE Frontiers in Education Conference (FIE), Uppsala, Sweden, 8–11 October 2022; pp. 1–3. [CrossRef]
- El-Sabagh, H.A. Adaptive e-learning environment based on learning styles and its impact on development students' engagement. *Int. J. Educ. Technol. High. Educ.* **2021**, *18*, 53. [CrossRef]
- Wang, H.; Li, Y.; Hu, X.S.; Yang, Y.; Meng, Z.; Chang, K.K. Using EEG to Improve Massive Open Online Courses Feedback Interaction. *CEUR Workshop Proc.* **2013**, *1009*, 59–66. Available online: <https://experts.illinois.edu/en/publications/using-eeg-to-improve-massive-open-online-courses-feedback-interac> (accessed on 8 April 2024).
- Wang, H. Confused Student EEG Brainwave Data. Kaggle, 28 March 2018. Available online: <https://www.kaggle.com/datasets/wanghaohan/confused-eeg/code> (accessed on 11 February 2023).
- Xu, T.; Wang, J.; Zhang, G.; Zhou, Y. Confused or not: Decoding brain activity and recognizing confusion in reasoning learning using EEG. *J. Neural Eng.* **2023**, *20*, 026018. [CrossRef] [PubMed]

7. Xu, T.; Wang, X.; Wang, J.; Zhou, Y. From Textbook to Teacher: An Adaptive Intelligent Tutoring System Based on BCI. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine & Biology Society, Mexico City, Mexico, 1–5 November 2021; pp. 7621–7624. [CrossRef]
8. Yang, J.; Wang, H.; Zhu, J.; Eric, P.X. SeDMiD for Confusion Detection: Uncovering Mind State from Time Series Brain Wave Data. *arXiv* **2016**, arXiv:1611.10252.
9. Ni, Z.; Yuksel, A.C.; Ni, X.; Mandel, M.I.; Xie, L. Confused or not Confused? Disentangling Brain Activity from EEG Data Using Bidirectional LSTM Recurrent Neural Networks. In Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB'17), Boston, MA, USA, 20–23 August 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 241–246. [CrossRef]
10. Zhou, Y.; Xu, T.; Li, S.; Li, S. Confusion State Induction and EEG-based Detection in Learning. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; pp. 3290–3293. [CrossRef]
11. Ibtihaz, N.; Naznin, M. Determining Confused Brain Activity from EEG Sensor Signals. In Proceedings of the 8th International Conference on Networking, Systems and Security (NSysS'21), Cox's Bazar, Bangladesh, 21–23 December 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 91–96. [CrossRef]
12. Tahmassebi, A.; Gandomi, A.H.; Meyer-Baese, A. An Evolutionary Online Framework for MOOC Performance Using EEG Data. In Proceedings of the 2018 IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–8. [CrossRef]
13. Zhou, Y.; Xu, T.; Li, S.; Shi, R. Beyond engagement: An EEG-based methodology for assessing user's confusion in an educational game. *Univ. Access Inf. Soc.* **2019**, *18*, 551–563. [CrossRef]
14. Kumar, B.; Gupta, D.; Goswami, R.S. Classification of student's confusion level in e-learning using machine learning: Sciencegate. *Int. J. Innov. Technol. Explor. Eng.* **2019**, *9*, 346–351. Available online: <https://www.sciencegate.app/document/10.35940/ijitee.b1092.1292s19> (accessed on 1 June 2023).
15. Erwianda, M.S.F.; Kusumawardani, S.S.; Santosa, P.I.; Rimadana, M.R. Improving Confusion-State Classifier Model Using XGBoost and Tree-Structured Parzen Estimator. In Proceedings of the 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 5–6 December 2019; pp. 309–313. [CrossRef]
16. Wang, J.; Wang, M. Review of the emotional feature extraction and classification using EEG signals. *Cogn. Robot.* **2021**, *1*, 29–40. [CrossRef]
17. Renosa, C.R.M.; Bandala, A.A.; Vicerra, R.R.P. Classification of Confusion Level Using EEG Data and Artificial Neural Networks. In Proceedings of the 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), Laoag, Philippines, 29 November–1 December 2019; pp. 1–6. [CrossRef]
18. Dakoure, C.; Benlamine, M.S.; Frasson, C. Confusion detection using cognitive ability tests. *Int. Flairs Conf. Proc.* **2021**, *34*. [CrossRef]
19. Benlamine, M.S.; Frasson, C. Confusion Detection within a 3D Adventure Game. *Intell. Tutoring Syst.* **2021**, *12677*, 387–397. [CrossRef]
20. He, S.; Xu, Y.; Zhong, L. EEG-based Confusion Recognition Using Different Machine Learning Methods. In Proceedings of the 2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE), Hangzhou, China, 5–7 November 2021; pp. 826–831. [CrossRef]
21. Dagheriri, T.; Rustam, F.; Aljedaani, W.; Bashiri, A.H.; Ashraf, I. Electroencephalogram Signals for Detecting Confused Students in Online Education Platforms with Probability-Based Features. *Electronics* **2022**, *11*, 2855. [CrossRef]
22. Abu-Gellban, H.; Zhuang, Y.; Nguyen, L.; Zhang, Z.; Imhmed, E. CSDL EEG: Identifying Confused Students Based on EEG Using Multi-View Deep Learning. In Proceedings of the 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC), Los Alamitos, CA, USA, 1–27 July 2022; pp. 1217–1222. [CrossRef]
23. Men, X.; Li, X. Detecting the Confusion of Students in Massive Open Online Courses Using EEG. *Int. J. Educ. Humanit.* **2022**, *4*, 72–77. [CrossRef]
24. Islam, R.; Moni, M.A.; Islam, M.; Mahfuz, R.A.; Islam, S.; Hasan, K.; Hossain, S.; Ahmad, M.; Uddin, S.; Azad, A.; et al. Emotion Recognition From EEG Signal Focusing on Deep Learning and Shallow Learning Techniques. *IEEE Access* **2021**, *9*, 94601–94624. [CrossRef]
25. Background on Filters for EEG. Clinical Gate, 24 May 2015. Available online: <https://clinicalgate.com/75-clinical-neurophysiology-and-electroencephalography/> (accessed on 13 February 2023).
26. Preprocessing. NeurotechEDU. Available online: <http://learn.neurotechedu.com/preprocessing/> (accessed on 13 February 2023).
27. Al-Fahoum, A.S.; Al-Fraihat, A.A. Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains. *ISRN Neurosci.* **2014**, *2014*, 730218. [CrossRef]
28. Geng, X.; Li, D.; Chen, H.; Yu, P.; Yan, H.; Yue, M. An improved feature extraction algorithms of EEG signals based on motor imagery brain-computer interface. *Alex. Eng. J.* **2022**, *61*, 4807–4820. [CrossRef]
29. Nandi, A.; Ahamed, H. Time-Frequency Domain Analysis. *Cond. Monit. Vib. Signals* **2019**, 79–114. [CrossRef]
30. Harpale, V.K.; Bairagi, V.K. Seizure detection methods and analysis. In *Brain Seizure Detection and Classification Using Electroencephalographic Signals*; Academic Press: Cambridge, MA, USA, 2022; pp. 51–100. [CrossRef]

31. Suviseshamuthu, E.S.; Handiru, V.S.; Allexandre, D.; Hoxha, A.; Saleh, S.; Yue, G.H. EEG-based spectral analysis showing brainwave changes related to modulating progressive fatigue during a prolonged intermittent motor task. *Front. Hum. Neurosci.* **2022**, *16*, 770053. [[CrossRef](#)]
32. Kim, D.W.; Im, C.H. EEG spectral analysis. In *Biological and Medical Physics, Biomedical Engineering*; Springer: Singapore, 2018; pp. 35–53.
33. Tsipouras, M.G. Spectral information of EEG signals with respect to epilepsy classification. *EURASIP J. Adv. Signal Process.* **2019**, *2019*, 10. [[CrossRef](#)]
34. Vanhollebeke, G.; De Smet, S.; De Raedt, R.; Baeken, C.; van Mierlo, P.; Vanderhasselt, M.-A. The neural correlates of Psychosocial Stress: A systematic review and meta-analysis of spectral analysis EEG studies. *Neurobiol. Stress* **2022**, *18*, 100452. [[CrossRef](#)]
35. Mainieri, G.; Loddo, G.; Castelnovo, A.; Balella, G.; Cilea, R.; Mondini, S.; Manconi, M.; Provini, F. EEG activation does not differ in simple and complex episodes of disorders of arousal: A spectral analysis study. *Nat. Sci. Sleep* **2022**, *14*, 1097–1111. [[CrossRef](#)]
36. An, Y.; Hu, S.; Duan, X.; Zhao, L.; Xie, C.; Zhao, Y. Electroencephalogram emotion recognition based on 3D feature fusion and convolutional autoencoder. *Front. Comput. Neurosci.* **2021**, *15*, 743426. [[CrossRef](#)] [[PubMed](#)]
37. Uyanık, H.; Ozelcik, S.T.A.; Duranay, Z.B.; Sengur, A.; Acharya, U.R. Use of differential entropy for automated emotion recognition in a virtual reality environment with EEG signals. *Diagnostics* **2022**, *12*, 2508. [[CrossRef](#)] [[PubMed](#)]
38. Ding, L.; Duan, W.; Wang, Y.; Lei, X. Test-retest reproducibility comparison in resting and the mental task states: A sensor and source-level EEG spectral analysis. *Int. J. Psychophysiol.* **2022**, *173*, 20–28. [[CrossRef](#)] [[PubMed](#)]
39. Roy, Y.; Banville, H.; Albuquerque, I.; Gramfort, A.; Falk, T.H.; Faubert, J. Deep learning-based electroencephalography analysis: A systematic review. *J. Neural Eng.* **2019**, *16*, 051001. [[CrossRef](#)]
40. Ding, Y.; Robinson, N.; Zeng, Q.; Chen, D.; Wai, A.A.P.; Lee, T.-S.; Guan, C. TSception: A Deep Learning Framework for Emotion Detection Using EEG. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–7. [[CrossRef](#)]
41. Wang, J.; Song, Y.; Mao, Z.; Liu, J.; Gao, Q. EEG-Based Emotion Identification Using 1-D Deep Residual Shrinkage Network With Microstate Features. *IEEE Sensors J.* **2023**, *23*, 5165–5174. [[CrossRef](#)]
42. Altaheri, H.; Muhammad, G.; Alsulaiman, M.; Amin, S.U.; Altuwaijri, G.A.; Abdul, W.; Bencherif, M.A.; Faisal, M. Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review. *Neural Comput. Appl.* **2021**, *35*, 14681–14722. [[CrossRef](#)]
43. Lun, X.; Yu, Z.; Chen, T.; Wang, F.; Hou, Y. A simplified CNN classification method for Mi-EEG via the electrode pairs signals. *Front. Hum. Neurosci.* **2020**, *14*, 338. [[CrossRef](#)] [[PubMed](#)]
44. Amin, S.U.; Alsulaiman, M.; Muhammad, G.; Mekhtiche, M.A.; Hossain, M.S. Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion. *Futur. Gener. Comput. Syst.* **2019**, *101*, 542–554. [[CrossRef](#)]
45. Joshi, V.M.; Ghongade, R.B.; Joshi, A.M.; Kulkarni, R.V. Deep BiLSTM neural network model for emotion detection using cross-dataset approach. *Biomed. Signal Process. Control* **2021**, *73*, 103407. [[CrossRef](#)]
46. Wang, H.; Wu, Z.; Xing, E.P. Removing Confounding Factors Associated Weights in Deep Neural Networks Improves the Prediction Accuracy for Healthcare Applications. *Biocomputing* **2018**, *24*, 54–65. [[CrossRef](#)]
47. Khan, S.M.; Liu, X.; Nath, S.; Korot, E.; Faes, L.; Wagner, S.K.; Keane, P.A.; Sebire, N.J.; Burton, M.J.; Denniston, A.K. A global review of publicly available datasets for ophthalmological imaging: Barriers to access, usability, and generalizability. *Lancet Digit. Health* **2021**, *3*, 51–66. [[CrossRef](#)] [[PubMed](#)]
48. Lodge, J.M.; Kennedy, G.; Lockyer, L.; Arguel, A.; Pachman, M. Understanding difficulties and resulting confusion in Learning: An integrative review. *Front. Educ.* **2018**, *3*, 49. [[CrossRef](#)]
49. Young, A.T.; Xiong, M.; Pfau, J.; Keiser, M.J.; Wei, M.L. Artificial intelligence in dermatology: A primer. *J. Investig. Dermatol.* **2020**, *140*, 1504–1512. [[CrossRef](#)] [[PubMed](#)]
50. Maria, S.; Chandra, J. Preprocessing Pipelines for EEG. *SHS Web Conf.* **2022**, *139*, 03029. [[CrossRef](#)]
51. Donoghue, T.; Dominguez, J.; Voytek, B. Electrophysiological Frequency Band Ratio Measures Conflate Periodic and Aperiodic Neural Activity. *eNeuro* **2020**, *7*, ENEURO.0192-20.2020. [[CrossRef](#)]
52. Ganepola, D.; Maduranga, M.; Karunaratne, I. Comparison of Machine Learning Optimization Techniques for EEG-Based Confusion Emotion Recognition. In Proceedings of the 2023 IEEE 17th International Conference on Industrial and Information Systems (ICIIS), Peradeniya, Sri Lanka, 25–26 August 2023; pp. 341–346. [[CrossRef](#)]
53. Ganepola, D.; Karunaratne, I.; Maduranga, M.W.P. Investigation on Cost-Sensitivity in EEG-Based Confusion Emotion Recognition Systems via Ensemble Learning. In *Asia Pacific Advanced Network*; Herath, D., Date, S., Jayasinghe, U., Narayanan, V., Ragel, R., Wang, J., Eds.; APANConf 2023; Communications in Computer and Information Science; Springer: Cham, Switzerland, 2024; Volume 1995. [[CrossRef](#)]
54. Arguel, A.; Lockyer, L.; Lipp, O.V.; Lodge, J.M.; Kennedy, G. Inside Out: Detecting Learners’ Confusion to improve Interactive Digital Environments. *J. Educ. Comput. Res.* **2016**, *55*, 526–551. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.