Article

Objective Detection of Trust in Automated Urban Air Mobility: A Deep Learning-Based ERP Analysis

Yuhan Li [1,2], Shuguang Zhang [3,*], Ruichen He [2,3] and Florian Holzapfel [2]

1   School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China; ffflora@buaa.edu.cn
2   Institute of Flight System Dynamics, Technical University of Munich, 80333 Munich, Germany; heruichen@buaa.edu.cn (R.H.); florian.holzapfel@tum.de (F.H.)
3   School of Transportation Science and Engineering, Beihang University, Beijing 100191, China
*   Correspondence: gnahz@buaa.edu.cn

Abstract: Urban Air Mobility (UAM) has emerged in response to increasing traffic demands. As UAM involves commercial flights in complex urban areas, well-established automation technologies are critical to ensure a safe, accessible, and reliable flight. However, the current level of acceptance of automation is insufficient. Therefore, this study sought to objectively detect the degree of human trust toward UAM automation. Electroencephalography (EEG) signals, specifically Event-Related Potentials (ERP), were employed to analyze and detect operators' trust towards automated UAM, providing insights into cognitive processes related to trust. A two-dimensional convolutional neural network integrated with an attention mechanism (2D-ACNN) was also established to enable the end-to-end detection of trust through EEG signals. The results revealed that our proposed 2D-ACNN outperformed other state-of-the-art methods. This work contributes to enhancing the trustworthiness and popularity of UAM automation, which is essential for the widespread adoption and advances in the UAM domain.

Keywords: automated vehicles; UAM; trust in automation; cognition; event-related potentials; CNN; ergonomics

1. Introduction

Some urban areas are facing increasingly serious traffic congestion issues with the development of society. Conventional ground transportation methods might be insufficient in addressing this congestion [1]. To provide more possibilities for urban transport, the idea of "Urban Air Mobility (UAM)" was introduced, having been initially put forth by the National Aeronautics and Space Administration (NASA) [2]. UAM involves flying in complex urban airspace and densely populated areas. Many stakeholders have been attempting to develop UAM. For example, the Federal Aviation Administration (FAA) provided a broader UAM blueprint [3] and instruction for vertiport design [4]. Rahman et al. explored how to integrate vertiports with existing public transport systems, as well as design guidelines for vertiports [5]. Rostami et al. [6] and Kadhiresan [7] investigated a way of designing the configuration of Electric Vertical Take-off and Landing (eVTOL) aircraft.

However, just like operators of other vehicles, human operators of UAM vehicles are also prone to experiencing fatigue and mental workload in complicated conditions, leading to decreased cognitive abilities and human errors [8]. Additionally, UAM involves commercial flights and has a broad application in the mass market [9]. Hence, well-established automation technologies are critical to ensure safe, accessible, and reliable UAM. However, public trust in automated vehicles remains precarious at present [10,11]. Concerns about the reliability and safety of current automation, a lack of understanding of automation decisions, and past accidents and incidents of automated vehicles all hinder people's trust in automation [12]. People's trust in automation can affect their acceptance and use of the system, such as the time they use the automated system and the level at

which they use it [13]. Ascertaining the operator's trust in UAM automation can effectively prevent misuse or abuse of the automation, thereby improving UAM safety, increasing public acceptance of UAM vehicles, and thus motivating the development of UAM.

Therefore, it is imperative to ascertain the level of human trust in UAM automation. The UAM Coordination and Assessment Team (UCAT) of NASA categorized UAM automation into five levels [14], as shown in Figure 1. The Assistive level represents a lower tier of automation, where operational safety is entirely entrusted to humans. The Comprehensive Safety-assurance level involves automation offering safety-related aid, like collision avoidance, yet humans remain entirely accountable for operational safety. Meanwhile, the Collaborative and Responsible level entails the automation system being proficient in executing specific functions, so humans can be relieved, and training on such functions can be reduced. However, human supervision of the entire system remains crucial to ensure safety, which means that human involvement is still required. The Highly-integrated Automated Networks level no longer requires real-time human involvement but can be improved through human intervention. At the System-wide Automation Optimization level, human supervision and intervention in the system are no longer required [15].
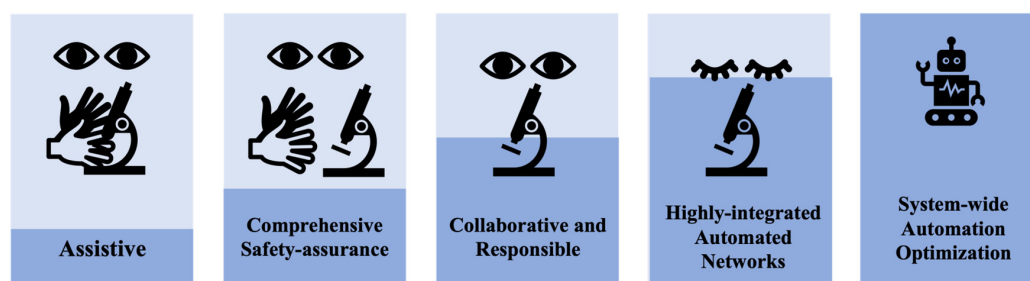

Figure 1. UAM automation levels according to UML [15].

Currently, automation technology in the aviation domain has reached a certain maturity level. The aircraft can fly automatically in most flight phases, leading to a reduced workload for pilots and improved flight efficiency and accuracy [16]. However, the current automation technology still has certain limitations. For example, there are potential safety risks [17]. Additionally, the criteria for ethics are unclear, such as how automation should set priorities when continuing to drive may threaten other lives but swerving may threaten the lives of the onboard passengers [18]. Hence, it is generally believed that achieving the Collaborative and Responsible level could be gradually realized in the next few decades [19]. Hence, this study mainly focuses on detecting trust in automation at the Collaborative and Responsible level.

It is widely believed that the safety of automation systems plays a crucial role in their acceptance, as automated systems involve human safety [20]. The safety assessment is based on comprehensive safety data throughout the entire flight process, including flight data, training data, procedures, forecasting data, etc. Another important factor that influences human trust towards automation is the user's psychological perception of automated systems [21]. It emphasizes the user's immediate perceptions when utilizing automation. This paper primarily investigates the latter aspect, delving into objective measurements beyond the traditional subjective evaluations of user trust in automation systems.

Traditionally, the measurement of trust levels is commonly conducted through subjective surveys [22], including questions about the acceptance level [23], usability [24], and expectations of future use [25]. There are also studies employing objective real-time measurements of human trust. For example, Highland et al. [12] used objective indicators, including eye-tracking data, mental workload, and task performance, to measure human trust in autonomy in dog-fighting. Dsouza et al. [26] adopted electroencephalography (EEG) to evaluate passenger trust in drivers. He et al. [27] employed electrocardiogram (ECG) and pupil diameter to investigate drivers' risk perception and trust under the Society

of Automotive Engineers (SAE) Level 2. Perello-March et al. [28] measured functional near-infrared spectroscopy (fNIRS), establishing the connection between human neural activities and trust within a highly automated driving context. However, most such studies mainly focus on ground vehicle drivers or fighter pilots, and research on the trust of commercial UAM operators remains largely unexplored. Compared to ground vehicles and traditional aircraft, the detection of trust for UAM vehicles is much more of a challenge. On the one hand, UAM vehicles usually run in more complex scenarios, which involve three-dimensional flying in busy urban areas, so the cognitive process for trusting the automation will be more sophisticated and the key features more difficult to capture [29]. On the other hand, since operators of UAM vehicles tend not to be as well trained as traditional civil or military aviation pilots, their misuse or abuse of automation may be greater [30].

Trust is a cognitive process with neural correlates related to emotion, cognition, and decision-making [31]. Regarding this, it can be said that trust can be linked to physiological indicators, including eye-tracking data, ECG, EEG, fNIRS, etc. [12,26–28]. Among a broad spectrum of physiological signals, EEG is considered to be the most responsive indicator of neural activities [32] and has been used in many studies [33–35]. It is believed that trust and distrust are different cognitive processes because different brain patterns would be activated [36]. Given that trust is related to attention, perception, and cognition, it is a complex concept involving multidimensional information. By leveraging the advantages of neural networks for non-linear and complex hidden feature extraction, effective detection of trust might be enabled [37]. Deep learning (DL) has found extensive application across diverse domains owing to its good performance and has shown satisfying results in applications such as mental load assessment [38], motor imagery [39], and emotion recognition [40].

This research aims to apply such methods to detect human trust in automated UAM within the realm of human–computer interaction by analyzing Event-Related Potentials (ERP), an important component of EEG. In other words, the aim of this study is to establish a trust detection framework in automated UAM by developing an attention-based convolutional neural network through ERP. Subsequently, the efficacy of the proposed framework will be validated by conducting a performance comparison with three state-of-the-art deep learning models. In summary, the research questions are:

➢ Does the ERP reflect human trust in automated systems in flight scenarios?
➢ Is deep learning superior to traditional statistical analysis or machine learning for processing and analyzing EEG signals? If so, for what reasons?
➢ Would the integration of the attention mechanism and 2D convolution optimize the deep learning model? If so, for what reasons?

2. Methodology

Trust refers to a complex psychosocial concept that involves the reliance and dependence of one entity on another [41]. In this study, trust is defined as the acceptance and attitude towards automation, rather than dependence on the system. ERP has demonstrated its ability to reflect social trust [26,42]. ERP typically occurs within 600 ms of stimulus appearance; they can reflect brain response [43]. This study specifically focuses on the P300 component. The P300 component is a positive potential that reaches its peak at approximately 300–600 ms after a stimulus and is regarded as closely associated with emotional and cognitive states [44]. By analyzing its latency, wave amplitude, and spectrum, insights into the neural processes related to human trust in automated UAM can be gained. A two-dimensional convolutional neural network integrated with the attention mechanism (2D-ACNN) was also developed to further detect trust. The overall experiment setup is shown in Figure 2 [45].
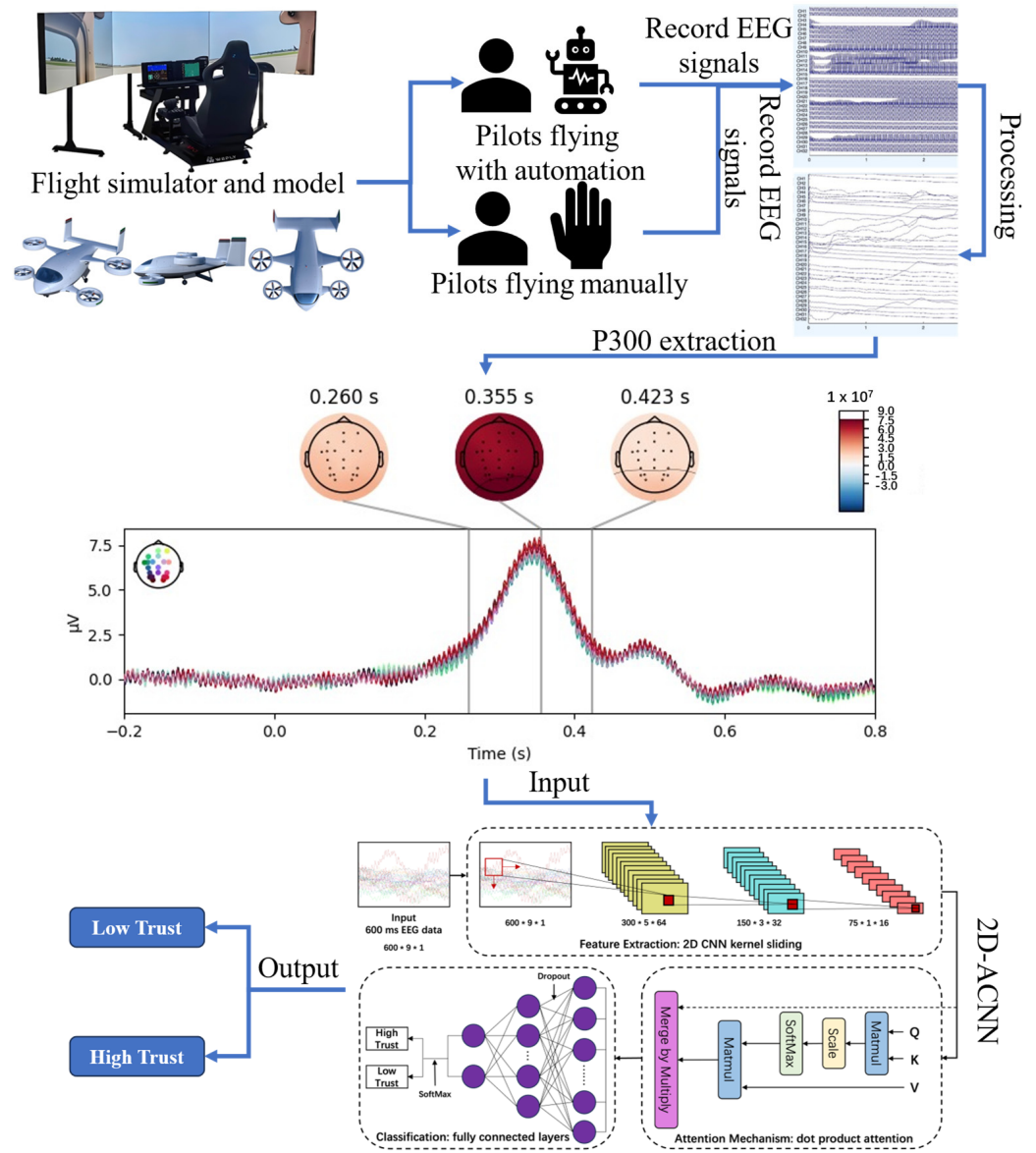
Figure 2. The overall experiment setup [45,46].

2.1. Data Description

The data for this study are part of the Simplified Vehicle Operations (SVO) project [47], which has been adapted for further research beyond its original goals. The experiments were conducted using a simulator with three screens, a pushrod to control throttle, a joystick to control direction, an avionics instrument display, a cockpit seat, and a flight model of the aeroG Aviation aG-4 "Liberty" Electric Vertical Take-off and Landing (eVTOL) aircraft, as shown in Figure 2, upper left part. Physiological data were collected by the BIOPAC system [48], where EEG signals were acquired from 21 locations on the scalp, within a dry electrode cap following the international 10–20 system. The electrode distribution is shown in Figure 3.

This project involved 40 participants to ensure a sufficient amount of data. They were asked to operate the simulated aircraft to take off from the suburb, then fly through the outskirts to the urban airspace, and eventually land on the top floor of a skyscraper. The flight duration was approximately 40 min, as depicted in Figure 4.
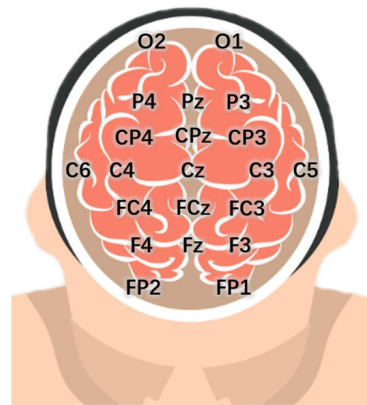
Figure 3. Electrode distribution.

Figure 4. Flight path, take-off point, and landing point for the experiment.

Flying the eVTOL in a city scenario is a challenging task because the pilot needs to control the aircraft in a complex urban environment, dealing with various complex factors such as other aircraft operating nearby, including other eVTOLs or helicopters, certain authorized and not authorized flying objects including drones, and the urban area architectonic landscape. The vertical take-off and landing also pose a great challenge to the pilot, as they require highly precise control of the aircraft's attitude and position. To reduce the pilot's workload, many eVTOLs are equipped with advanced automation systems such as automatic obstacle avoidance, automatic cruise, etc. [49]. However, a crewless flight can only be currently achieved in a controlled environment.

Before the experiment, participants were trained to operate the eVTOL and practiced for the secondary task, which is the oddball task [50], to evaluate workload. In this project, half of the participants had to fly manually. The rest of the participants were engaged in flight with a highly automated system, wherein they only needed to monitor the flight situation and did not need to fly manually. As this paper aims to analyze trust in automated UAM, we mainly focus on participants who fly with an automated system.

Although the automated system is introduced to reduce the pilot's workload, automation accidents are induced artificially. In other words, the automation system could randomly fail and trigger an alarm, requiring participants to press a button and take control of the aircraft. Additionally, to investigate the effects of automation failure on operators' driving behavior, 10 participants experienced a "catastrophic accident" during the automated flight, where the automation failed and resulted in the crash of the aircraft after 10 min of the flight. After the crash, the aircraft was reborn in the same location and participants were asked to continue the experiment. During the flight, participants were also required to perform a non-driving-related oddball task to better evoke ERP [50]. Within the oddball paradigm, participants received a sequence of auditory stimuli with a 5 s interval, wherein 85% were standard stimuli and 15% were oddball stimuli. Participants were asked to step on a sensor under their feet when the oddball stimuli occurred. The ERP waveforms elicited by the oddball paradigm convey crucial information to understand cognitive mechanisms regarding trust [51]. Since participants were asked to allocate their cognitive resources primarily to control the simulator, it was assumed that the oddball paradigm would not impose an additional flight workload on them.

During the experiment, there were regular pauses every 10 min, during which participants were interviewed about their subjective workload (NASA Task Load Index), ease of operation, and feelings about the automation [52].

2.2. EEG Signal Processing

For this study, EEG data from 20 participants who utilized the automation system for flying will be processed and analyzed to demonstrate the feasibility of applying ERP to detect trust in UAM automation. These participants were divided into two groups because different levels of trust can be induced through the automation reliability perceived by naive participants [28]. Participants who experienced a "catastrophic accident" were regarded as the Low Trust (LT) group, while the rest were assigned to the High Trust (HT) group. For the LT group, we only analyzed data after the "catastrophic accident". A t-test of subjective questionnaires was conducted to prove that the categorization was acceptable [53].

The EEG data underwent processing by the MNE-Python package [54]. Initially, the data were subjected to filtration via a second-order Butterworth filter, within the range of 0.1–40 Hz. Following this, noise separation was performed through Independent Component Analysis (ICA) [55]. The EEG signals before and after processing can be found in Figure 5.
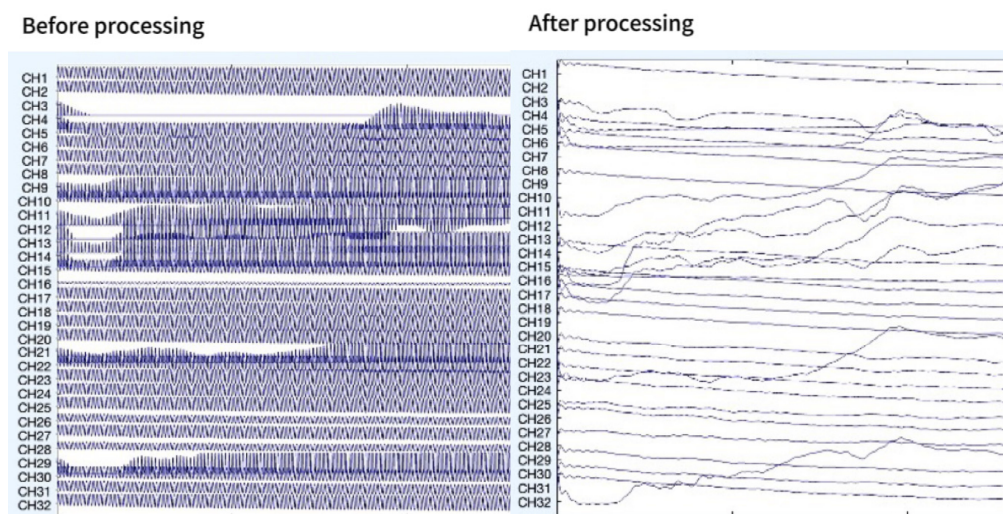

Figure 5. The EEG signals before and after processing.

As the P300 waveform is most prominent in the frontal and parietal lobe [51], we selected data from these patterns, including the CPz, Pz, Fz, FCz, Cz, P3, F3, P4, and F4 electrodes for further statistical analysis. The EEG data were then segmented with an epoch of 800 ms time-locked to the oddball stimulus, spanning from 200 ms prior to the oddball stimulus to 600 ms following it. For each epoch, the data from the first 200 ms were used as the baseline to normalize the data from the subsequent 600 ms.

The processed EEG data were exported into Excel files, including the EEG information for each selected channel, the trust condition (HT or LT), and the participant IDs. The MNE-Python package was adopted to compute the average ERP waveform for each participant to enhance the stability and reliability of the waveform [56]. Overlaying the ERP reduces the effect of random noise and better reflects the neural responses, so the average ERP waveform of all participants was then overlayed for each trust condition. Finally, the amplitude and latency of ERP waveforms were measured and compared across different trust conditions.

2.3. Trust Detection Framework

Although there are some studies analyzing human trust using EEG, few devote effort into developing DL frameworks for trust detection [37]. The key to utilizing DL on trust detection is to efficiently extract crucial EEG features and perform classification. Shafiei et al. used a linear SVM in combination with kernel-target alignment (KTA) and kernel class separability (KTS criteria) to detect a surgeon's trust in a robot-assisted interface in robot-assisted surgery (RAS) [57]. The method is simple and effective but performs well only in binary classification. Choo and Nam utilized a convolutional neural network (CNN) to estimate operators' trust in automation while performing Air Force multi-attribute task battery (AF-MATB) by extracting and analyzing image features of electroencephalogram (EEG) signals [37]. This network showed great results. However, the experimental data for this network were collected under specific tasks where the EEG waveforms were more pronounced. It may not be as effective as such in a perturbed real flight scenario and modification is required.

For this research endeavor, our plan involves the utilization of a convolutional neural network (CNN) to accomplish this task. A CNN is a deep neural network with a convolutional structure and multiple building blocks, which serves the dual purpose of feature extraction and classification [58,59]. As a CNN requires minimal preprocessing of the input data, it is well suited for EEG signal analysis and has been utilized in tasks such as brain disease detection [60], emotion classification [61], BCI identification [62], etc. We believe such dominance could also work in trust detection.

Typically, deep learning studies for EEG signal recognition mostly utilize Conv1D or Long Short-Term Memory (LSTM) [63]. However, these structures can only extract features along a single direction, potentially ignoring useful spatial information. The Conv2D was utilized, which possesses the capability to extract interconnections among multiple EEG channels, not solely concurrently but also across neighboring time segments, for trust detection [64]. The EEG signals, consisting of electrical activities at different electrode positions, are transformed into a time series of 2D data. The data are represented as Equation (1):

$$E_i = [c^i_t, c^i_{t+1}, \ldots, c^i_{t+n}], \tag{1}$$

where the variable "i" signifies the i-th channel, "t" corresponds to the starting time, and "t+n" is the ending time.

Subsequently, 2D convolution and pooling techniques [65] were adopted to automatically derive temporal and spatial features from the EEG data. It is noteworthy that EEG signals encompass an extensive array of multi-dimensional information, and the aim is to extract only the information relevant to trust [66]. Hence, we integrated an attention module, enabling the network to adaptively adjust its attention to different features [67]. This attention mechanism helps emphasize crucial features while ignoring noise or unimportant information.

In general, the proposed trust detection framework consists of three parts: CNN feature extraction modules, attention module, and classification module, as illustrated in Figure 6.

To be more specific, the feature extraction module comprises three blocks, each encompassing a sequence of components: a 2D convolutional layer, a 2D max-pooling layer, and a batch normalization layer. This arrangement is illustrated in Figure 7. The convolution kernel glides through the time steps, capturing time domain features, and performing convolution operations on different spatial locations to extract spatial features. The Max-Pooling layers incorporate 2 × 2 max-pooling filters, employing a stride of 2. Batch normalization is utilized to prevent overfitting. For more intricate insights into convolution and pooling methodologies, in-depth information can be found in [68]. After processing by the feature extraction module, several features are extracted.
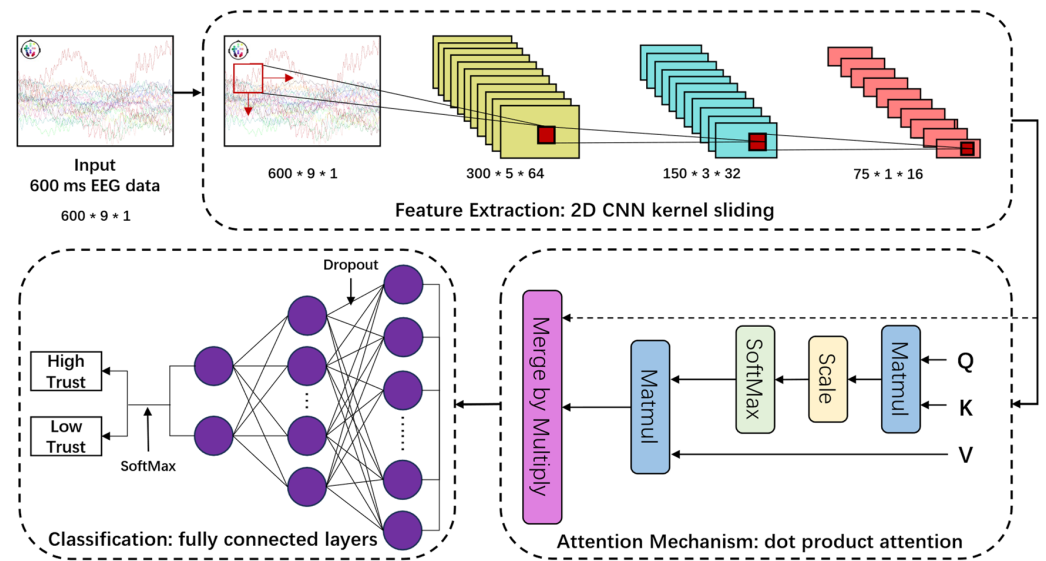

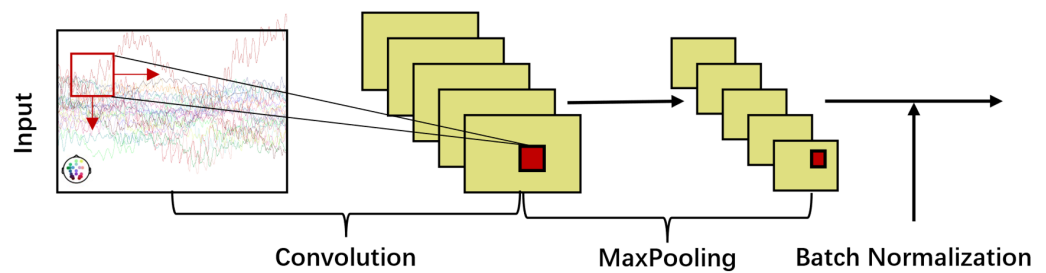**Figure 6.** The structure of the 2D-ACNN [59].

Figure 7. The architecture of the feature extraction module [64].

Subsequent to the completion of the feature extraction module, an attention module is introduced. By calculating the correlation between the Query and Key, the attention mechanism determines the importance of each feature and concentrates more attention on crucial ones. The Value is then weighted and summed according to the computed attention weights. In this way, the model automatically fuses information from different time steps and channels based on the correlation between Q and K, obtaining more informative feature representations. The expression of the attention weights is represented by Equation (2):

$$Q = f_Q(Y), K = f_K(Y), V = f_V(Y),$$
$$\text{Attention } (Q,\ K,\ V) = \text{Softmax } \left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{2}$$

where Y is the feature vectors, $f_Q(Y')$, $f_K(Y')$, and $f_V(Y')$ represent three different linear transformations, and d_k is the length of the feature vector.

Finally, fully connected layers made up of 512 and 64 hidden neurons are used for the final classification. The output layer is a Dense layer with Softmax activation.

During the process of model training, the batch size was set at 128, the initial learning rate was established as 0.001, and the Adam optimizer was utilized [69]. Early stopping was adopted, where training stops when the model's accuracy no longer improves over 50 epochs, and the optimal model was saved [70]. The model was implemented using PyTorch.

2.4. Evaluation Method

To gauge the efficacy of the proposed DL model, we employ metrics including Accuracy, Recall, Precision, and F1 score for evaluation, following the approaches outlined by [71] and [72]. These metrics are defined as Equation (3):

$$
\begin{aligned}
\text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN}, \\
\text{Recall} &= \frac{TP}{TP + FN}, \\
\text{Precision} &= \frac{TP}{TP + FP}, \\
\text{F1 score} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\end{aligned}
\tag{3}
$$

where TP stands for True Positive, TN signifies True Negative, FP denotes False Positive, and FN represents False Negative.

Furthermore, to gauge the efficacy of the proposed trust detection framework, a comparative analysis was carried out. Several state-of-the-art DL models were incorporated for comparison with regard to the above metrics. The compared models are as follows:

- IIFTWSVM [73]: IIFTWSVM is a modified version of the Intuitionistic Fuzzy Twin Support Vector Machines (IFTWSVM) method. This approach integrates both membership and non-membership weights, aimed at minimizing the impact of outliers. It introduces diverse Lagrangian functions as a strategy to circumvent the need for matrix inverse computations.
- Cluster-based KNN [74]: this method involves calculating the distance from the clustered mean of the test sample. The test samples are categorized according to the primary vote of their neighboring points, determined by the first K rows in ascending order.
- Bi-LSTM [75]: Bi-directional LSTM involves training two LSTM layers on the input sequence instead of one. The key difference is that the first recurrent layer in the network is duplicated, creating two layers placed side by side.

3. Results

In this part, we investigated the subjective perceptions of participants' trust towards the automated system and participants' performance in the oddball task. As the dataset was originally used for workload analysis, we further analyzed it to determine whether the data were capable of reflecting trust. The ERP from the HT group and LT group were also compared. Furthermore, an analysis of the performance of the proposed trust detection model was conducted, along with an evaluation of other DL models.

3.1. Subjective Trust Rating

Prior to the experiment and every 10 min during the experiment, participants were requested to rate their trust toward the automation and workload. It was anticipated that variations in flight scenarios and disparate levels of automation trustworthiness would yield distinct ratings among participants.

The outcomes are presented in Figure 8. It is evident that the subjective trust ratings of both groups differed only by 0.15 before the experiment. However, after the "catastrophic accident", the trust rating in the LT group dropped sharply, followed by a slow recovery, but remained lower than at the beginning. On the other hand, the trust level in the HT group exhibited fluctuations and gradually increased after the start of the experiment.
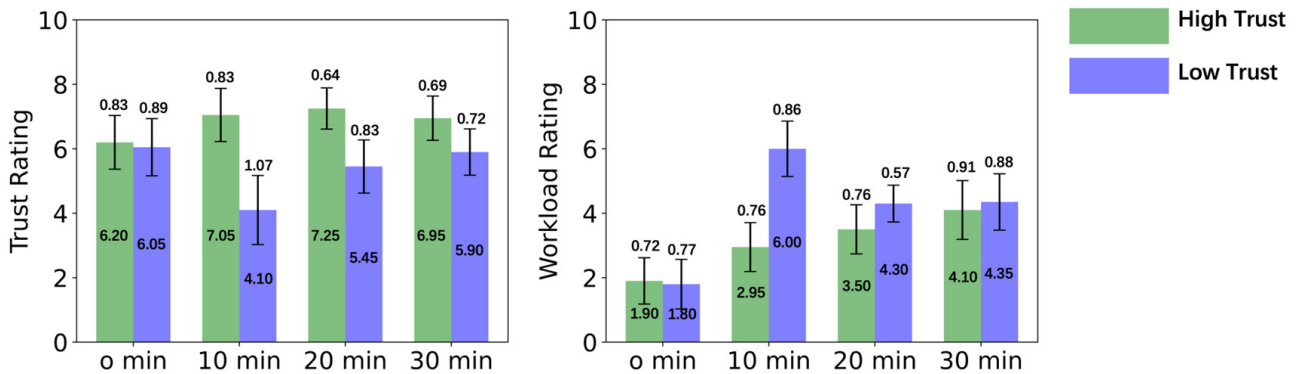

Figure 8. The subjective rating of trust and workload at 0 min, 10 min, 20 min, and 30 min during the experiment.

3.2. Oddball Task Performance

As trust level would affect participants' cognitive resources, it was assumed that oddball task performance might vary accordingly. The reaction time and error rate of the oddball task between the two groups are shown in Figure 9. An independent samples t-test was performed, indicating a noteworthy distinction regarding the level of trust between the two groups. Participants exhibited generally longer reaction times and higher error rates when their trust towards automation was lower.
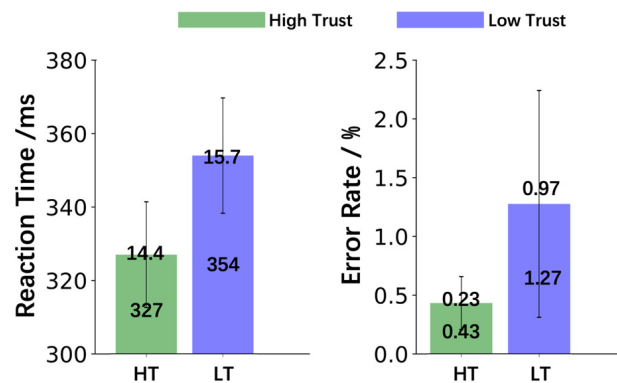

Figure 9. The reaction time and error rate of the oddball task for HT and LT.

3.3. ERP Results

The ERPs were analyzed for both trust conditions. Figure 10A exhibits ERP waveforms synchronized with the initiation of oddball stimuli for each channel within the two groups. Topographic maps illustrating the P300 effect can be observed in Figure 10B. Figure 10C showcases superimposed waveforms of the P300 response across various channels within both groups. Figure 10D illustrates the average amplitude and latency of the P300 response. It can be observed that for the HT group, the P300 response exhibited a more positive amplitude and shorter latency.
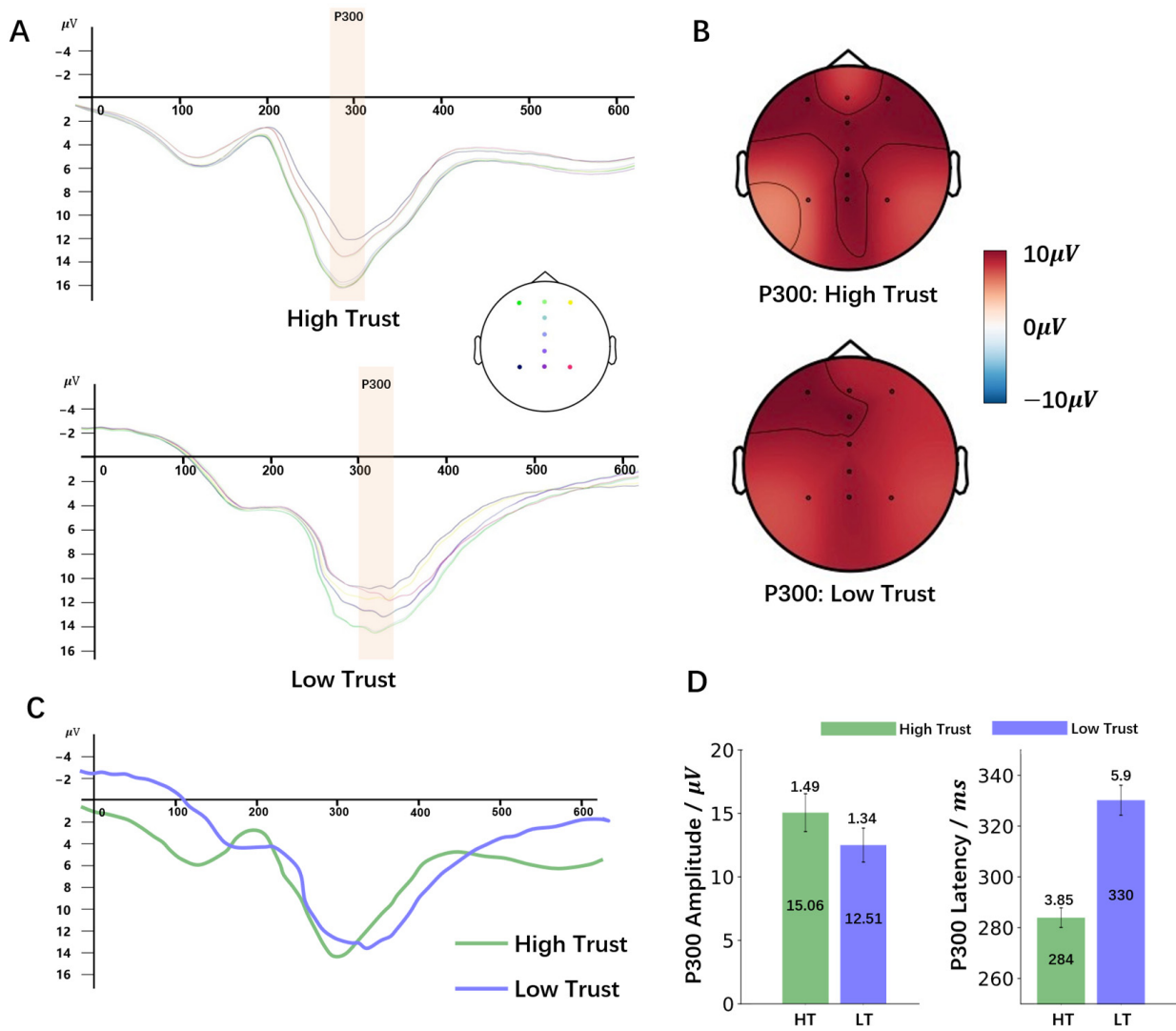
Figure 10. (**A**). ERP waveforms time-locked to the oddball stimuli for CPz, Pz, Fz, FCz, Cz, P3, F3, P4, and F4 electrodes. (**B**). The topographic maps for the P300 effect. (**C**). The average overlaid waveforms of the P300 response for CPz, Pz, Fz, FCz, Cz, P3, F3, P4, and F4 electrodes. (**D**). The average amplitude and latency of the P300 response.

3.4. Training Results

The data described in 2.2 previously were used as the training set for the proposed 2D-ACNN. *Train_Test_Split* [76] was used to randomly divide the data, allocating 80% of it for the training set and reserving 20% for the test set. Several models, including IIFTWSVM, Cluster-based KNN, and Bi-LSTM, were compared through the evaluation method. Table 1 and Figure 11 show the five-fold cross-validation results of these models. It can be observed that 2D-ACNN outperforms the other classifiers across all evaluation metrics.

Table 1. Five-fold cross validation results of IIFTWSVM, Cluster-based KNN, Bi-LSTM, and 2D-ACNN according to their accuracy, recall, precision, and F1 score.

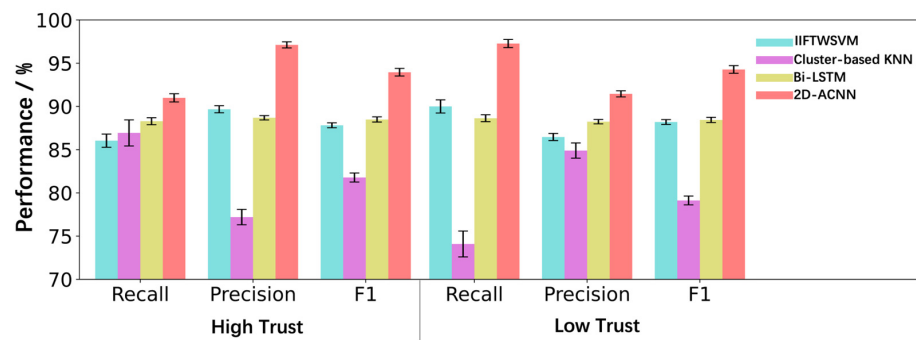| Evaluation	IIFTWSVM	Cluster-Based KNN	Bi-LSTM	2D-ACNN
Accuracy	88.01 ± 0.26	80.54 ± 1.12	88.46 ± 0.37	94.12 ± 0.46
Recall	88.02 ± 0.76	80.51 ± 1.51	88.46 ± 0.40	94.13 ± 0.47
Precision	88.01 ± 0.41	81.05 ± 0.89	88.46 ± 0.25	94.28 ± 0.36
F1 score	88.01 ± 0.29	80.45 ± 0.52	88.46 ± 0.52	94.11 ± 0.44

Figure 11. Comparison among IIFTWSVM, Cluster-based KNN, Bi-LSTM, and 2D-ACNN in relation to recall, precision, and F1 score for each class.

Furthermore, we performed an ablation study to demonstrate the efficacy of the attention block. Table 2 shows the model performance with and without the attention module. Figure 12 demonstrates feature maps with and without the attention module. The left map is the output feature map of the attention layer and the right map is the output feature map of the third convolutional layer. Although it is difficult to directly discern the usefulness of the attention module from the feature maps, the feature maps with and without the attention module differ. Coupled with the fact that the model with the attention module outperforms the one without, as shown in Table 2, we therefore hypothesize that the inclusion of the attention module helped the model to focus on key features.

Table 2. Five-fold cross-validation results with and without attention module.

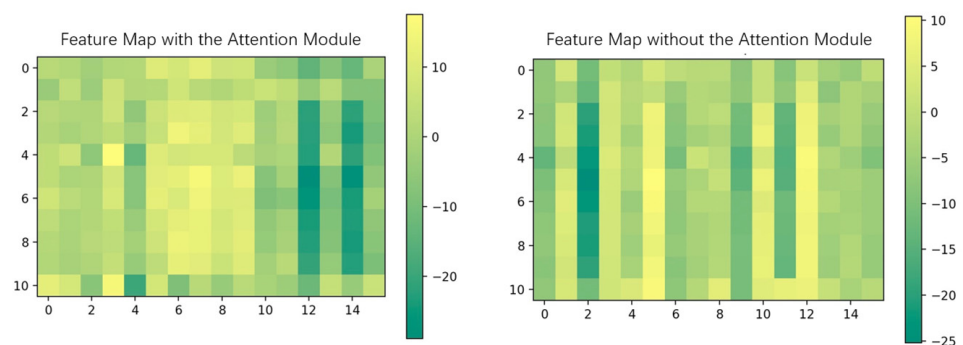| Evaluation	With Attention Module	Without Attention Module
Accuracy	94.12 ± 0.46	91.63 ± 0.55
Recall	94.13 ± 0.47	91.95 ± 0.78
Precision	94.28 ± 0.36	87.74 ± 0.55
F1 score	94.11 ± 0.44	89.49 ± 0.45

Figure 12. The feature maps with the attention module (**left**) and without the attention module (**right**).

4. Discussion

4.1. ERP Features for Trust

This study analyzes the EEG signals of two groups with different trust levels for an automated UAM vehicle. The ERP results revealed that when participants have higher trust in the automation system (15.06 millivolts), their P300 responses are more positive compared to the responses of participants who have lower trust (12.51 millivolts). In addition, as trust decreases, participants' risk perception increases, leading to heightened vigilance and a higher workload. Consequently, participants required more time to respond to stimuli, resulting in longer latency in the LT condition (330 ms) compared to the HT condition (284 ms). These findings align with earlier studies showing that lower trust levels can induce increased cognitive load and negative emotions [77].

4.2. Trust Detection Based on DL

A DL framework for trust detection in automated UAM using ERP was developed. Both Bi-LSTM and 2D-ACNN showed good performance in detecting trust, suggesting that there is a correlation between trust and the ERP components and that this correlation can be captured by DL models. Considering that trust is a complicated neural process, the relationship between trust and ERP is likely to be complex and non-linear. The DL models, consisting of multiple layers, can learn hierarchical and abstract representations, which is crucial when dealing with the intricacies of trust and its neural correlates. By leveraging the deep architecture, the non-linear relationship between trust and ERP can be effectively modeled, thereby achieving better performance in trust detection.

4.3. The 2D-ACNN Structure

The proposed 2D-ACNN exhibits better performance compared to the other-mentioned DL methods, perhaps benefiting from decoding ERP signals in a better way. The utilization of 2D convolution in the feature extraction module facilitates the extraction of temporal and spatial attributes at multiple scales. Furthermore, a scaled dot-product attention module is integrated with the CNN, which can appropriately assign weights, indicating that paying more attention to specific electrodes and time steps may contribute to enhancing classification performance.

5. Conclusions

This study utilizes a DL framework to detect operators' trust in automated UAM through objective neurophysiological methods. The operator's trust was greatly reduced after a "catastrophic accident" caused by automation. As trust reduces, the plotted P300 waveform amplitude also decreases, indicating that patterns in ERP can reveal the operator's trust in the automated UAM. The proposed 2D-ACNN can effectively extract key features from ERP using an attention mechanism, while preserving both temporal and spatial features, enabling trust detection. It outperforms other DL classifiers with respect to accuracy, recall, precision, and F1 score. The accuracy of 2D-ACNN is 94.12%, which is 5.66% higher than the widely used Bi-LSTM model. The research findings contribute to the following areas:

- Demonstrating the neurocognitive differences between trust and distrust, and revealing the close association between trust and risk perception during automated UAM;
- Proposing a novel 2D-ACNN model for trust detection through EEG signals, and exhibiting superior performance to other models due to its great feature extraction;
- Integrating the attention mechanism with 2D-CNN, thus achieving better trust detection by focusing on key time points and EEG channels.

In summary, the detection of trust can help the operators avoid over-trust or distrust in automation, thereby optimizing the design of automated vehicles. It is also an important step to promote UAM development. In future work, the limitations of this study will be further addressed, including increasing the sample size for analysis and training, investigating and modifying more advanced classification networks, and performing data augmentation, etc.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: www.kaggle.com/datasets.leonieschneider/p300-dataset-for-workload-detection (accessed on 10 November 2023).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kim, N.; Yoon, Y. Regionalization for urban air mobility application with analyses of 3D urban space and geodemography in San Francisco and New York. In Proceedings of the 12th International Conference on Ambient Systems, Networks and Technologies (ANT), Warsaw, Poland, 23–26 March 2021.
2. Mendonca, N.; Murphy, J.; Patterson, M.D.; Alexander, R.; Juarex, G.; Harper, C. Advanced Air Mobility Vertiport Considerations: A List and Overview. In Proceedings of the AIAA AVIATION 2022 Forum, Chicago, IL, USA, 27 June–1 July 2022.
3. Urban Air Mobility (UAM). *Concept of Operations, Version 2.0*; FAA: Washington, DC, USA, 2023.
4. Federal Aviation Administration. *Engineering Brief No. 105, Vertiport Design*; FAA: Washington, DC, USA, 2023.
5. Rahman, B.; Bridgelall, R.; Habib, M.F.; Motuba, D. Integrating Urban Air Mobility into a Public Transit System: A GIS-Based Approach to Identify Candidate Locations for Vertiports. *Vehicles* **2023**, *5*, 1803–1817. [CrossRef]
6. Rostami, M.; Bardin, J.; Neufeld, D.; Chung, J. EVTOL Tilt-Wing Aircraft Design under Uncertainty Using a Multidisciplinary Possibilistic Approach. *Aerospace* **2023**, *10*, 718. [CrossRef]
7. Kadhiresan, A.R.; Duffy, M.J. Conceptual design and mission analysis for eVTOL urban air mobility flight vehicle configurations. In Proceedings of the AIAA AVIATION 2019 Forum, Dallas, TX, USA, 17–21 June 2019.
8. Vajda, P.; Maris, J. *A Systematic Approach to Developing Paths Towards Airborne Vehicle Autonomy*; Report No. 20210019878; NASA, LLC: Merritt Island, FL, USA, 2021.
9. Wing, D.J.; Chancey, E.T.; Politowicz, M.S.; Ballin, M.G. Achieving resilient in-flight performance for advanced air mobility through simplified vehicle operations. In Proceedings of the AIAA AVIATION 2020 Forum, Virtual Event, 15–19 June 2020.
10. Cohen, A.P.; Shaheen, S.A.; Farrar, E.M. Urban Air Mobility: History, Ecosystem, Market Potential, and Challenges. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 6074–6087. [CrossRef]
11. Schweiger, K.; Preis, L. Urban Air Mobility: Systematic Review of Scientific Publications and Regulations for Vertiport Design and Operations. *Drones* **2022**, *6*, 179. [CrossRef]
12. Highland, P.; Schnell, T.; Woodruff, K.; Avdic-McIntire, G. Towards Human Objective Real-Time Trust of Autonomy Measures for Combat Aviation. *Int. J. Aerosp. Psychol.* **2023**, *33*, 1–34.
13. Merritt, S.M.; Heimbaugh, H.; LaChapell, J.; Lee, D. I Trust It, but I Don't Know Why: Effects of Implicit Attitudes Toward Automation on Trust in an Automated System. *Hum. Factors* **2013**, *55*, 520–534. [CrossRef] [PubMed]
14. Li, C.; Qu, W.; Li, Y.; Huang, L.; Wei, P. Overview of traffic management of urban air mobility (UAM) with eVTOL aircraft. *Traffic Transp.* **2020**, *20*, 35–54.
15. Goodrich, K.H.; Theodore, C.R. Description of the NASA urban air mobility maturity level (UML) scale. In Proceedings of the AIAA Scitech 2021 Forum, Virtual Event, 11–21 January 2021.
16. Gil, G.H.; Kaber, D.; Kaufmann, K.; Kim, S.H. Effects of Modes of Cockpit Automation on Pilot Performance and Workload in a Next Generation Flight Concept of Operation. *Hum. Factors Ergon. Manuf. Serv. Ind.* **2012**, *22*, 395–406. [CrossRef]
17. Degani, A.; Shmueli, Y.; Bnaya, Z. Equilibrium of Control in Automated Vehicles: Driver Engagement Level and Automation Capability Levels. In Proceedings of the 4th IFAC Workshop on Cyber-Physical and Human Systems (CPHS), Houston, TX, USA, 1–2 December 2022.
18. Mueller, A.S.; Reagan, I.J.; Cicchino, J.B. Addressing Driver Disengagement and Proper System Use: Human Factors Recommendations for Level 2 Driving Automation Design. *J. Cogn. Eng. Decis. Mak.* **2021**, *15*, 3–27. [CrossRef]
19. Vempati, L.; Woods, S.; Winter, S.R. Pilots' willingness to operate in urban air mobility integrated airspace: A moderated mediation analysis. *Drone Syst. Appl.* **2022**, *2*, 59–76. [CrossRef]
20. Al Haddad, C.; Chaniotakis, E.; Straubinger, A.; Plötner, K.; Antoniou, C. Factors affecting the adoption and use of urban air mobility. *Transp. Res. Part A Policy Pract.* **2020**, *132*, 696–712. [CrossRef]
21. Khastgir, S.; Birrell, S.; Dhadyalla, G.; Jennings, P. Calibrating trust through knowledge: Introducing the concept of informed safety for automation in vehicles. *Transp. Res. Part C Emerg. Technol.* **2018**, *96*, 290–303. [CrossRef]
22. Lu, Y.; Sarter, N. Eye Tracking: A Process-Oriented Method for Inferring Trust in Automation as a Function of Priming and System Reliability. *IEEE Trans. Hum. Mach. Syst.* **2019**, *49*, 560–568. [CrossRef]
23. Drewitz, U.; Wilbrink, M.; Oehl, M.; Jipp, M.; Ihme, K. Subjective certainty to increase the acceptance of automated and connected driving. *Forsch. Im Ingenieurwesen-Eng. Res.* **2021**, *85*, 997–1012. [CrossRef] [PubMed]
24. Bowden, V.K.; Griffiths, N.; Strickland, L.; Loft, S. Detecting a Single Automation Failure: The Impact of Expected (But Not Experienced) Automation Reliability. *Hum. Factors* **2023**, *65*, 533–545. [CrossRef] [PubMed]
25. Barg-Walkow, L.H.; Rogers, W.A. The Effect of Incorrect Reliability Information on Expectations, Perceptions, and Use of Automation. *Hum. Factors* **2016**, *58*, 242–260. [CrossRef]
26. Dsouza, K.M.; Dang, T.; Metcalfe, J.S.; Bhattacharya, S. Brain-Based Indicators of Passenger Trust During Open-Road Driving. In Proceedings of the VTC2021-Fall, Virtual Event, 27 September–8 October 2021.

27. He, X.; Stapel, J.; Wang, M.; Happee, R. Modelling perceived risk and trust in driving automation reacting to merging and braking vehicles. *Transp. Res. Part F Traffic Psychol. Behav.* **2022**, *86*, 178–195. [CrossRef]

28. Perello-March, J.R.; Burns, C.G.; Woodman, R.; Elliott, M.T.; Birrell, S.A. Using fNIRS to Verify Trust in Highly Automated Driving. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 739–751. [CrossRef]

29. Michael, S.; John, K.; Thomas, L.; Kimberlee, S.; Loran, H. Evaluation of Novel eVTOL Aircraft Automaton Concepts. In Proceedings of the AIAA AVIATION Forum and Exposition, San Diego, CA, USA, 12–16 June 2023.

30. Mrinmoy, S.; Xuyang, Y.; Abenezer, G.; Abdollah, H. A Comprehensive eVTOL Performance Evaluation Framework in Urban Air Mobility. *Intell. Syst. Appl.* **2022**, *542*, 459–474.

31. Lebiere, C.; Blaha, L.M.; Fallon, C.K.; Jefferson, B. Adaptive Cognitive Mechanisms to Maintain Calibrated Trust and Reliance in Automation. *Front. Robot. AI* **2021**, *8*, 652776. [CrossRef]

32. Liu, Y.; Ayaz, H.; Shewokis, P.A. Multisubject "Learning" for Mental Workload Classification Using Concurrent EEG, fNIRS, and Physiological Measures. *Front. Hum. Neurosci.* **2017**, *11*, 389. [CrossRef] [PubMed]

33. Kakkos, I.; Dimitrakopoulos, G.N.; Sun, Y.; Yuan, J.; Matsopoulos, G.K.; Bezerianos, A.; Sun, Y. EEG Fingerprints of Task-Independent Mental Workload Discrimination. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 3824–3833. [CrossRef] [PubMed]

34. Lee, D.H.; Jeong, J.H.; Kim, K.; Yu, B.W.; Lee, S.W. Continuous EEG Decoding of Pilots' Mental States Using Multiple Feature Block-Based Convolutional Neural Network. *IEEE Access* **2020**, *8*, 121929–121941. [CrossRef]

35. Liu, S.; Wang, X.; Zhao, L.; Li, B.; Hu, W.; Yu, J.; Zhang, Y.D. 3DCANN: A Spatio-Temporal Convolution Attention Neural Network for EEG Emotion Recognition. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 5321–5331. [CrossRef]

36. Lin, J.; Chen, Y.; Xie, J.; Mo, L. Altered Brain Connectivity Patterns of Individual Differences in Insightful Problem Solving. *Front. Behav. Neurosci.* **2022**, *16*, 905806. [CrossRef]

37. Choo, S.; Nam, C. Detecting Human Trust Calibration in Automation: A Convolutional Neural Network Approach. *IEEE Trans. Hum. Mach. Syst.* **2022**, *52*, 774–783. [CrossRef]

38. Kim, H.S.; Hwang, Y.; Yoon, D.; Choi, W.; Park, C.H. Driver Workload Characteristics Analysis Using EEG Data from an Urban Road. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 1844–1849. [CrossRef]

39. Hossain, K.M.; Islam, A.; Hossain, S.; Nijholt, A.; Ahad, A.R. Status of deep learning for EEG-based brain–computer interface applications. *Front. Comput. Neurosci.* **2023**, *16*, 1006763. [CrossRef]

40. Wang, Y.; Zhang, L.; Xia, P.; Wang, P.; Chen, X.; Du, L.; Fang, Z.; Du, M. EEG-Based Emotion Recognition Using a 2D CNN with Different Kernels. *Bioengineering* **2022**, *9*, 231. [CrossRef]

41. Lee, J.D.; See, K.A. Trust in automation: Designing for appropriate reliance. *Hum. Factors* **2004**, *46*, 50–80. [CrossRef]

42. Blue, P.R.; Hu, J.; Zhou, X. Higher Status Honesty Is Worth More: The Effect of Social Status on Honesty Evaluation. *Front. Psychol.* **2018**, *9*, 350. [CrossRef]

43. Yang, J.; Qiu, J.; Zhang, Q. The neural basis of risky decision making in a blackjack task. *Int. J. Psychol.* **2008**, *43*, 819–820. [CrossRef]

44. Long, Y.; Jiang, X.; Zhou, X. To believe or not to believe: Trust choice modulates brain responses in outcome evaluation. *Neuroscience* **2012**, *200*, 50–58. [CrossRef]

45. Pang, Y.; Hu, J.; Lieber, C.; Cooke, N.; Liu, Y. Air traffic controller workload level prediction using conformalized dynamical graph learning. *Adv. Eng. Inform.* **2023**, *57*, 102113. [CrossRef]

46. aeroG Aviation. Available online: https://aerogaviation.com/x-plane-11-%26-p3d-aircraft (accessed on 28 July 2023).

47. Kaggle. Available online: https://www.kaggle.com/datasets/leonieschneider/p300-dataset-for-workload-detection (accessed on 24 July 2023).

48. Palmer, A.R.; Distefano, R.; Leneman, K.; Berry, D. Reliability of the BodyGuard2 (FirstBeat) in the Detection of Heart Rate Variability. *Appl. Psychophysiol. Biofeedback* **2021**, *46*, 251–258. [CrossRef]

49. Xiang, S.; Xie, A.; Ye, M.; Yan, X.; Han, X.; Niu, H.; Li, Q.; Huang, H. Autonomous eVTOL: A summary of researches and challenges. *Green Energy Intell. Transp.* **2024**, *3*, 100140. [CrossRef]

50. Horat, S.K.; Herrmann, F.R.; Favre, G.; Terzis, J.; Debatisse, D.; Merlo, M.C.G.; Missonnier, P. Assessment of mental workload: A new electrophysiological method based on intra-block averaging of ERP amplitudes. *Neuropsychologia* **2016**, *82*, 11–17. [CrossRef]

51. van Dinteren, R.; Arns, M.; Jongsma, M.L.A.; Kessels, R.P.C. Combined frontal and parietal P300 amplitudes indicate compensated cognitive processing across the lifespan. *Front. Aging Neurosci.* **2014**, *6*, 294. [CrossRef] [PubMed]

52. Evans, D.C.; Fendley, M. A multi-measure approach for connecting cognitive workload and automation. *Int. J. Hum. Comput. Stud.* **2017**, *97*, 182–189. [CrossRef]

53. Asgher, U.; Khalil, K.; Khan, M.J.; Ahmad, R.; Butt, S.I.; Ayaz, Y.; Naseer, N.; Nazir, S. Enhanced Accuracy for Multiclass Mental Workload Detection Using Long Short-Term Memory for Brain–Computer Interface. *Front. Neurosci.* **2020**, *14*, 584. [CrossRef] [PubMed]

54. Gramfort, A.; Luessi, M.; Larson, E.; Engemann, D.A.; Strohmeier, D.; Brodbeck, C.; Hnen, M.S. MNE software for processing MEG and EEG data. *Neuroimage* **2014**, *86*, 446–460. [CrossRef]

55. Haresign, I.M.; Phillips, E.; Whitehorn, M.; Noreika, V.; Jones, E.; Leong, V.; Wass, S. Automatic classification of ICA components from infant EEG using MARA. *Dev. Cogn. Neurosci.* **2021**, *52*, 101024. [CrossRef]

56. Boudewyn, M.A.; Luck, S.J.; Farrens, J.L.; Kappenman, E.S. How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology* **2018**, *55*, 13049. [CrossRef]
57. Shafiei, S.B.; Hussein, A.A.; Muldoon, S.F.; Guru, K.A. Functional Brain States Measure Mentor-Trainee Trust during Robot-Assisted Surgery. *Sci. Rep.* **2018**, *8*, 3667. [CrossRef] [PubMed]
58. Rajwal, S.; Aggarwal, S. Convolutional Neural Network-Based EEG Signal Analysis: A Systematic Review. *Arch. Comput. Methods Eng.* **2023**, *30*, 3585–3615. [CrossRef]
59. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **2018**, *9*, 611–629. [CrossRef] [PubMed]
60. Kadhim, K.A.; Mohamed, F.; Sakran, A.A.; Adnan, M.M.; Salman, G.A. Early Diagnosis of Alzheimer's Disease using Convolutional Neural Network-based MRI. *Malays. J. Fundam. Appl. Sci.* **2023**, *19*, 362–368. [CrossRef]
61. Zou, Z.B.; Ergan, S. Towards emotionally intelligent buildings: A Convolutional neural network based approach to classify human emotional experience in virtual built environments. *Adv. Eng. Inform.* **2023**, *55*, 101868. [CrossRef]
62. Huang, J.X.; Hsieh, C.Y.; Huang, Y.L.; Wei, C.S. Toward CNN-Based Motor-Imagery EEG Classification with Fuzzy Fusion. *Int. J. Fuzzy Syst.* **2022**, *24*, 3812–3823. [CrossRef]
63. Sun, Y.; Lo, F.P.W.; Lo, B. EEG-based user identification system using 1D-convolutional long short-term memory neural networks. *Expert Syst. Appl.* **2019**, *125*, 259–267. [CrossRef]
64. Wang, J.; Cheng, S.; Tian, J.; Gao, Y. A 2D CNN-LSTM hybrid algorithm using time series segments of EEG data for motor imagery classification. *Biomed. Signal Process. Control.* **2023**, *83*, 104627. [CrossRef]
65. Escottá, T.; Beccaro, W.; Ramírez, M.A. Evaluation of 1D and 2D Deep Convolutional Neural Networks for Driving Event Recognition. *Sensors* **2022**, *22*, 4226. [CrossRef] [PubMed]
66. Zhou, H.; Zhao, X.; Zhang, H.; Kuang, S. The mechanism of a multi-branch structure for EEG-based motor imagery classification. In Proceedings of the 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO), Dali, China, 6–8 December 2019.
67. Shaw, P.; Uszkoreit, J.; Vaswani, A. Self-attention with relative position representations. *arXiv* **2018**, arXiv:1803.02155.
68. O'Shea, K.; Nash, R. An introduction to convolutional neural networks. *arXiv* **2015**, arXiv:1511.08458.
69. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference for Learning Representations, San Diego, CA, USA, 7–9 May 2015.
70. Caruana, R.; Lawrence, S.; Giles, L. Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. *Adv. Neural Inf. Process. Syst.* **2001**, *13*, 402–408.
71. Lavie, A.; Sagae, K.; Jayaraman, S. The significance of recall in automatic metrics for MT evaluation. In Proceedings of the Machine Translation: From Real Users to Research: 6th Conference of the Association for Machine Translation in the Americas, AMTA 2004, Washington, DC, USA, 28 September–2 October 2004; pp. 134–143.
72. Arlot, S.; Celisse, A. A survey of cross-validation procedures for model selection. *Statist. Surv.* **2020**, *4*, 40–79. [CrossRef]
73. Ganaie, M.A.; Kumari, A.; Malik, A.K.; Tanveer, M. EEG signal classification using improved intuitionistic fuzzy twin support vector machines. *Neural Comput. Appl.* **2022**, *36*, 163–179. [CrossRef]
74. Rafiammal, S.; Najumnissa Jamal, D.; Kaja Mohideen, S. Detection of Epilepsy Seizure in Adults Using Discrete Wavelet Transform and Cluster Nearest Neighborhood Classifier. *Iran. J. Sci. Technol. Trans. Electr. Eng.* **2021**, *45*, 1103–1115. [CrossRef]
75. Rabby, M.K.M.; Eshun, R.B.; Belkasim, S.; Islam, A.K.M.K. Epileptic Seizure Detection Using EEG Signal Based LSTM Models. In Proceedings of the 2021 IEEE Fourth International Conference on Artificial Intelligence and Knowledge Engineering, Virtual Event, 1–3 December 2021.
76. Feurer, M.; Eggensperger, K.; Falkner, S.; Lindauer, M.; Hutter, F. Auto-sklearn 2.0: Hands-free automl via meta-learning. *J. Mach. Learn. Res.* **2022**, *23*, 11936–11996.
77. Du, N.; Haspiel, J.; Zhang, Q.; Tilbury, D.; Pradhan, A.K.; Yang, X.J.; Robert, L.P. Look who's talking now: Implications of AV's explanations on driver's trust, AV preference, anxiety and mental workload. *Transp. Res. Part C Emerg. Technol.* **2019**, *104*, 428–442. [CrossRef]