

Article

EEG-Based Emotion Classification in Financial Trading Using Deep Learning: Effects of Risk Control Measures

Bhaskar Tripathi * and Rakesh Kumar Sharma 

School of Humanities and Social Sciences, Thapar Institute of Engineering and Technology, Patiala 147004, India

* Correspondence: btripathi_phdp16@thapar.edu

Abstract: Day traders in the financial markets are under constant pressure to make rapid decisions and limit capital losses in response to fluctuating market prices. As such, their emotional state can greatly influence their decision-making, leading to suboptimal outcomes in volatile market conditions. Despite the use of risk control measures such as stop loss and limit orders, it is unclear if these strategies have a substantial impact on the emotional state of traders. In this paper, we aim to determine if the use of limit orders and stop loss has a significant impact on the emotional state of traders compared to when these risk control measures are not applied. The paper provides a technical framework for valence-arousal classification in financial trading using EEG data and deep learning algorithms. We conducted two experiments: the first experiment employed predetermined stop loss and limit orders to lock in profit and risk objectives, while the second experiment did not employ limit orders or stop losses. We also proposed a novel hybrid neural architecture that integrates a Conditional Random Field with a CNN-BiLSTM model and employs Bayesian Optimization to systematically determine the optimal hyperparameters. The best model in the framework obtained classification accuracies of 85.65% and 85.05% in the two experiments, outperforming previous studies. Results indicate that the emotions associated with Low Valence and High Arousal, such as fear and worry, were more prevalent in the second experiment. The emotions associated with High Valence and High Arousal, such as hope, were more prevalent in the first experiment employing limit orders and stop loss. In contrast, High Valence and Low Arousal (calmness) emotions were most prominent in the control group which did not engage in trading activities. Our results demonstrate the efficacy of our proposed framework for emotion classification in financial trading and aid in the risk-related decision-making abilities of day traders. Further, we present the limitations of the current work and directions for future research.

Keywords: behavioral finance; emotion classification; deep learning; electroencephalography (EEG); neuro-finance; decision-making



Citation: Tripathi, B.; Sharma, R.K. EEG-Based Emotion Classification in Financial Trading Using Deep Learning: Effects of Risk Control Measures. *Sensors* **2023**, *23*, 3474. <https://doi.org/10.3390/s23073474>

Academic Editor: Georg Fischer

Received: 14 January 2023

Revised: 14 March 2023

Accepted: 22 March 2023

Published: 26 March 2023



Copyright: © 2023 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/>).

1. Introduction

Emotions are physiological states associated with the neurological system that influence feelings and rational behavior [1]. Emotions and their management are seen as essential factors for efficient and intelligent decision-making [2]. Affective computing is an area of artificial intelligence that focuses on human-computer interactions, such as recognizing human behavior and emotional states [3].

In recent years, automatic emotion detection has been applied in various areas such as emotion recognition from movies [4], audio [5], text [6], and facial expressions [7]. With the development of low-cost wearable technology, non-invasive electroencephalography (EEG)-based techniques for automatic emotion identification have achieved widespread popularity and acceptance [8]. With their high temporal precision, EEG waves directly represent the brain's neural activity. The information these signals provide is more dependable than that provided by facial expressions or text since they cannot be falsified or replicated to feign an emotional state [9]. High-speed computing has allowed machine learning

techniques to be applied with EEG data to detect emotions even more accurately [10]. EEG has been widely used in areas such as medical research to assess brain function and neurological conditions. In recent years, machine learning (ML) models have been applied to EEG signals with explainable AI to develop prediction models for diseases such as stroke and to evaluate mental workload. Examples of such applications include based stroke prediction [11,12], assessment of task-induced neurological outcomes after stroke [13], detection of driving-induced neurological biomarkers [14], and prediction of sleep stages based on EEG biomarkers [15]. In emotion recognition, each emotion combines valence (a spectrum of negative to positive emotions) and arousal (the intensity) associated with the nervous system [16].

In financial markets, day traders are motivated to profit from the market's price movement in the shortest time feasible while avoiding capital loss. During high-pressure trading sessions, they seek to adjust to changing market conditions and respond to fluctuating market conditions, such as volatility, trend reversals, market sentiment, and events, such as profits and losses. These responses consistently elicit feelings including fear, greed, hope, calmness, and regret [17,18]. Depending on the trade performance, the emotional intensity of traders fluctuates at various periods and may influence their rational decision-making [19]. Day traders operate in a volatile market environment and make use of market price movements to enter or exit their positions. However, they are constantly pressured to make split-second judgments. An erroneous entry or exit decision can lead to significant capital loss [20]. Understanding and classifying their emotions can help them comprehend their emotional state and reactions to various events and make rational decisions to minimize future losses. The psychological and emotional effects of simulated trading can be comparable to actual trading, as most traders attach emotional importance to money-related decisions [17]. It has been observed in past literature that some professional traders tend to hold their positions for too long and sell their winning investments too soon. In addition, many traders apply limit orders to obtain better prices for their assets [21] and use stop-loss strategies to limit excessive losses [22].

Being confronted with regular market volatility, risks, and significant pressure to generate profits while avoiding loss, traders require a systematized framework with ambient intelligence to comprehend their emotional state when making crucial financial judgments to optimize their rational decision-making capacity under market risks [23]. In recent years, many studies have applied deep learning techniques to analyze EEG signals and use them for emotion classification [8,24–27]. However, the literature on the application of EEG-based analysis in the context of the stock market and financial trading is relatively limited [10,17,28,29]. There is a need for research on the identification and classification of the emotions of day traders in order to gain a deeper understanding of their emotional states and reactions to the results of their trading decisions. Such understanding has the potential to aid in the formulation of strategies that can assist day traders in making more rational decisions, ultimately leading to a reduction in future losses. This is particularly relevant in the context of volatile markets and the high-pressure environment in which day traders operate, where the ability to make informed decisions is crucial for minimizing capital losses.

The objective of this research paper is to determine the valence-arousal state of traders when they use limit orders and stop loss to their trades and compare their emotional states when these risk control measures are not applied. Our research question investigates whether there is a substantial difference between traders' emotional states in the two scenarios. We propose a novel framework for multi-trader emotion classification that uses brainwave signals and deep learning algorithms and signal processing methods. We contribute to the electroencephalography-based emotion classification literature in financial trading by creating an EEG database of participants executing trades with and without stop loss and limit orders in two experiments to classify their emotional states. We recorded participants' emotional state using a Self-assessment Manikin in a valence-arousal environment to answer our research question. The main aims of this research are:

- (1) To provide a technical framework for emotion classification in financial trading using EEG data and deep learning algorithms.
- (2) To determine the most appropriate neural network architecture and optimize it for improved EEG emotion classification and to achieve state-of-the-art accuracy.
- (3) To investigate the influence of limit orders and stop loss on traders' emotional states and to generate an electroencephalography dataset for financial trading scenarios.

1.1. Contributions

The major contributions of this work are summarized as follows:

1. This paper contributes to the field of emotion classification in financial trading by presenting an EEG database collected from 20 participants executing real-time trade and a framework for emotion classification utilizing EEG data and deep learning algorithms.
2. A hybrid neural architecture is proposed and found to outperform previous studies.
3. The study provides evidence that the use of risk control measures (limit orders and stop loss) has a significant impact on the emotional state of traders, and these measures can provoke affective states associated with trading.
4. The empirical results demonstrate that the trading groups exhibit significantly different affective states from the control group.

1.2. Paper Structure

The remaining sections of the paper are organized as follows: Section 2 discusses the data collection procedure, data pre-processing, and the experimental design for the participants; Section 3 discusses the proposed framework and the deep learning-based emotion classification strategies utilized; Section 4 analyzes the empirical results; and Section 5 concludes the paper.

2. Dataset

2.1. Study Participants

We used electroencephalography (EEG) data in this study to validate the self-reported emotional states of twenty healthy individuals ($n = 20$, 10 males and 10 females, aged 25 to 60). All participants had prior stock trading experience to maintain the homogeneity of the group. Each participant was subjected to a single trial of two experiments of 30 min each. Prior to conducting the experiments, participants received an hour of hands-on training in the user interface of the simulated trading environment.

2.2. Data Acquisition and Recording Process

The EEG data were collected using the Interaxon Muse 2 headset. The Muse 2 EEG headset records brainwaves from four channels (AF7, AF8, TP9, and TP10) and a reference electrode (FPz) that acquires data from the Frontal (AF7 and AF8) and temporoparietal (TP9 and TP10) regions of the brain. Figure 1 depicts the Interaxon Muse 2 EEG headset used for recording brainwave signals. A new column was added to the dataset to label emotions at specified timestamps. The raw EEG signals recorded by the Muse 2 headband were transmitted from a mobile application, called Mind Monitor, to a Windows 10 machine using Open Sound Control streaming on port 5000 and saved in a flat file. To facilitate the downsampling of the data and generate a uniform stream frequency, all signals were recorded with a Unix timestamp. To minimize external noise and distractions, participants were not exposed to any other external noise or disturbances while conducting the study.

To ensure the integrity of the EEG signals, participants were instructed to minimize their movement and to keep their eyes open throughout the tasks. When looking at technical indicator charts, eye blink variability and lateral eye movement events can be associated with varied attentiveness levels; classification algorithms can account for these patterns of signal spikes [30]. The eye blink rate, which may affect AF7 and AF8 sensors, was not encouraged or discouraged to maintain a natural condition.

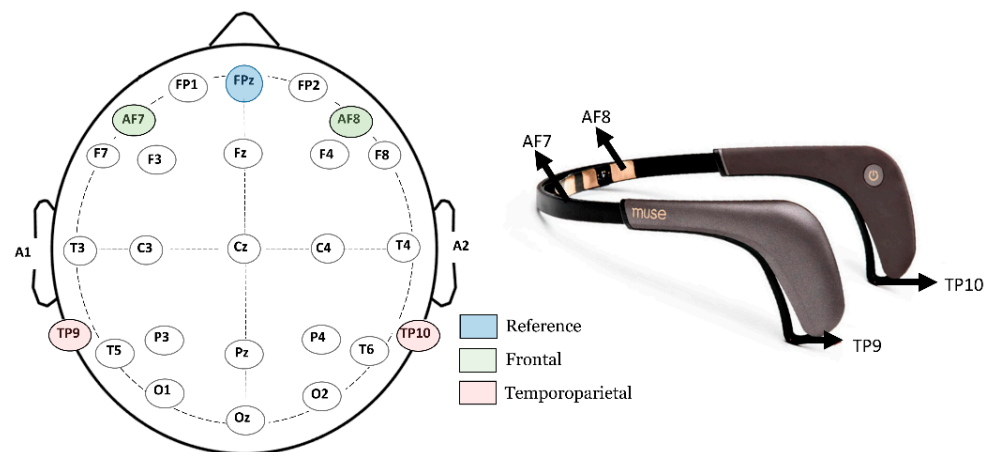


Figure 1. The four channels AF7, AF8, TP9, and TP10, highlighted in the 10–20 electrode placement scheme, and the EEG headband utilized to acquire EEG data.

Throughout the 30-min experiment, participants were shown a Self-Assessment Manikin (SAM) [31] every minute and asked to rate their emotions on a scale of 1 to 10. The SAM is a tool that facilitates the rapid labeling of emotions without inherent bias and with less participant fatigue. These self-reported emotions were used as labels by the emotion classification algorithms. We used the digital version of SAM [32] in this study. Figure 2 shows the categorization of emotions that are captured according to the valence-arousal model via a SAM. Valence refers to the positive or negative dimension of emotion, while arousal refers to the degree or strength of the emotion. The combinations of valence and arousal can be transposed into several emotional states, namely High Valence High Arousal (HVHA), Low Valence High Arousal (LVHA), Low Valence Low Arousal (LVLA), and High Valence Low Arousal (HVLA). Hope, fear, regret, and calmness are the high-level emotions described by the valence-arousal framework, as depicted in Figure 2.

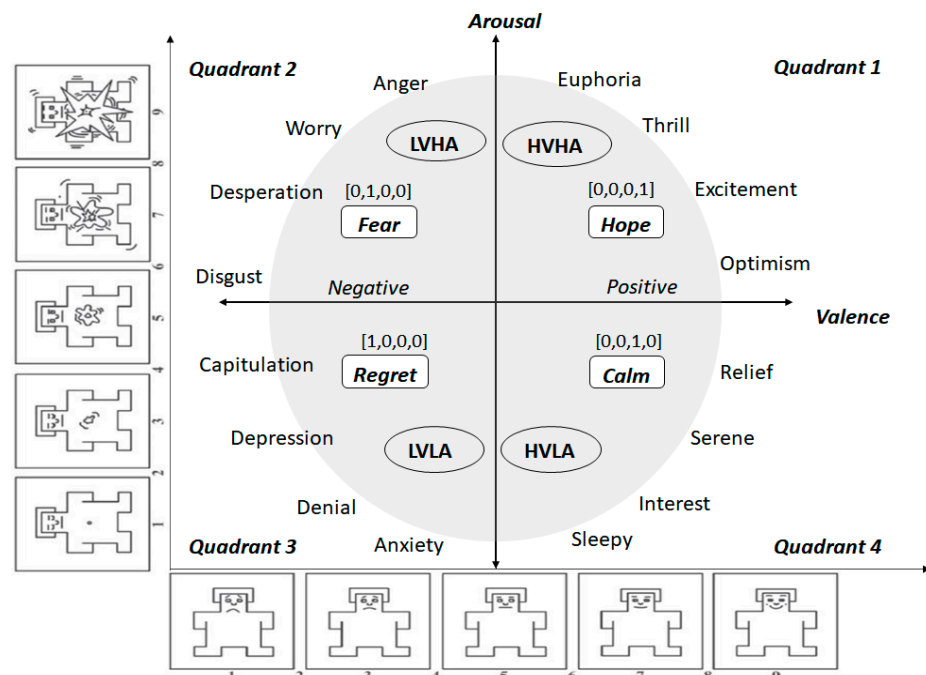


Figure 2. A schematic illustration of the self-labeling of emotions using SAM within the framework of valence and arousal.

3. Proposed Methodology

This section describes the data pre-processing, feature extraction, emotion classification processes, and comparison of performance metrics with the benchmark dataset. The proposed framework for emotion classification consists of several essential steps to ensure high accuracy in emotion recognition. The first step involves the acquisition and recording of EEG signals along with the self-labeling of emotions using a Self-Assessment Manikin. Next, the acquired data undergoes a thorough cleaning and pre-processing stage described in Section 3.1. This step includes filtering and artifact removal at multiple levels to ensure that the data is high quality and free from noise and other disturbances that could affect the results. The pre-processed data undergoes feature extraction through the Independent Component Analysis (ICA) of various frequency bands, including gamma, beta, alpha, theta, and delta. This step is crucial as it helps to extract the most relevant information from the EEG signals for further analysis. Following feature extraction, the data is split into a training set and a testing set; the training set is then used to train various neural network models, including LSTM, CNN, CNN-LSTM, CNN-BiLSTM, and CNN-BiLSTM-CRF. Utilizing Bayesian Optimization, the hyperparameters of these classifiers are optimized to ensure their ideal performance. The final step in the proposed framework involves classifying emotions into the HVHA, LVHA, LVLA, and HVLA classes using the trained models. The performance of the models is then evaluated using various performance metrics, including accuracy, precision, sensitivity, specificity, F-score, ROC, and AUC, and compared with the previous literature. Figure 3 illustrates a self-explanatory flowchart of the overall experimental design approach.

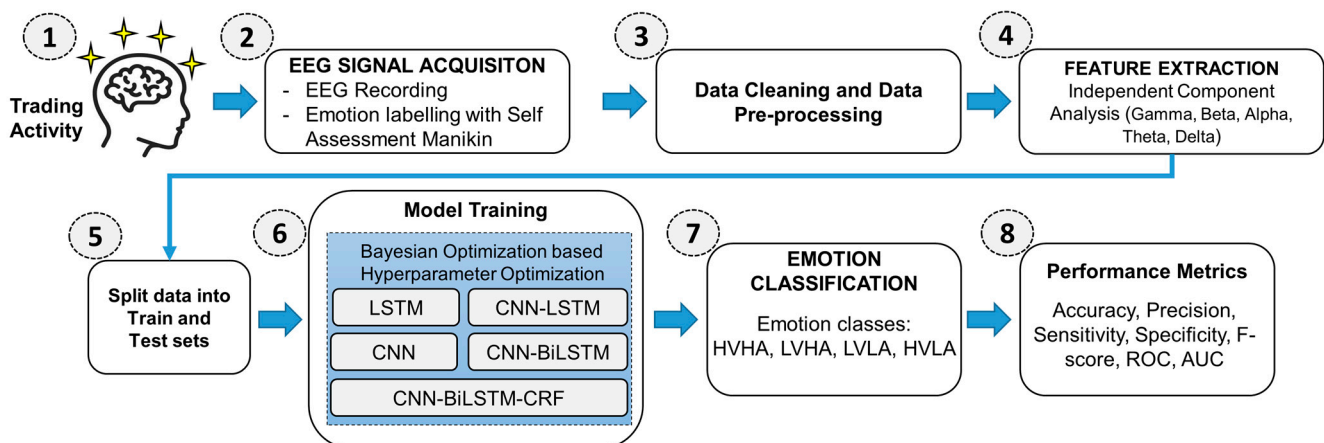


Figure 3. Workflow schematic of the proposed emotion classification framework.

3.1. Pre-Processing of EEG Data

We applied the following process for EEG data preprocessing and artifact removal. Figure 4 illustrates the steps involved in the data preprocessing procedure:

1. Data collection: We imported the raw EEG data recorded from the Muse 2 headset, streamed it from the Mind Monitor application (mind-monitor.com, accessed on 22 November 2022), and saved it as a comma-separated file.
2. Downsampling and data cleaning: The raw EEG data was recorded at 256 Hz, but we downsampled it to 128 Hz using EEGLab software, version 2022.1. To ensure high-quality data, we removed the first 30 s of data from the beginning of each participant's signal which was designated for relaxation. We also removed non-informative components of the signal, such as missing data, non-numeric values, and variables related to the heart rate, accelerometer, and gyroscope that were collected but not studied.
3. Low pass and high pass filtering: During EEG recording, the signals are interrupted by artifacts caused by eyeblinks, heart rate, and muscular movements such as jaw

clenches [33]. We eliminated these artifacts from the resampled signal using two Butterworth IIR (Infinite Impulse Response) filters given by the following expression:

$$H(s) = 1/(1 + (s/\omega c)(2n)) \quad (1)$$

where $H(s)$ is the transfer function of the filter, s is the complex frequency variable, ωc is the cutoff frequency of the filter, and n is the order of the filter (i.e., the number of poles in the filter).

We used two Butterworth IIR filters to eliminate artifacts from the resampled signal caused by eyeblinks, heart rate, and muscular movements. A low-pass filter was applied at 45 Hz and then a high-pass filter at 1 Hz. Before filtering, we visually inspected the data and deleted any parts containing significant artifacts to minimize the spread of these artifacts during filtering. We only filtered continuous data segments and avoided filtering across boundaries to prevent filtering artifacts.

4. Notch filtering: To reduce noise from the 50 Hz frequency in the EEG signals, we used a notch filter to eliminate line noise originating from the electrical power supply. This step ensured that the EEG data was free from such noise and ready for further analysis.
5. Channel rejection: We conducted checks to remove faulty channels by examining channels with no EEG activity for more than 5 s, channels with high noise (high standard deviation relative to other channels), and channels with low correlation with other channels (low correlation to other channels using a rejection threshold of 0.70).
6. Automated Signal rejection (ASR) algorithm: We employed the ASR algorithm [34] for additional rejection of bad data segments based on the number of channels exceeding a 20 (default value) standard deviation threshold within a time window of 5 s. This allowed us to further reject portions of data that may have been missed by the earlier steps.
7. Independent Component Analysis: We applied the ICA method [35] to further minimize the effects of artifacts such as ECG, EMG, EOG, and others that may still be present in the data despite multiple levels of cleaning. ICA is a successful method for dealing with artifacts such as muscle movements, eye blinks, lateral eye movements, heart rate, and others. However, to avoid strong artifacts affecting the data, we filtered them out in previous steps as ICA may not be able to remove them effectively.

Visual inspection: Finally, we removed any remaining artifacts manually using the visual inspection method in EEGLab with a sliding window of 30 s. Figures A1 and A2 in the Appendix A show the EEG signals before and after visual inspection and artifact rejection.

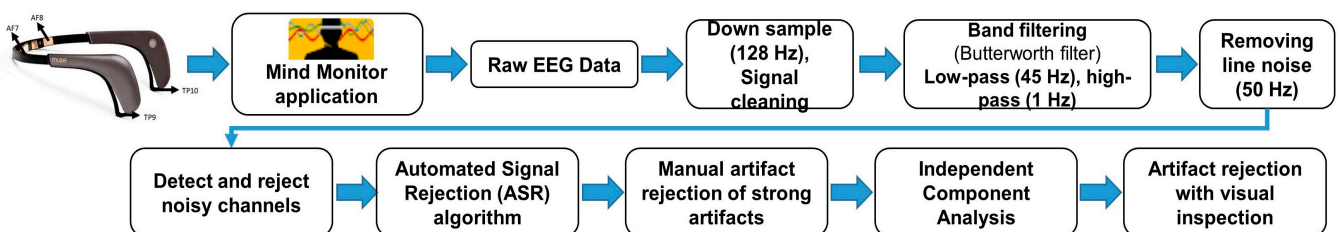


Figure 4. Schematic Illustration of data pre-processing for artifact removal and signal denoising.

3.2. Feature Extraction

After obtaining the cleaned EEG signals, we applied Independent Component Analysis (ICA) [35] on all four recorded channels to extract the features. ICA is a feature extraction method that decomposes a multivariate signal into independent components. In addition to the extraction of features, ICA also eliminates embedded artifacts that could be missed out in the manual artifact removal process. The ICA algorithm decomposes a non-stationary signal contained with several mixed frequencies into distinct, independent components

(signals), each corresponding to a different frequency band. Each frequency band is associated with different emotional states. Table 1 shows the power spectrum of frequency bands that are associated with different emotional states [36]. The ICA method extracts gamma, beta, alpha, theta, and delta frequencies for each of the four channels. Consequently, each experiment with a single trial generated 20 features per dataset.

Table 1. Emotional state association of brainwaves at different frequency bands.

Frequency Band	Frequency Range	Emotional State Association
Gamma	>30 Hz	Problem-solving, concentrating, associated with positive valence. Arousal increases with high-intensity stimuli
Beta	13–20 Hz	Awake, busy, normal activities, and alert
Alpha	8–13 Hz	Relaxed, calm, and reflective
Theta	4–7 Hz	Deep relaxation
Delta	0–4 Hz	REM sleep and dreaming

3.3. Emotion Classification Algorithms

This section discusses the neural network classifiers used for emotion classification following the input data's successful preprocessing and feature extraction. To perform emotion classification, we used five different types of neural networks, namely, LSTM, CNN, CNN-LSTM, CNN-BiLSTM, and CNN-BiLSTM-CRF. Bayesian Optimization optimizes the hyperparameters of these classifiers.

3.3.1. Long Short-Term Memory

Long Short-Term Memory (LSTM) is a type of recurrent neural network that is designed to model long-term dependencies in sequential data. Each LSTM cell takes in a sequence of input vectors x_1, x_2, \dots, x_T and produces a corresponding sequence of output vectors h_1, h_2, \dots, h_T . The input vectors x_t and output vectors h_t at each time step t can have different shapes and dimensions depending on the specific task and dataset. Mathematically, the LSTM model can be represented by the below equations:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (2)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (3)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (4)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (5)$$

$$h_t = o_t \tanh(c_t) \quad (6)$$

where i_t, f_t, c_t , and o_t are intermediate variables; h_t represents the output of the LSTM cell at time step t ; x_t and h_{t-1} are the input data and previous hidden state at time step t , respectively; U and W are weight matrices, and b is a bias vector. The sigmoid function is represented by σ , and the hyperbolic tangent function is represented by \tanh .

3.3.2. Convolutional Neural Network

A Convolutional Neural Network (CNN) consists of three layers, namely, a convolutional layer, a pooling layer, and a fully connected layer [37]. The output of every layer is connected to a small neighborhood in the input through a weight matrix also known as a filter or kernel. Each filter is moved around the input giving rise to one 2D output.

The outputs corresponding to each filter are stacked, giving rise to an output volume. The convolution operation of a 2D input signal $x(p, q)$ can be expressed as follows:

$$y(m, n) = \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} x(p, q)h(m - p, n - q) \quad (7)$$

where $y(m, n)$ is the output obtained after the convolution operation and h is the kernel coefficient of the system.

Following a series of convolutions, there is a pooling layer that is used to reduce the dimensionality of feature vectors. The main job of a pooling layer is to provide translational invariance by subsampling the input feature vector. The pooling layers are followed by a fully connected layer that acts as a classifier. Figure 5 shows the basic structure of a CNN model.

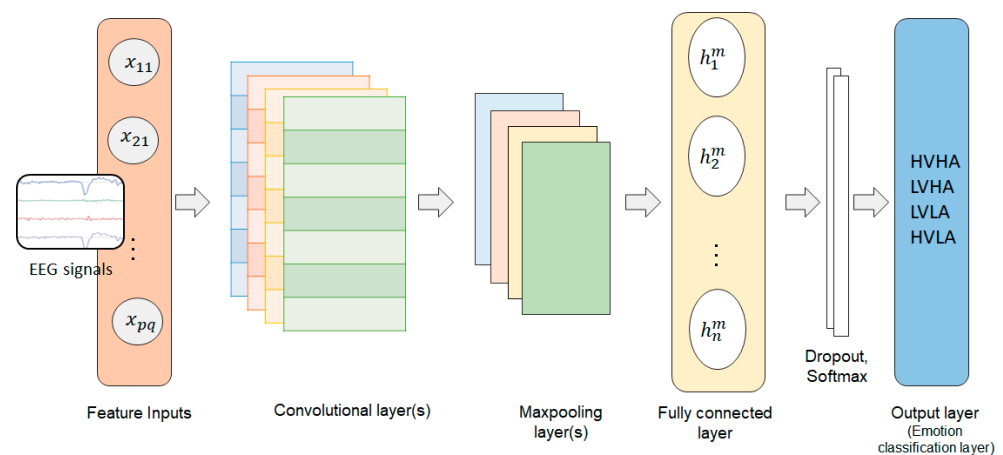


Figure 5. Block diagram of a convolutional neural network.

3.3.3. CNN-LSTM

A CNN-LSTM model integrates the CNN model for feature selection with a Long Short-Term Memory (LSTM) model for time series prediction. A convolutional layer in this model identifies the important features in the input data [38]. Pooling reduces the data dimensions. The pooling layer interfaces with LSTM and linked layers (also known as dense layers). The combined CNN-LSTM model can be expressed as follows:

$$h_t = LSTM(x_t, h_{t-1}) \quad (8)$$

$$y_t = CNN(h_t) \quad (9)$$

where x_t represents the input data at time t , h_{t-1} represents the hidden state of the LSTM at time $t - 1$, and h_t represents the hidden state of the LSTM at time t .

3.3.4. CNN-BiLSTM

A CNN-BiLSTM [39] model incorporates the CNN for feature selection and the bidirectional LSTM (BiLSTM) model for time series forecasting. A BiLSTM model has an auxiliary LSTM layer that reverses the information flow. As the direction of the input sequence is reversed due to the additional LSTM layer, the model can consider incoming information from both forward and reverse directions. Figure 6 shows a schematic structure of a CNN model combined with an LSTM or BiLSTM model.

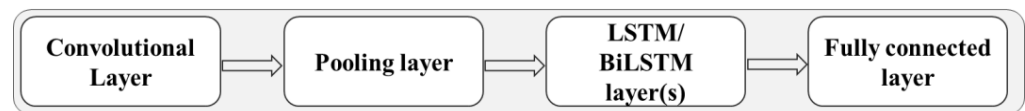


Figure 6. CNN structure depicting the interface between the pooling layer and an LSTM or BiLSTM.

The CNN-BiLSTM model can be expressed as follows:

$$h_t^f = LSTM(x_t, h_{t-1}^f) \quad (10)$$

$$h_t^b = LSTM(x_t, h_{t+1}^b) \quad (11)$$

$$y_t = CNN([h_t^f, h_t^b]) \quad (12)$$

where x_t represents the input data at time t , h_{t-1}^f and h_{t+1}^b represent the forward and backward hidden states of the BiLSTM at time $t - 1$ and $t + 1$, respectively, and h_t^f and h_t^b represent the forward and backward hidden states of the BiLSTM at time t . The function LSTM represents the LSTM network, which processes the input data x_t and hidden state h_{t-1}^f or h_{t+1}^b to produce the hidden state h_t^f or h_t^b . The function CNN represents the CNN network, which processes the concatenated hidden states $[h_t^f, h_t^b]$ to produce the output y_t .

3.3.5. CNN-BiLSTM with Conditional Random Field

Conditional Random Field (CRF) is a type of probabilistic graphical model that can be used to model the relationships between different variables in a dataset. EEG data contains complicated temporal correlations between the data and labels as a result of fluctuating emotional states in response to price changes and losing or winning a transaction, therefore, establishing the appropriate label sequence can be a challenging task. The addition of a CRF layer facilitates the modeling of these dependencies and thus helps to improve the accuracy of predictions in a CNN-BiLSTM model for the classification of emotions from the EEG data. The CRF layer [40] can be defined as follows:

Let $y = (y_1, \dots, y_T)$ be the sequence of labels and $x = (x_1, \dots, x_T)$ be the sequence of input features. The CRF layer computes the conditional probability of the label sequence given the input features as follows:

$$p(y | x) = \frac{1}{Z(x)} \prod_{t=1}^T \psi_t(y_t, y_{t-1}, x) \quad (13)$$

where $Z(x)$ is the normalization factor, also known as the partition function, and ψ_t is the transition score from label y_t to label y_{t-1} at time step t .

The transition scores are defined as:

$$\psi_t(y_t, y_{t-1}, x) = \exp\left(\sum_{k=1}^K \text{tr}(k, y_t, y_{t-1}) \cdot \text{emit}(k, y_t, x_t)\right) \quad (14)$$

where K is the number of classes, $\text{tr}(k, y_t, y_{t-1})$ is the transition weight from class k at time step $t - 1$ to class k at time step t , and $\text{emit}(k, y_t, x_t)$ is the emission weight from class k at time step t to the input features x_t .

3.3.6. Hyperparameter Optimization

We applied the Bayesian Optimization [41] approach to find the best possible values for the hyperparameters of the specified models. Bayesian Optimization is well-suited for tackling black-box optimizations and noisy functional evaluations, and it is more efficient than manual network-tuning or traditional approaches such as the Grid-Search

and Randomized-Search in high-dimensional datasets. Table A1 in Appendix A shows the search space employed by the Bayesian Optimization technique for all classifiers. The optimization process considers the learning rate, dropout rate, count of hidden layers, neuron count per layer, batch size, activation function, optimizer, number of CNN filters, kernel size, pooling size, and type, among other hyperparameters.

In this paper, we used the same search space configuration for both experiments. To optimize the hyperparameters, we utilized the Adaptive Experimentation (AE) library [42] in Python programming language. AE is a widely adopted multi-objective optimization framework that has been employed effectively by Facebook in numerous real-world online experiments.

3.4. Performance Metrics and Experimental Software

In this research, accuracy, precision, recall (sensitivity), specificity, and negative predicted values from the confusion matrix are used as the performance metrics for classification algorithms.

Accuracy is the ratio of the number of true positives and true negatives to the total number of predictions made by the model. It is given by the below expression:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (15)$$

Precision refers to the proportion of correctly classified positive instances classified as positive by the classifier. Mathematically, it is given by:

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FN}) \quad (16)$$

Recall (Sensitivity, True Positive Rate) is the proportion of positive instances that were correctly classified as positive by the classifier. The recall is given by:

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \quad (17)$$

The NPV (Negative Predicted Value) measures the proportion of negative instances that were correctly classified as negative by the classifier. NPV is expressed as follows:

$$\text{NPV} = \text{TN}/(\text{TN} + \text{FN}) \quad (18)$$

The F1-Measure represents the harmonic mean of precision and recall. F1-Measure provides a comprehensive evaluation of the classifier's performance considering both precision and recall. It is given by:

$$\text{F1 - Measure} = 2 \times (\text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall}) \quad (19)$$

where TP = True Positives (number of instances correctly classified as positive), TN = True Negatives (number of instances correctly classified as negative), FP = False Positives (number of instances incorrectly classified as positive), and FN = False Negatives (number of instances incorrectly classified as negative).

4. Experiments

The proposed emotion classification framework compares participants' emotional state when trading with predetermined risk locking to their emotional state when trading without risk locking. Each experiment was simulated using live Bitcoin market data with 1-min interval candlestick charts for a duration of 30 min. In the experiments, participants were given a starting balance of USD 10,000 for use in simulated trades. The initial trading amount was uniform across all participants and was not adjusted based on individual characteristics or prior trading experience. The initial amount was set at a relatively large sum in order to give participants the opportunity to experience a wide range of potential

outcomes and to allow for the analysis of their emotional responses under different trading actions (buy, hold, and sell).

In the first experiment, participants were given a set of predetermined risk-locking measures to use while making simulated trades. These measures included a pre-defined stop loss of 1% of the traded value and a 1% purchase limit order of the traded value, which were set based on the anticipated 30-min price volatility. A stop loss is a risk management tool that automatically closes a trade when the loss reaches a predetermined price threshold, helping to prevent the trader from incurring further losses. A buy limit order is an order to buy a security at a specified price or lower. It is automatically executed when the price reaches the upper limit of the specified threshold. These risk-locking measures were intended to help participants manage their potential losses and make informed trading decisions.

In contrast, the second experiment did not have any predetermined buy limit orders or stop loss measures to lock in risks. Participants were given the freedom to make trading decisions based on their own assessment of the market and to profit from fluctuations in Bitcoin's price by buying, holding, or selling their holdings.

A control group was employed in this study to establish a baseline for comparison with the trading groups. The control group refrained from any trading activity and was not provided with the options of buy, sell, or hold actions. Instead, the affective states of the control group were derived solely from observing the market data. We aimed to discern any significant distinctions in emotional responses that could be attributed to trading activity by comparing the affective states of the control group with those of the trading groups.

The subsequent section describes how each electrode's raw EEG data was converted into brainwaves (delta, theta, alpha, beta, and gamma frequencies) which were then categorized into different emotional states using machine learning algorithms.

All experiments were executed on a Windows 10 operating system, equipped with an Intel Core i5 processor with 8 cores operating at a frequency of 2.30 GHz, 20 GB of RAM, and an NVIDIA GeForce GTX 1050 graphics processing unit. The data preprocessing and deep learning techniques were implemented utilizing Python 3.7.3 (64-bit), TensorFlow 2.6, and CUDNN 10.1.105. The EEG analysis was performed utilizing the MNE library 1.3.0 (<https://mne.tools>, accessed on 23 December 2022) and Matlab's EEGLab toolbox.

5. Results and Analysis

5.1. Network Architecture Suggested by Bayesian Optimization

Table 2 presents the suggested network architecture for both Experiment 1 and Experiment 2 based on the hyperparameter values obtained using Bayesian Optimization. We utilized these hyperparameter values for the deep learning models used for emotion classification. The CNN portion of the models used 128 CNN filters with a kernel size of 3, a pooling layer with a size of 2, and a max pooling type as obtained from Bayesian Optimization.

Table 2. The suggested network architecture based on the hyperparameter values obtained using Bayesian Optimization for both experiments.

Experiment	Model	Dropout Rate	Learning Rate	Hidden Layers	Neurons per Layer	Activation Function	Batch Size	Optimizer
Experiment 1	LSTM	0.022	0.0001	3	461	linear	32	adam
Experiment 1	CNN	0.262	0.001	3	463	relu	8	adamx
Experiment 1	CNN-LSTM	0.045	0.0010	2	450	relu	8	adam
Experiment 1	CNN-BiLSTM	0.220	0.0310	2	515	relu	32	adam
Experiment 1	CNN-BiLSTM-CRF	0.210	0.0128	3	535	relu	64	adam
Experiment 2	LSTM	0.025	0.0105	3	444	linear	64	adamx
Experiment 2	CNN	0.202	0.0610	3	401	relu	4	adam
Experiment 2	CNN-LSTM	0.039	0.0120	2	494	relu	8	adamx
Experiment 2	CNN-BiLSTM	0.210	0.0262	2	463	relu	8	adam
Experiment 2	CNN-BiLSTM-CRF	0.097	0.0121	3	470	relu	32	adamx

5.2. Distribution of Valence-Arousal for Experimental Groups

Figure 7 presents the distribution of valence-arousal for both experimental groups. In both experiments, the most dominant emotion is High Valence High Arousal (HVHA), followed by Low Valence Low Arousal (LVLA), Low Valence High Arousal (LVHA), and High Valence Low Arousal (HVLA). Upon further analysis, it can be observed that while the overall emotion experienced by participants in both experiments is similar, there is a notable difference in the distribution of emotions experienced under different trading strategies. Specifically, Experiment 1, which employed predetermined stop loss and limit orders to lock in profit and risk objectives, exhibited a higher frequency of HVHA and LVLA compared to Experiment 2, which did not utilize such risk-locking techniques.

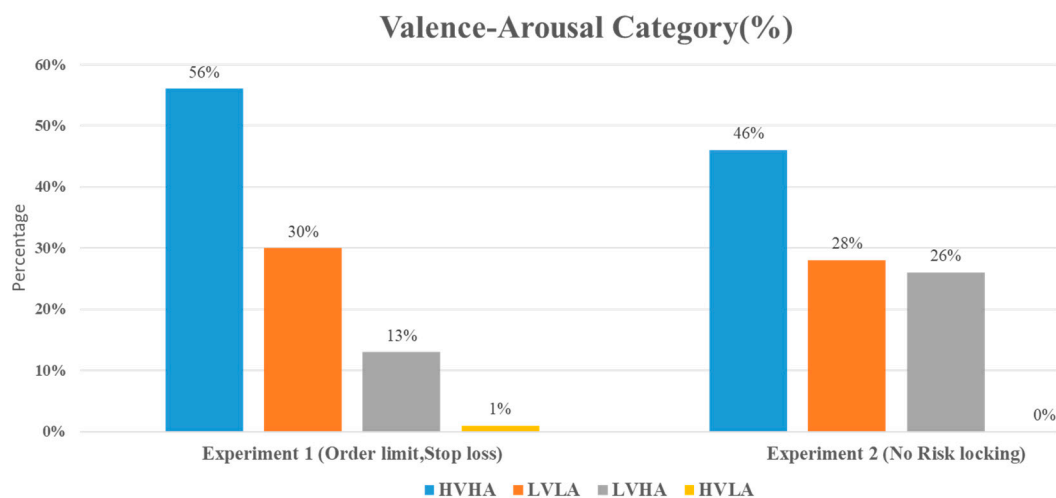


Figure 7. The percentage of valence-arousal category distribution obtained for both experiments conducted with and without risk locking via limit order and stop loss. The second group, which did not use stop loss and limit orders to lock in the risks and book the profits, demonstrated twice (26%) as much Low Valence and High Arousal (the emotion of fear) as the first group (13%), which did utilize these measures.

5.3. Effect of Stop Loss and Limit Orders on Participant Emotions

The results of Experiment 1 showed that the use of stop loss and limit orders on trades was associated with a higher High Valence High Arousal (HVHA) emotion (56%) compared to the group in Experiment 2, which did not use these risk-locking criteria (46%). In other words, the test group that utilized stop loss and limit orders was 17.85% more hopeful than the group that did not use these criteria. The frequency of Low Valence Low Arousal (LVLA) emotions was similar between the two groups, although there was a marginally higher frequency in Experiment 1.

One of the most notable findings of this study was that the use of stop loss and limit orders were associated with a lower frequency of Low Valence High Arousal (LVHA, or fear) emotions. Specifically, participants in Experiment 1 who employed these risk-reward control measures reported LVHA emotions of 13%, compared to 26% in the group that traded without these measures in Experiment 2. This difference can be attributed to the ability of stop loss and rules to minimize the impact of volatility and significant losses [22]. Consequently, traders experience less fear when entering or exiting positions due to these risk control methods.

Another notable finding was the absence of High Valence Low Arousal (HVLA, or calmness) emotions in the second group. This may be expected, as it is rare for traders to exhibit calmness during highly volatile short-term trading sessions. In comparison, the frequency of HVLA emotions was low but detectable in the first group. Overall, these results suggest that the use of stop loss and limit orders is associated with higher frequencies

of HVHA and lower frequencies of LVHA emotions, indicating that managing risks in this way may lead to more hopeful and less fearful traders.

5.4. Analysis of Classification Accuracies in Experiments 1 and 2

The classification accuracies for each participant in Experiment 1 and Experiment 2 are presented in Tables 3 and 4, respectively. The classification accuracy of each participant is dependent on variances in their physiological features and the intensity or complexity of their conveyed emotions. The results of the ablation study demonstrate the mean classification accuracy for each participant using a 10-fold cross-validation technique for both Experiments 1 and 2. Results from Experiment 1 show that the CNN-BiLSTM-CRF model achieved the highest mean accuracy of 85.65%, followed by the CNN-BiLSTM, CNN-LSTM, and CNN models. In Experiment 2, the highest mean accuracy of 85.05% was also obtained by the CNN-BiLSTM-CRF model, with similar performance for the CNN-BiLSTM, CNN-LSTM, and CNN models. The results of this study suggest that the incorporation of a CRF layer in the CNN-BiLSTM architecture improves the overall accuracy of emotion classification.

Table 3. Mean classification accuracy with a 10-fold cross-validation technique for each participant in Experiment 1. Ten test runs were performed on each model, and the average of the results was calculated.

Participant	LSTM	CNN	CNN-LSTM	CNN-BiLSTM	CNN-BiLSTM-CRF
1	59.06	64.03	66.43	75.72	78.77
2	79.16	84.49	84.56	85.68	89.91
3	61.11	67.58	69.30	68.97	76.90
4	78.83	79.32	80.47	85.92	86.12
5	69.24	72.90	77.66	81.58	85.98
6	84.49	88.03	85.16	86.82	89.79
7	75.01	80.35	85.82	85.59	89.63
8	80.77	85.25	84.47	87.52	89.54
9	74.69	76.38	78.00	79.33	84.31
10	73.03	76.63	76.48	82.69	84.90
11	63.90	70.60	73.53	74.19	79.90
12	62.66	65.54	70.88	70.58	79.88
13	67.52	80.22	81.52	79.04	86.19
14	72.61	75.18	79.73	85.71	88.84
15	83.41	84.46	85.57	87.55	89.71
16	75.10	75.85	81.55	82.53	86.82
17	80.87	82.75	82.98	85.46	89.89
18	77.65	77.80	79.13	75.98	83.89
19	74.90	71.58	76.35	82.35	84.75
20	70.80	65.72	79.36	86.62	87.21
Mean Accuracy	73.24	76.23	78.95	81.49	85.65 **

The numbers in bold represent the mean accuracy; ** denotes the highest accuracy reported.

Furthermore, it can be observed from Figure 8 that there is a clear progression in terms of accuracy as the model complexity increases, with LSTM achieving the lowest accuracy and CNN-BiLSTM-CRF achieving the highest accuracy. The CNN-BiLSTM-CRF model achieved the highest mean accuracy with the least deviation in performance, unlike LSTM and the CNN models which exhibited high variability in their performance. The CNN-BiLSTM-CRF model produced the highest mean accuracy with the least performance variance in contrast to other competitive models. This can be attributed to the use of a CRF layer which helps to regularize the model and reduce overfitting, resulting in more consistent and reliable performance. Additionally, it allows the model to consider the context of previously predicted classes, thus enforcing smoothness constraints on the predicted sequence and aligning predictions from different parts of the sequence. This can help mitigate the effect of outliers in the data.

Table 4. Mean classification accuracy for each participant in Experiment 1 using a 10-fold cross-validation technique. Ten test runs were performed on each model, and the average of the results was calculated.

Participant	LSTM	CNN	CNN-LSTM	CNN-BiLSTM	CNN-BiLSTM-CRF
1	52.42	65.44	62.97	73.09	76.88
2	82.25	84.36	80.92	81.42	90.10
3	57.14	59.63	70.03	71.56	75.62
4	80.25	83.39	84.95	87.43	86.10
5	63.21	68.15	69.67	80.32	83.96
6	79.13	80.88	81.03	86.10	88.80
7	68.52	73.08	81.87	84.28	87.72
8	74.50	84.23	81.20	84.28	87.79
9	69.50	76.64	75.79	78.51	82.92
10	76.59	74.44	81.17	84.26	88.73
11	61.73	65.19	72.65	74.53	81.87
12	59.84	66.78	72.61	73.01	80.49
13	65.39	75.70	82.89	79.36	82.88
14	75.63	72.25	73.62	77.38	84.19
15	78.53	78.82	81.83	84.13	89.59
16	76.08	68.74	83.71	78.66	86.84
17	82.35	86.80	83.72	78.65	88.45
18	75.59	73.21	80.88	77.57	83.57
19	79.53	72.31	74.39	82.81	87.95
20	75.02	58.71	81.40	81.35	86.58
Mean Accuracy	71.66	73.44	77.87	79.94	85.05 **

The numbers in bold represent the mean accuracy. ** denotes the highest accuracy reported.

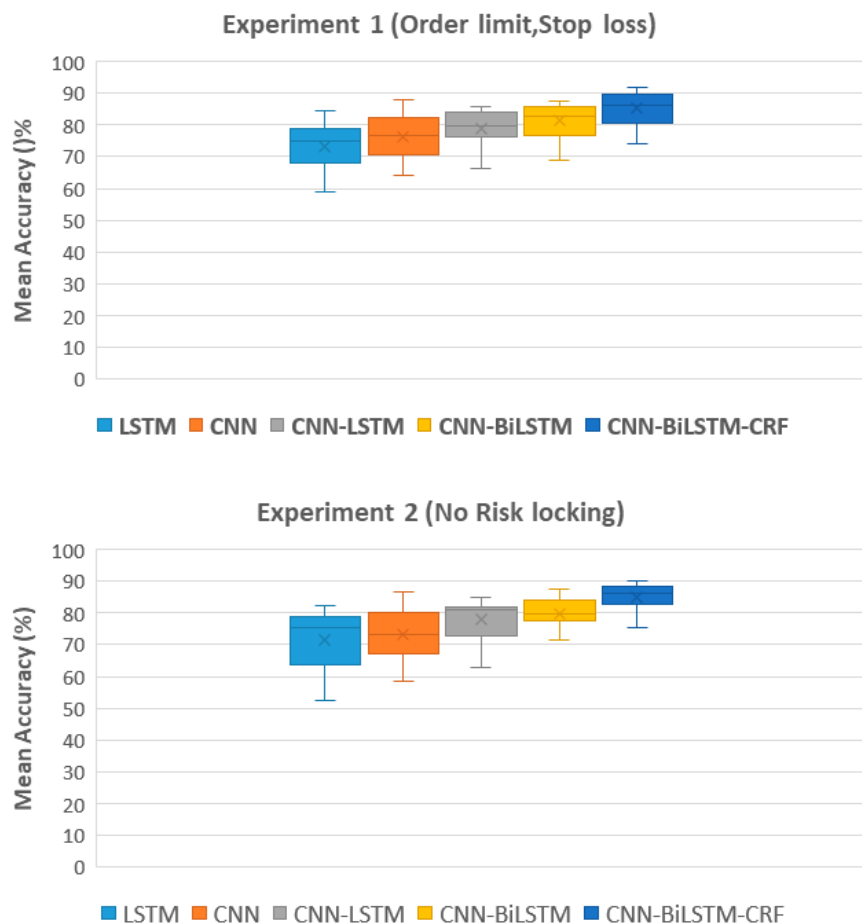


Figure 8. Comparison of the mean classification accuracy across different models for Experiment 1 with risk locking and Experiment 2 without risk locking, as determined by a 10-fold cross-validation technique.

5.5. Influence of Non-Systematic Factors and Subject Level Differences

In this study, an exhaustive methodology was applied to preprocess the EEG data for all participants. However, it is possible that other non-systematic factors may have affected the EEG signals of some participants. Therefore, the reported accuracy differs on an individual basis. Compared to other participants, subjects 1, 3, 11, and 12 reported marginally inferior accuracy in both experiments. Further investigation using spectral analysis of the EEG data for these participants may be necessary to understand this phenomenon.

5.6. Comparison of Precision, Recall, F1-Measure, Specificity, and NPV

Tables 5 and 6 present the mean classification performance metrics of EEG emotion classification models in Experiments 1 and 2. In Experiment 1, the model with the highest performance across all metrics (precision, recall, F1-measure, specificity, and NPV) is CNN-BiLSTM-CRF. Similarly, in Experiment 2, the performance metrics for various models reveal that CNN-BiLSTM-CRF is the most robust model. The results show that the addition of a Conditional Random Field (CRF) layer to the CNN-BiLSTM model improves the overall performance of the model by capturing the dependencies between the output labels and providing more accurate predictions.

Table 5. The mean classification performance metrics with a 10-fold cross-validation technique across all participants in Experiment 1. Ten test runs were performed on each model, and the average of the results was calculated.

Classifier	Accuracy	Precision	Recall	F1-Measure	Specificity	NPV
LSTM	73.24	0.8187	0.7335	0.7651	0.7369	0.7249
CNN	76.23	0.8380	0.7657	0.7912	0.7818	0.7760
CNN-LSTM	78.95	0.8532	0.7862	0.8138	0.8172	0.8000
CNN-BiLSTM	81.49	0.8708	0.8135	0.8364	0.8353	0.8207
CNN-BiLSTM-CRF	85.65	0.8981	0.8601	0.8725	0.8688	0.8575

Table 6. The mean classification performance metrics with a 10-fold cross-validation technique across all participants in Experiment 2. Ten test runs were performed on each model, and the average of the results was calculated.

Classifier	Accuracy	Precision	Recall	F1-Measure	Specificity	NPV
LSTM	71.66	0.8075	0.7128	0.7514	0.7045	0.7387
CNN	73.44	0.8200	0.7328	0.7673	0.7434	0.7629
CNN-LSTM	77.87	0.8471	0.7832	0.8046	0.7558	0.7830
CNN-BiLSTM	79.94	0.8598	0.7948	0.8228	0.8449	0.8214
CNN-BiLSTM-CRF	85.05	0.8949	0.8434	0.8678	0.8570	0.8531

5.7. ROC Analysis of Cross-Validated Emotion Classification Performance

Figures 9 and 10 depict the mean cross-validated Receiver Operating Characteristic (ROC) curves for the training and test data in Experiments 1 and 2, respectively. The ROC curve helps analyze the performance of binary classifiers and exhibits the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) (FPR). The Area Under Curve (AUC) is used to evaluate the overall performance of a classifier, with a larger AUC value indicating a superior classifier. In Figures 9 and 10, the cross-validated ROC curves of the models are displayed for the training and test data in Experiment 1 and Experiment 2, respectively. A thorough analysis of the figures highlights the robustness of the CNN-BiLSTM-CRF model, as it consistently achieves a significantly higher Area Under Curve (AUC) than the other models across all four classes in both experiments. This result suggests that the proposed model performs better in accurately classifying emotions.

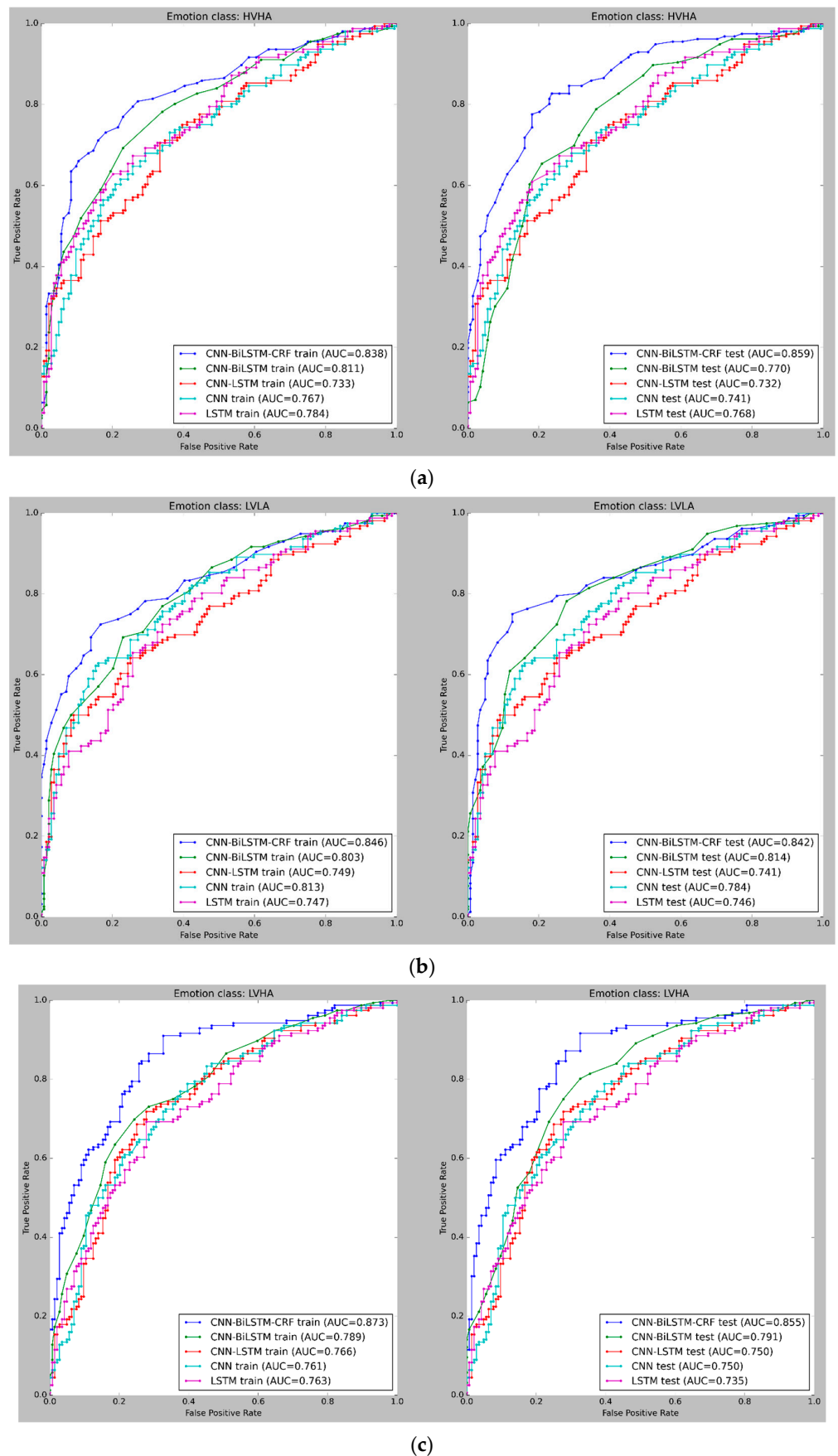
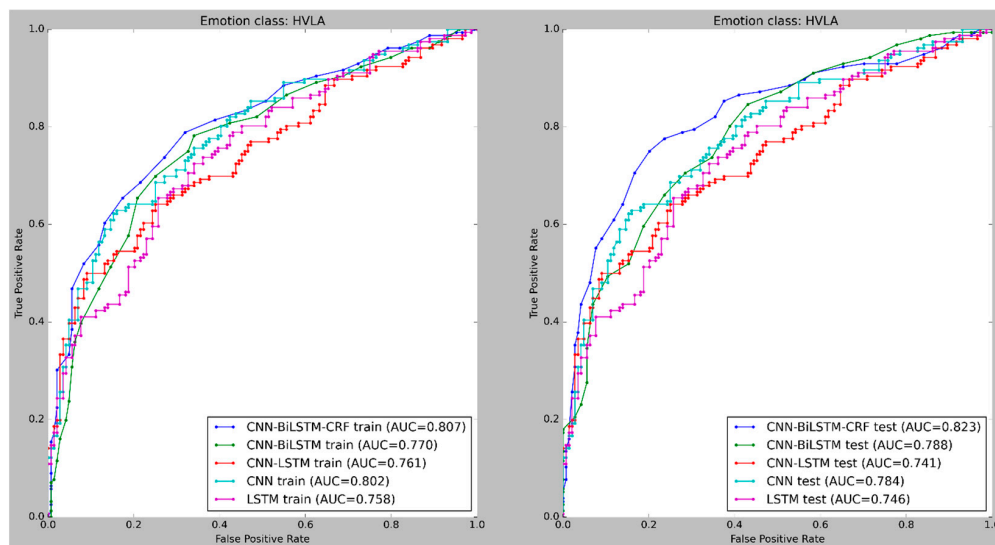
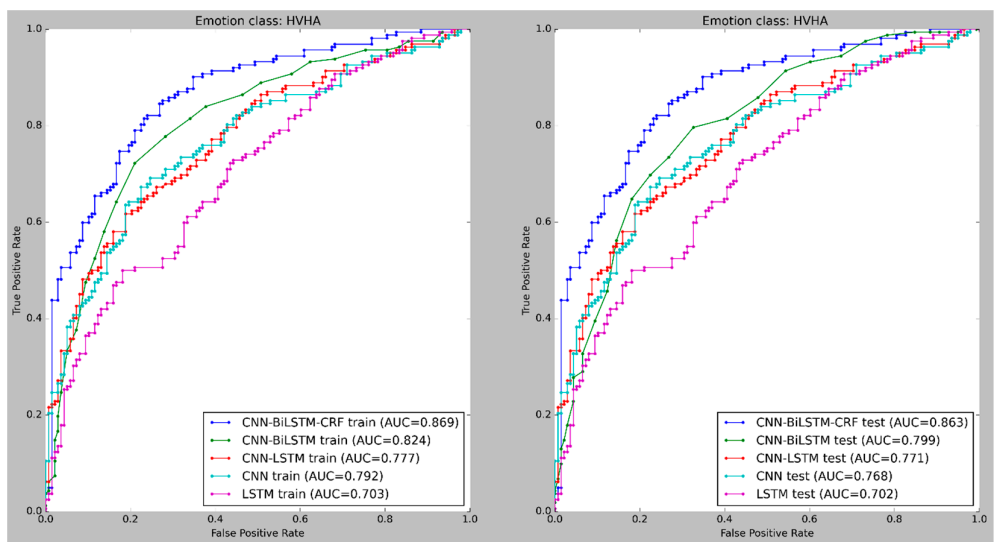


Figure 9. Cont.



(d)

Figure 9. (a) ROC curves for emotion classification in Experiment 1, using cross-validated training and test datasets for High Valence High Arousal (HVHA). The curve shift towards the upper-left quadrant indicates improved model performance. (b) ROC curves for emotion classification in Experiment 1, using cross-validated training and test datasets for Low Valence Low Arousal (LVLA) class. (c) ROC curves for emotion classification in Experiment 1, using cross-validated training and test datasets for Low Valence High Arousal (LVHA) class. (d) ROC curves for emotion classification in Experiment 1, using cross-validated training and test datasets for High Valence Low Arousal (HVLA) class.



(a)

Figure 10. Cont.

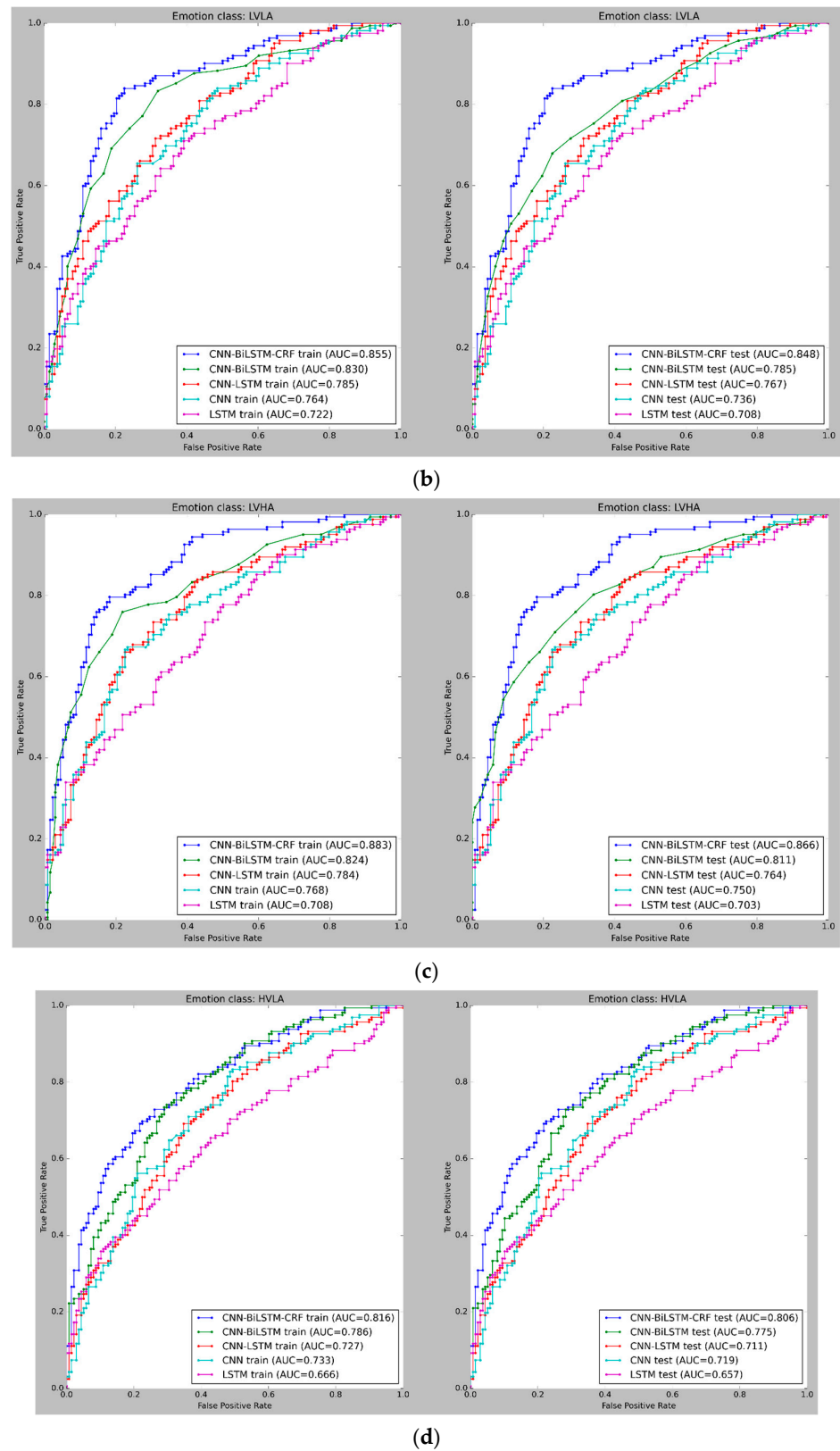


Figure 10. (a–c) ROC curves for emotion classification in Experiment 2, using cross-validated training and test datasets for all four emotion classes: High Valence High Arousal (HVHA), Low Valence Low Arousal (LVLA), and Low Valence High Arousal (LVHA). (d) ROC curves for emotion classification in Experiment 2, using cross-validated training and test datasets for all four emotion classes: High Valence Low Arousal (LVLA).

5.8. Contribution of EEG Features in the Proposed Model

In this section, we discuss the results of feature importance obtained for Experiment 1, Experiment 2, and the control group which performed no trading activity.

The results of the deepSHAP algorithm's feature importance analysis in Experiments 1 and 2 are shown in Figure 11a,b. The results in Experiment 1 indicate that the most significant features, in order of importance, were gamma_Af7, gamma_AF8, gamma_TP9, gamma_TP10, beta_AF8, alpha_AF7, delta_TP9, beta_TP10, and alpha_AF8. Out of the total 20 features present in the EEG dataset, the dominance of gamma-related features can be linked to their association with problem-solving, concentration, and positive valence, as well as the increased arousal that typically results from high-intensity feelings dominated by the outcomes of win or loss. Conversely, Theta-related features, which are associated with deep relaxation and REM sleep, were the least significant and can be closely tied to the HVLA emotion class. This is supported by the distributions of valence-arousal shown in Figure 7, which suggest that, despite having trading experience, the participants in the study did not exhibit a significant amount of HVLA emotion.

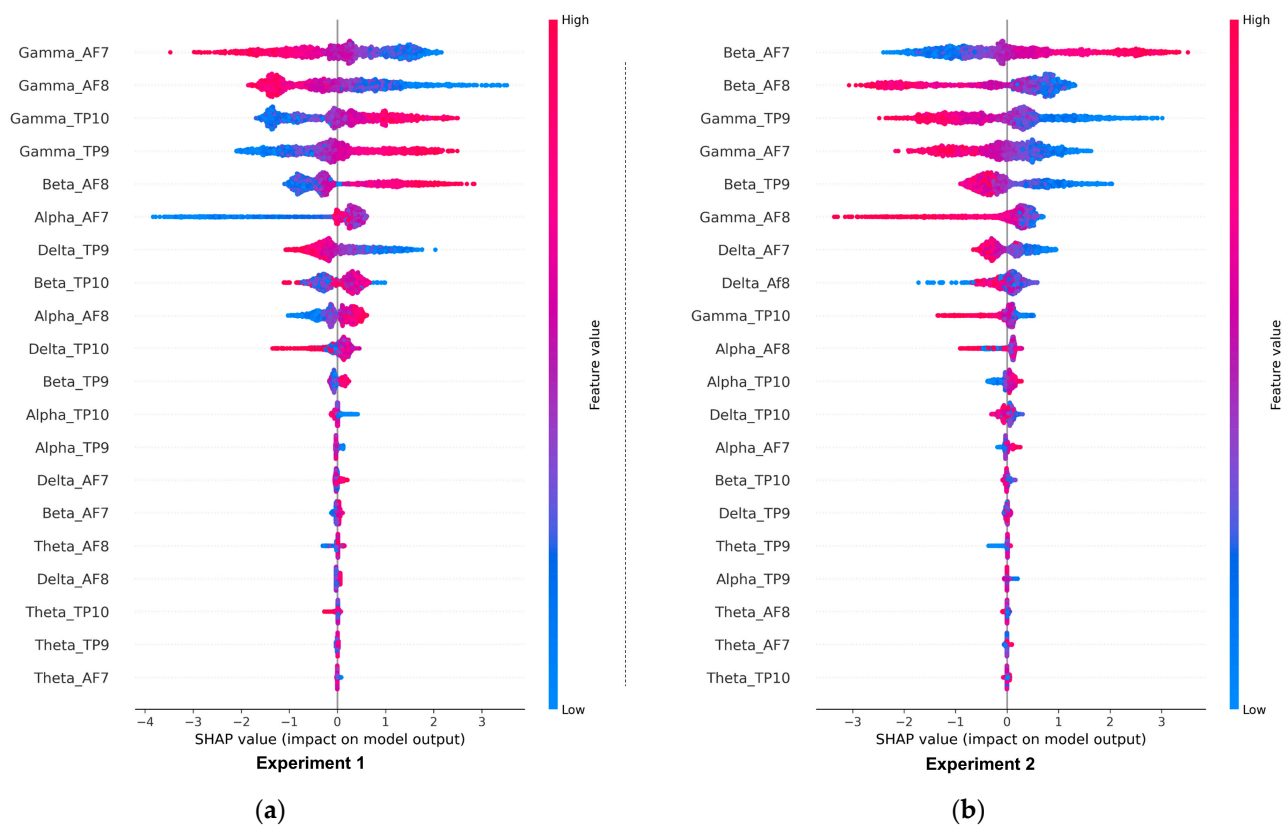


Figure 11. (a,b) Global feature importance of various EEG features in the proposed model (CNN-BiLSTM-CRF) for Experiment 1 and Experiment 2, respectively.

The feature importance analysis in Experiment 2, shown in Figure 11b, shows that beta-related features are the most dominant followed by gamma and alpha-related features, while theta and delta-related features are the least significant. According to Table 1, Beta related features are closely associated with concentration and decision-making. Figure 12a,b shows the proposed model's local feature contributions. The local approach offers a deeper understanding of the relationship between individual EEG features and their impact on class predictions for individual instances. The feature importance for specific instances is consistent with the results of the global method, with gamma and beta-related features remaining the most important features and theta and delta being less significant.

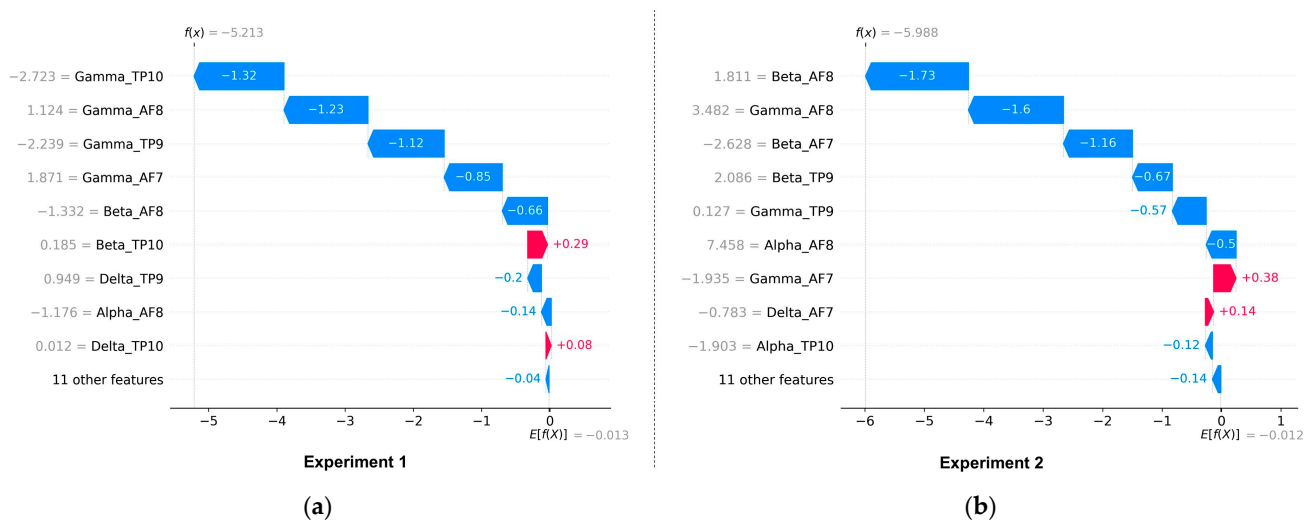


Figure 12. (a,b) Local contribution of various EEG features in the proposed model (CNN-BiLSTM-CRF) for Experiment 1 and Experiment 2, respectively.

The global and local feature importance for the control group is shown in Figure 13a,b. The participants in this group simply observed the market data without performing any actual trading activity. This group serves as a baseline for comparison with Experiment 1 and Experiment 2 which involved active trading activity. The global and local feature importance analysis of the control group shows that the Delta and Theta-related features are the most dominant features along with a minor contribution from Alpha AF8 (global) and Alpha AF7 (local). According to Table 1, Delta and Theta are associated with peaceful states and resting states. This state can be most closely linked to HVLA (calmness) as shown in Figure 2. This contrasts with the dominant features observed in Experiment 1 and Experiment 2 where gamma, beta, and alpha are associated with problem-solving, concentrating, and actively thoughtful states. Moreover, the results indicate that HVLA was negligible in Experiment 1 and absent in Experiment 2, which suggests that even experienced traders find it challenging to remain calm during the event of a win or loss. However, when they are not actively engaged in trading, HVLA becomes dominant. The difference in the feature importance between the control and trading groups can be attributed to the effect of trading activity.

5.9. Comparison with Previous Works

In addition to the ablation experiments, we also compared our research results to those of other closely related studies. These studies employed similar self-labeling and emotion identification methods as our own but used different experimental scenarios for financial trading. Table 7 presents this comparison of the classification performance of the methodology proposed in this paper and earlier EEG-based emotion recognition studies in financial trading. Table 7 shows that the CNN-BiLSTM-CRF classifier achieved a superior accuracy of 85.65% compared to previous works. In the context of similar multi-class emotion categorization studies, it can be inferred that the proposed methodology is both effective and competitive.

Table 7. Comparing our classification performance to previous EEG-based emotion recognition studies in financial trading.

Reference	Classifier	Accuracy (%)
[17]	Deep Learning and Random Forest	83.18
[10]	k-Nearest Neighbors(k-NN)	75.00
Our Proposed Method	CNN-BiLSTM-CRF	85.65

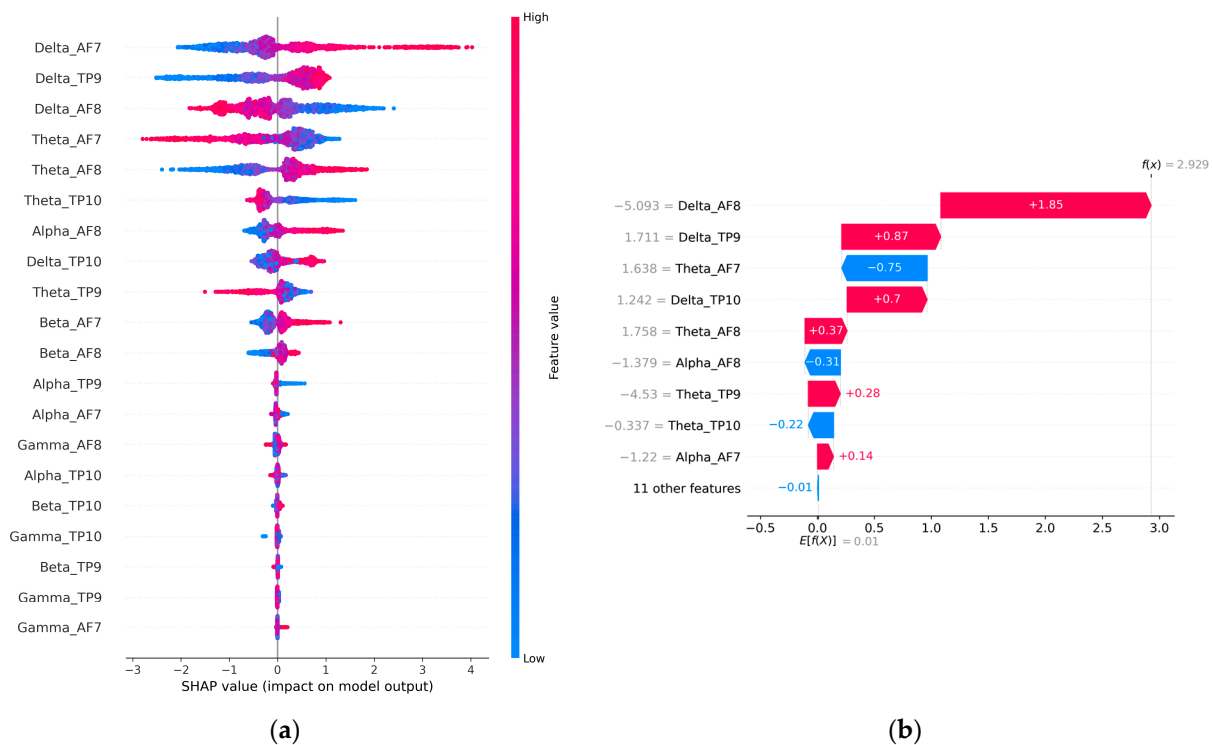


Figure 13. (a,b) Global and local contribution of various EEG features in the proposed model (CNN-BiLSTM-CRF) for the control group.

6. Discussion

The present study aimed to determine the emotional state of traders when utilizing limit orders and stop loss in their trades and compare it to when these risk control measures were not applied. A novel framework for multi-trader emotion classification using EEG signals and deep learning algorithms was proposed and utilized. The study made several important contributions to the field of emotion classification in financial trading. Firstly, a new EEG database was created and the proposed framework was shown to outperform previous studies, achieving state-of-the-art accuracy levels of 85.65% and 85.05% in two experiments. The model was robust with regard to artifacts, with multiple levels of artifact removal implemented, and its performance was evaluated across multiple performance metrics. Furthermore, feature contribution was explained using deepSHAP. The Self-assessment Manikin was used for self-labeling, providing a more reliable and accurate measurement of emotional states. The results demonstrated an association between the application of stop loss and limit orders as risk management measures and a reduction in the frequency of fear (measured by LVHA) among traders. This decrease in fear was linked to the effectiveness of stop loss and limit orders to offset the effects of market volatility, prospective losses, and significant gains. One limitation of this study is the exclusive use of a homogeneous group of experienced participants with prior trading experience. Future research may broaden the EEG data sample by incorporating participants from a broader range of backgrounds, including novice traders.

Table 8 summarizes various studies by presenting the models used, the feature extraction methods employed, the number of participants involved, and the advantages and disadvantages of each study. This comparison highlights each study's strengths and weaknesses and serves as a baseline for the performance of the current study and future works in this area.

Table 8. Comparing the proposed framework with past studies.

Study	Model(s)	Feature Extraction	Participants	Advantages	Disadvantages
[17]	Multi-layer Feedforward ANN and Random Forest	Differential Entropy (DE), Differential Asymmetry (DASM), and Rational Asymmetry (RASM).	10 (5 Males, 5 Females)	High accuracy (82.61%) with real market data and simulated money, feature correlation explained using statistical methods, and self-labeling performed with SAM.	Accuracy was reported as the only metric, with the performance of the deep learning algorithm lower than random forest. Little focus was on artifact removal.
[10]	kNN, Multi-layer Perceptron, and Random Forest	High-Order-Crossing (HOC)	8 (4 Males, 4 Females)	A simulated work environment for a competitive setting.	HVLA class was not considered, small dataset (eight participants), highest accuracy (75%) was lower compared to other relevant works.
[43]	Multi-layer Perceptron	Kernel Density Estimation (KDE)	1 Active investor	Displays emotion distribution in the valence-arousal quadrant and examines the impact of fear on investors' rational decision-making.	Only one participant was considered and did not report any classification metrics obtained from the MLP model; only emotion distributions were reported.
Our work	LSTM, CNN, CNN-LSTM, CNN-BiLSTM, and CNN-BiLSTM-CRF	Independent Component Analysis (ICA)	20 (10 Males, 10 Females)	High accuracy (85.65% and 85.05%) in two experiments, robust to artifacts with multi-level removal, rigorous evaluation procedure, feature contribution explained with deepSHAP, and new findings on LVHA emotions and stop loss/limit orders.	Did not explore the applicability of the proposed architecture to other domains. The work used experienced traders and future research needs to expand the EEG dataset to a more diverse population including beginners.

7. Limitations and Future Work Directions

While the findings of this study are promising for modulating the affective states of participants, it is essential to acknowledge and consider the limitations in the context of future research in this area.

Emotion is a complex subconscious process regulated by the amygdala in the brain [44]. This study uses the participants' self-reported feelings as a measure of emotion by employing the digital variant [32] of SAM [45] which is one of the most accepted measurement tools for human emotions. However, self-reported feelings characterized by valence and arousal provide a limited understanding of the complexity of emotions. Therefore, future studies should incorporate additional physiological features such as cardiovascular (heart rate, heart rate variability, and inter-beat intervals), electrodermal, and respiratory data, among others as suggested by [9]. This analysis may aid in further enhancing the understanding of the emotional elicitation process during trading activities.

Secondly, it is important to note that EEG signals are susceptible to various external factors such as noise, incorrect electrode placement, and environmental conditions. In addition, participants' physical and mental states can be influenced by individual biases, personal factors, mood, and past experiences, which can result in variations in their emotional responses. A small change in any of these factors can lead to signal distortion or changes in emotional response. While our study employed rigorous preprocessing methods to mitigate the effects of external factors on EEG signals, it is important to acknowledge that these factors can never be fully eliminated. Thus, future studies may benefit from continued efforts to optimize data collection and preprocessing techniques.

Third, the participant group in this study is a reasonably homogeneous group of traders to replicate everyday work activities in a competitive environment. However, future studies could investigate more heterogeneous participant groups to obtain more generalizable results and broaden the scope of our work.

8. Conclusions

The main objective of this study was to investigate the effect of using stop loss and limit orders on the emotional state of traders during short-term trading sessions. A deep learning-based valence-arousal framework for emotion classification was developed and applied to EEG data collected from 20 participants who participated in two single-trial experiments. The first group of participants was required to trade with stop loss and limit orders, while the second group was instructed to trade without using these risk control measures. The participants self-labeled their emotional states using a Self-assessment Manikin.

We proposed a hybrid neural architecture combining a Conditional Random Field layer with a CNN-BiLSTM model and used Bayesian Optimization to determine the optimal hyperparameters for multi-trader emotion classification. The results showed that the CNN-BiLSTM-CRF model achieved the highest mean accuracy and least deviation in performance. The model outperformed the previous literature in EEG-based multi-class emotion recognition. Our findings revealed that the group trading with stop loss and limit orders exhibited a greater HVHA (more hope) and a lower LVHA (less fear). In addition, HVLA (emotion of calmness) was found to be insignificant regardless of whether traders applied risk control via limit orders and stop loss or not.

The findings of this study have substantial implications for day traders and portfolio managers. The results of this research can help traders better understand their emotional state and develop positive decision-making skills and the EEG dataset generated can be used for future behavioral finance studies focusing on financial trading. In future research, we plan to expand our EEG dataset through additional testing on a larger sample size and more diverse population in order to gain a deeper understanding of the emotional states in financial trading scenarios. We also intend to explore the applicability of the proposed emotional classification architecture to other domains.

Author Contributions: Conceptualization, B.T. and R.K.S.; methodology, B.T. and R.K.S.; software, B.T.; validation, B.T. and R.K.S.; writing—original draft preparation, B.T.; writing—review and editing, B.T. and R.K.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the ARTHKALP Research Ethical Committee (IRB No. 23112022-10) of ARTHKALP Research Foundation (protocol code 23112022-10 and 23 November 2022). We acknowledge all participants and ARTHKALP Research for their contributions.

Informed Consent Statement: Informed consent was obtained from all participants involved in the study.

Data Availability Statement: Data available upon request.

Acknowledgments: The authors thank the anonymous editors and reviewers for their feedback to enhance the paper's quality.

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

Appendix A

Table A1. Search space configuration for hyperparameter optimization for both experiments.

Hyperparameter Name	Type	Bounds	Values	Description
Learning rate	range	{0.001, 0.01}	-	The learning rate for network weights
Dropout rate	range	{0.001, 0.25}	-	The use of a regularization dropout to ignore a subset of neurons
Count of hidden layers	range	{1, 5}	-	Number of hidden layers
Neuron count per layer	range	{350, 550}	-	Neurons in each layer
Batch size	choice	-	{8, 16, 32, 64}	Batch Size for a single pass
Activation function	choice	-	{linear, relu}	Activation function for the network
Optimizer	choice	-	{rms, adam, adamx, sgd}	Optimizer for reducing the loss function
CNN filters	choice	-	{32, 64, 128, 256}	Number of convolutional filters
Kernel size	choice	-	{2, 3, 4, 5}	Kernel size for convolutional layers
Pooling size and type	choice	-	{2, 3, 4} and {'max', 'average'}	Pooling size and type

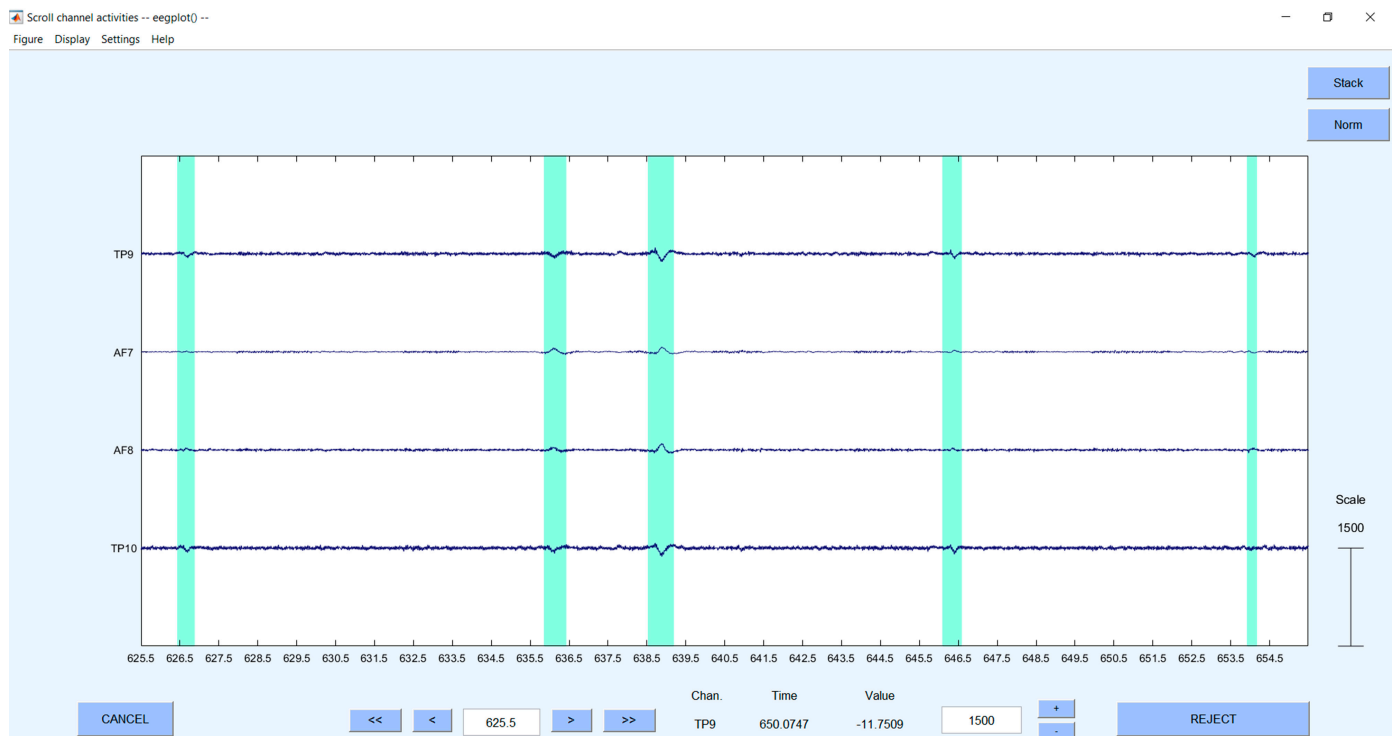


Figure A1. The raw EEG signal of Participant 1 after artifact removal using a sliding window size of 30 seconds. The green-highlighted boundaries indicate spikes caused by artifacts such as eye blinking. The hyphen (-) in the figure serves as a representation of the minus sign (–) in EEGLab.

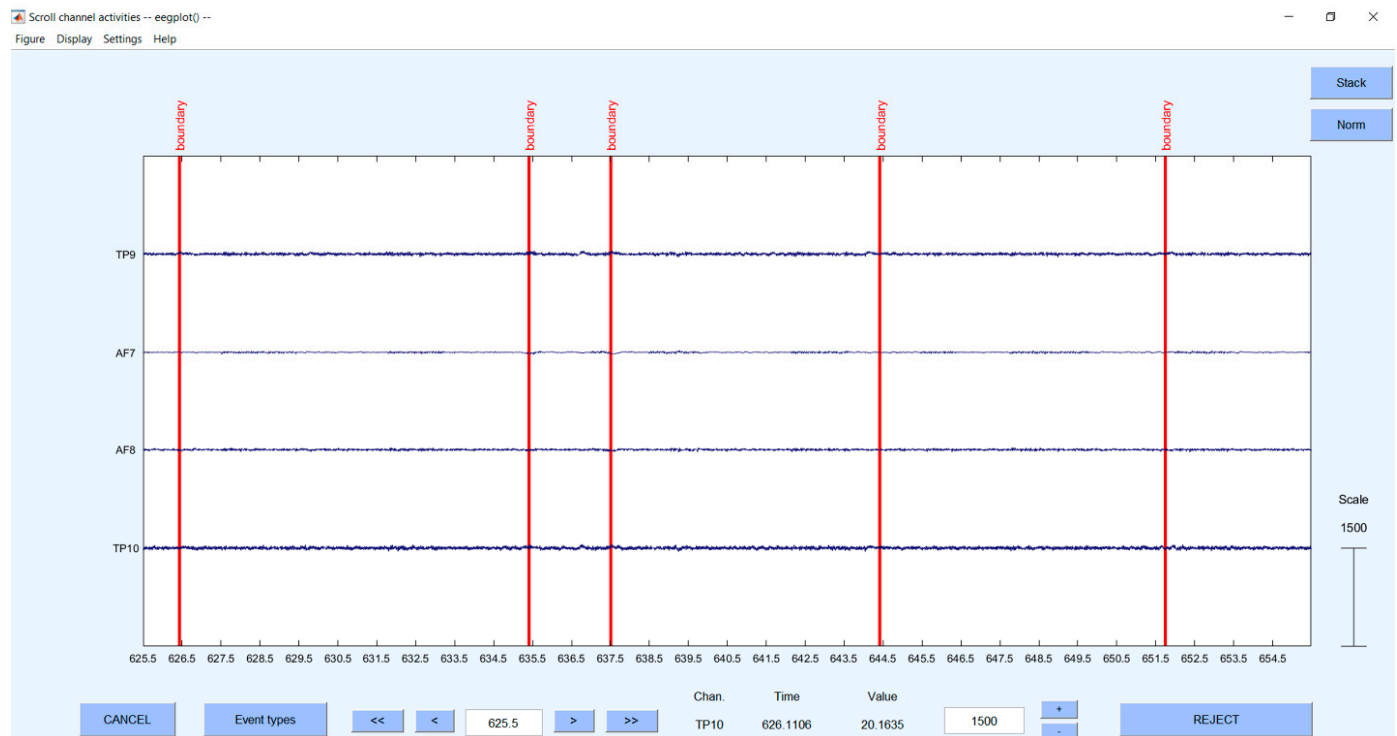


Figure A2. The raw EEG signal of Participant 1 after artifact removal using a sliding window size of 30 seconds. Regarding the boundary markers, it can be observed that the artifact-induced spikes have been eliminated and the signal appears clean.

References

1. Damasio, A.R. Emotion in the perspective of an integrated nervous system. *Brain Res. Rev.* **1998**, *26*, 83–86. [[CrossRef](#)] [[PubMed](#)]
2. Suhaimi, N.S.; Mountstephens, J.; Teo, J. A Dataset for Emotion Recognition Using Virtual Reality and EEG (DER-VREEG): Emotional State Classification Using Low-Cost Wearable VR-EEG Headsets. *Big Data Cogn. Comput.* **2022**, *6*, 16. [[CrossRef](#)]
3. Suhaimi, N.S.; Mountstephens, J.; Teo, J. EEG-Based Emotion Recognition: A State-of-the-Art Review of Current Trends and Opportunities. *Comput. Intell. Neurosci.* **2020**, *2020*, 8875426. [[CrossRef](#)]
4. Soleymani, M.; Lichtenauer, J.; Pun, T.; Pantic, M. A Multimodal Database for Affect Recognition and Implicit Tagging. *IEEE Trans. Affect. Comput.* **2012**, *3*, 42–55. [[CrossRef](#)]
5. Subramanian, R.R.; Sireesha, Y.; Reddy, Y.S.P.K.; Bindamrutha, T.; Harika, M.; Sudharsan, R.R. Audio Emotion Recognition by Deep Neural Networks and Machine Learning Algorithms. In Proceedings of the 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 8–9 October 2021; pp. 1–6.
6. Bharti, S.K.; Varadhaganapathy, S.; Gupta, R.K.; Shukla, P.K.; Bouye, M.; Hingaa, S.K.; Mahmoud, A. Text-Based Emotion Recognition Using Deep Learning Approach. *Comput. Intell. Neurosci.* **2022**, *2022*, 2645381. [[CrossRef](#)]
7. Bonassi, A.; Ghilardi, T.; Gabrieli, G.; Truzzi, A.; Doi, H.; Borelli, J.L.; Lepri, B.; Shinohara, K.; Esposito, G. The recognition of cross-cultural emotional faces is affected by intensity and ethnicity in a Japanese sample. *Behav. Sci.* **2021**, *11*, 59. [[CrossRef](#)]
8. Torres, P.; Edgar, P.; Torres, E.A.; Hernández-Álvarez, M.; Yoo, S.G. EEG-based BCI emotion recognition: A survey. *Sensors* **2020**, *20*, 5083. [[CrossRef](#)] [[PubMed](#)]
9. Shu, L.; Xie, J.; Yang, M.; Li, Z.; Li, Z.; Liao, D.; Xu, X.; Yang, X. A Review of Emotion Recognition Using Physiological Signals. *Sensors* **2018**, *18*, 2074. [[CrossRef](#)]
10. Torres, P.; Edgar, P.; Torres, H.E.; Hernández-Álvarez, M.; Yoo, S.G. EEG-Based BCI Emotion Recognition Using the Stock-Emotion Dataset. In *BT-Advances in Emerging Trends and Technologies*; Botto-Tobar, M., Gómez, O.S., Miranda, R.R., Cadena, A.D., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 226–235.
11. Islam, M.S.; Hussain, I.; Rahman, M.M.; Park, S.J.; Hossain, M.A. Explainable Artificial Intelligence Model for Stroke Prediction Using EEG Signal. *Sensors* **2022**, *22*, 9859. [[CrossRef](#)]
12. Hussain, I.; Park, S.J. HealthSOS: Real-Time Health Monitoring System for Stroke Prognostics. *IEEE Access* **2020**, *8*, 213574–213586. [[CrossRef](#)]
13. Hussain, I.; Park, S.J. Quantitative evaluation of task-induced neurological outcome after stroke. *Brain Sci.* **2021**, *11*, 900. [[CrossRef](#)] [[PubMed](#)]
14. Hussain, I.; Young, S.; Park, S.J. Driving-induced neurological biomarkers in an advanced driver-assistance system. *Sensors* **2021**, *21*, 6985. [[CrossRef](#)] [[PubMed](#)]
15. Hussain, I.; Hossain, M.A.; Jany, R.; Bari, M.A.; Uddin, M.; Kamal, A.R.M.; Ku, Y.; Kim, J.-S. Quantitative Evaluation of EEG-Biomarkers for Prediction of Sleep Stages. *Sensors* **2022**, *22*, 3079. [[CrossRef](#)] [[PubMed](#)]
16. Posner, J.; Russell, J.A.; Peterson, B.S. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* **2005**, *17*, 715–734. [[CrossRef](#)]
17. Torres, E.P.; Torres, E.A.; Hernández-Álvarez, M.; Yoo, S.G. Emotion Recognition Related to Stock Trading Using Machine Learning Algorithms with Feature Selection. *IEEE Access* **2020**, *8*, 199719–199732. [[CrossRef](#)]
18. Lo, A.W.; Repin, D.V.; Steenbarger, B.N. Fear and Greed in Financial Markets: A Clinical Study of Day-Traders. *Am. Econ. Rev.* **2005**, *95*, 352–359. [[CrossRef](#)]
19. Gottman, J.M.; Coan, J.; Carrere, S.; Swanson, C.; Gottman, J.M.; Coan, J.; Carrere, S.; Swanson, C. Predicting Marital Happiness and Stability from Newlywed Interactions. *J. Marriage Fam.* **1998**, *60*, 5–22. [[CrossRef](#)]
20. Shefrin, H.; Statman, M. The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence. *J. Financ.* **1985**, *40*, 777–790. [[CrossRef](#)]
21. Hollifield, B.; Miller, R.A.; Sandås, P. Empirical analysis of limit order markets. *Rev. Econ. Stud.* **2004**, *71*, 1027–1063. [[CrossRef](#)]
22. Klement, J. Assessing Stop-Loss and Re-Entry Strategies. *J. Trading* **2013**, *8*, 44–53. [[CrossRef](#)]
23. Fernández, J.M.; Augusto, J.C.; Seepold, R.; Madrid, N.M. Why Traders Need Ambient Intelligence. In *BT-Ambient Intelligence and Future Trends-International Symposium on Ambient Intelligence (ISAml 2010)*; Augusto, J.C., Corchado, J.M., Novais, P., Analide, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 229–236.
24. Du, R.; Zhu, S.; Ni, H.; Mao, T.; Li, J.; Wei, R. Valence-arousal classification of emotion evoked by Chinese ancient-style music using 1D-CNN-BiLSTM model on EEG signals for college students. *Multimed. Tools Appl.* **2022**. [[CrossRef](#)]
25. Shen, F.; Peng, Y.; Dai, G.; Lu, B.; Kong, W. Coupled Projection Transfer Metric Learning for Cross-Session Emotion Recognition from EEG. *Systems* **2022**, *10*, 47. [[CrossRef](#)]
26. Chen, Z.; Duan, S.; Peng, Y. EEG-Based Emotion Recognition by Retargeted Semi-Supervised Regression with Robust Weights. *Systems* **2022**, *10*, 236. [[CrossRef](#)]
27. Cui, F.; Wang, R.; Ding, W.; Chen, Y.; Huang, L. A Novel DE-CNN-BiLSTM Multi-Fusion Model for EEG Emotion Recognition. *Mathematics* **2022**, *10*, 582. [[CrossRef](#)]
28. Toma, F.M.; Miyakoshi, M. Left frontal eeg power responds to stock price changes in a simulated asset bubble market. *Brain Sci.* **2021**, *11*, 670. [[CrossRef](#)]
29. Vieito, J.P.; da Rocha, A.F.; Rocha, F.T. Brain Activity of the Investor’s Stock Market Financial Decision. *J. Behav. Financ.* **2015**, *16*, 220–230. [[CrossRef](#)]

30. Wedel, M.; Pieters, R.; van der Lans, R. Modeling Eye Movements During Decision Making: A Review. *Psychometrika* **2022**. [[CrossRef](#)] [[PubMed](#)]
31. Bradley, M.M.; Lang, P.J. Measuring emotion: The self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* **1994**, *25*, 49–59. [[CrossRef](#)] [[PubMed](#)]
32. Betella, A.; Verschure, P.F.M.J. The affective slider: A digital self-assessment scale for the measurement of human emotions. *PLoS ONE* **2016**, *11*, e0148037. [[CrossRef](#)]
33. Minguillon, J.; Lopez-Gordo, M.A.; Pelayo, F. Trends in EEG-BCI for daily-life: Requirements for artifact removal. *Biomed. Signal Process. Control* **2017**, *31*, 407–418. [[CrossRef](#)]
34. Plechawska-Wojcik, M.; Kaczorowska, M.; Zapala, D. The Artifact Subspace Reconstruction (ASR) for EEG Signal Correction. In *A Comparative Study BT-Information Systems Architecture and Technology: Proceedings of 39th International Conference on Information Systems Architecture and Technology-ISAT 2018*; Świątek, J., Borzemski, L., Wilimowska, Z., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 125–135.
35. Stone, J. V Independent component analysis: An introduction. *Trends Cogn. Sci.* **2002**, *6*, 59–64. [[CrossRef](#)] [[PubMed](#)]
36. Abhang, P.A.; Gawali, B.W.; Mehrotra, S.C. Chapter 2-Technological Basics of EEG Recording and Operation of Apparatus. In *Introduction to EEG- and Speech-Based Emotion Recognition*; Abhang, P.A., Gawali, B.W., Mehrotra, S.C., Eds.; Academic Press: Cambridge, MA, USA, 2016; pp. 19–50.
37. Patil, A.; Rane, M. Convolutional Neural Networks: An Overview and Its Applications in Pattern Recognition. *Smart Innov. Syst. Technol.* **2021**, *195*, 21–30.
38. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 802–810.
39. Kavianpour, P.; Kavianpour, M.; Jahani, E.; Ramezani, A. A CNN-BiLSTM Model with Attention Mechanism for Earthquake Prediction. *arXiv* **2021**, arXiv:2112.13444.
40. Sutton, C.; McCallum, A. An Introduction to Conditional Random Fields for Relational Learning. *Introd. Stat. Relat. Learn.* **2019**, *4*, 267–373.
41. Snoek, J.; Larochelle, H.; Adams, R.P. Practical Bayesian optimization of machine learning algorithms. *Adv. Neural Inf. Process. Syst.* **2012**, *4*, 2951–2959.
42. Bakshy, E.; Dworkin, L.; Karrer, B.; Kashin, K.; Letham, B.; Murthy, A.; Facebook, S.S. AE: A domain-agnostic platform for adaptive experimentation. *Nips* **2018**, *8*, 1–8.
43. Razi, N.I.M.; Othman, M.; Yaacob, H. EEG-based emotion recognition in the investment activities. In Proceedings of the 6th International Conference on Information and Communication Technology for the Muslim World, ICT4M 2016, Jakarta, Indonesia, 22–24 November 2016; pp. 325–329.
44. Sergerie, K.; Chochol, C.; Armony, J.L. The role of the amygdala in emotional processing: A quantitative meta-analysis of functional neuroimaging studies. *Neurosci. Biobehav. Rev.* **2008**, *32*, 811–830. [[CrossRef](#)] [[PubMed](#)]
45. Bradley, M.; Lang, P.J. Self-Assessment Manikin (SAM). *J. Behav. Ther. Exp. Psychiat.* **1994**, *25*, 49–59. [[CrossRef](#)] [[PubMed](#)]

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.