

Highly Accurate EEG Signal Classification Using Multiple Feature Extraction and LSTM Networks

Deshmukh Deepika^{1,2*} and Gillala Rekha¹

¹Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Aziz Nagar, Hyderabad, Telangana 500075, India

²Department of Computer Science and Engineering, Mahatma Gandhi Institute of Technology, Hyderabad, Telangana 500075, India

*e-mail: deshmukhdeepika@gmail.com

Abstract. Brain-computer interface (BCI) was first only thought of as a control channel for end users with simple disabilities, like those locked in a room. Nevertheless, the spectrum of BCI applications has significantly increased due to the multidisciplinary advancements made over the past ten years. Today's BCI technology can combine this artificial output with muscle-based natural products, directly translating brain impulses into control signals. Therefore, combining several biological signals for in-the-moment communication can benefit a far bigger population than first anticipated. New generations of assistive devices could aid individuals with preserved residual functions. Electroencephalography (EEG) signals can effectively perform BCI with maximum accuracy. This work will implement a new multiple-feature extraction-based BCI with a deep convolutional neural network to progress accuracy and reduce system complexity. The EEG signal is extracted from a database and then preprocessed using de-noising and smoothing techniques. Extracting features from the signal, then concatenating the features that have been extracted. After that, the Long Short-Term Memory (LSTM) network is used to train the signals. A proper preprocessing technique will remove the artefacts from the captured EEG signal. Sensitivity evaluation metrics, such as sensitivity, specificity, accuracy, etc., will be evaluated to validate the training and testing performance. This work achieves 99.35% accuracy, 96.38% sensitivity, 99.18% specificity, and 99.21% precision. By combining a range of feature extraction methods, we have effectively harnessed the complementary strengths of these techniques, resulting in a more holistic representation of the underlying neural activities. © 2024 Journal of Biomedical Photonics & Engineering.

Keywords: brain-computer interfaces; electroencephalography (EEG); deep convolutional neural networks; multiple feature extraction; sensitivity evaluation.

Paper #9029 received 21 Oct 2023; revised manuscript received 22 Jan 2024; accepted for publication 23 Jan 2024; published online 3 Mar 2024. doi: [10.18287/JBPE24.10.010304](https://doi.org/10.18287/JBPE24.10.010304).

1 Introduction

The electroencephalography (EEG) records electrical impulses that exit the brain and makes it possible to retrieve valuable data about how the brain works. EEG-based brain-computer interface (BCI) is employed

in many biological and medical procedures, such as detecting and treating central nervous diseases, assessing mental workload, and diagnosing brain tumors. In particular, BCI can assist in communicating brain instructions for those with motor neuron disease, forebrain stroke, brain trauma injury, palsy,

musculoskeletal problems, or other diseases affecting the control and response system between the brain and muscles and facilitate the authority of prosthetic joints [1].

The procedure provides control settings that may be utilised by healthy or disabled people to construct a new brain-computer connection. Either Event-Related Potentials (ERPs), which are time-locked reactions to an external event, or Event-Related Oscillatory Changes (EROCs), internally caused modulations in the continuing EEG, can typically be utilized as input to a BCI. One particularly well-liked and frequently employed mental technique is the imagination of movements. Electrodes affixed to the top of the head record the brain's electrical activity as EEG signals. The implementation of EEG signal analysis is based on curves, a graphical representation of collected data. The visual evaluation of these curves allows doctors to diagnose neural disorders [2].

Even when dealing with inexperienced doctors, relying solely on visual assessment might prove insufficient. The analysis of EEG signals holds significant importance, requiring algorithms to not only gather information from the recorded EEG signals but also process it effectively to predict and classify various brain states [3]. A notable challenge in EEG data processing lies in identifying epilepsy events. Research has indicated that by employing machine learning techniques to analyze the dynamics of task-related Independent Components (ICs), the performance of BCIs in predicting human patients' cognitive functions can be notably improved. To illustrate, ICs located in the temporoparietal region can aid in comprehending intended movement directions, while those in the posterior brain area can simulate individual performance, and sensorimotor ICs can serve as crucial markers for categorizing EEG patterns [4].

BCIs were once thought to be the only control channel whose intended end-users were people with severe impairments, such as those locked in. However, these initial objectives have been greatly expanded due to the transdisciplinary development made over the past ten years. Today's modern BCI technologies can translate brain signals directly into fresh output and mix this kind of synthetic production with a muscle-based prolactin level. Therefore, integrating numerous biochemical signals for real-time interaction has the opportunity to boost a much wider population than initially thought, from end users with maintained residue left features who can benefit from new assistive devices centuries to healthy individuals who could enhance their musculoskeletal achievement further than their normal capabilities [5].

In earlier investigations, factors like emotions, psychopathology, and focus were predictors of classification and impacted EEG data. In contrast, focus, attention level, relaxation, mood, and hand-grasping imagination are often the EEG signal variables translated in BCI. In most cases, the categorization method in the investigations only used one characteristic variable. In

earlier research, BCI was utilized to operate computer applications, move characters in arcade games, and drive robotic wheelchairs via focus feedback. After feature extraction in pattern recognition, the classification system is used [6]. Several strategies have already been put forth to achieve BCI. The main issues with BCI are memory usage and computational complexity.

With limited and abundant samples of EEG training data focused on motor and mental imagery, Muhammad et al. introduced a novel automated framework based on pretrained convolutional neural networks (CNNs). This framework aims to enhance the robustness of BCI systems. Fifty-four individuals who performed MI on two separate days submitted the results of 21600 trials for the MI task to the database. It is composed of spectral-spatial insight, which embeds this same diversity. To evaluate the robustness of the features in two classes, an observation model based on information theory must first consider the discriminative frequency bands. This model uses spectral-spatial inputs. Spectrum-spatial inputs with the distinct features of brain signal patterns are produced from discriminative frequency bands and then converted into a covariance matrix. Spectral-spatial inputs are independently trained using a CNN during the feature representation phase, and they are then joined using a concatenation fusion technique [7].

Han et al. [8] proposed that compared to single-image EEG- and NIRS-based brain changes, a hybrid EEG/NIRS switch's performance was better. The fictional online research's findings typically showed the same overarching trend as those of the offline study. There was no noticeable difference in performance indicators between the disconnected and pseudo-online analytic methodologies when the quantity of keeping fit data was similar [8]. Zhang et al. [9] proposed a sparse Bayesian method for EEG classification by exploiting Laplace. Within a Bayesian evidence framework, a sparse discriminant vector undergoes hierarchical training with a Laplace prior. Comprehensive comparisons between the SBLaplace algorithm and various competing techniques are conducted using two EEG datasets [9].

Gupta et al. [10] proposed a unique feature-creation method for categorizing multi-mental tasks. The suggested approach extracts characteristics in two stages. The EEG signal is first broken down using the wavelet transform. Each feature component acquired in the second step is compactly expressed using eight parameters. The Optimal Decision Tree-Based Support Vector Machine (ODT-SVM) classifier is utilized for categorizing several mental tasks [10].

The motivation of this work is to enhance the accuracy and effectiveness of EEG signal classification, which can lead to a better understanding and utilization of brain signals for various applications, ranging from medical diagnoses to brain-controlled technologies.

The main objectives of the paper are

- to investigate brain activity throughout sensory perception and hand movement phases, ensuring EEG data matches neurological function.

- to enhance the classification performance of the BCI process for several classes while dealing with enormous datasets.

2 Proposed Method

The process begins by obtaining EEG signals from a database, which are subsequently subjected to preprocessing involving techniques like noise reduction and signal smoothing. Once the signals are refined, meaningful features are extracted from them. These extracted features are then combined by concatenation, forming comprehensive feature vectors. Subsequently, a CNN is employed to train the processed EEG signals. This involves feeding the concatenated feature vectors into the LSTM architecture, where convolutional layers identify intricate patterns, pooling layers aid in downsampling and feature selection, and fully connected layers make final predictions or classifications. The network is trained using labeled data, with optimization techniques refining its weights and parameters to minimize the difference between predicted outcomes and actual labels. This comprehensive process amalgamates signal preprocessing, feature extraction, concatenation, and LSTM-based training for a range of applications like medical diagnosis and cognitive analysis. Fig. 1 shows the block diagram of the proposed method.

2.1 Dataset

This dataset consists of a comprehensive collection of over 1500 EEG recordings, ranging from one to two minutes in length, and involving 109 individuals as participants. The EEG data was recorded using a 64-channel EEG setup facilitated by the BCI2000 system (<https://archive.physionet.org/pn4/eegmmidb/>).

During the recording sessions, the participants were involved in a variety of tasks. Each participant completed a total of 14 testing rounds, which included two initial one-minute baseline runs and three successive runs for

each of the four distinct tasks. The dataset is fully explained in Table 1.

2.2 Signal Preprocessing

Signal averaging is a time-domain signal processing method used to boost a signal's intensity compared to noise obscuring it [11]. The signal-to-noise ratio (SNR), ideally proportional to the square root of the number of measurements, will be improved by averaging a set of replicate measurements [12].

In order to smooth out signals using the time-fractional diffusion equation, it is necessary to search for a solution to Eq. (1) that is able to fulfil the starting requirements that have been specified. In other words, $u(x, 0)$ is a noisy signal, and $u(x, T)$ is the smoothed signal.

$$\begin{cases} \frac{\partial^\alpha u(x, t)}{\partial t^\alpha} = g(x, t, u) \frac{\partial^2 u(x, t)}{\partial x^2}, \alpha \in (0, 1), 0 < x < L, 0 < t < T, \\ u(x, 0) = f(x), 0 < x < L, \\ u(0, t) = u(L, t) = 0, 0 < t < T. \end{cases} \quad (1)$$

where $g(x, t) = \exp[-(|\frac{d^2 u(x, t)}{\lambda dx^2}|)^2]$ is a diffusion function. In general, $g(x, t)$ is a smooth decreasing function with $g(0, t) = 1$, $g(x, t) \geq 0$ and $g(x, t)$.

Normalization is the alteration of the indicator to a scale relative to a known and repeatable worth. The values in all four columns of the datasets span from positive immensity to negative immensity [13]. Min-max normalization is used in this study to restrict the values to those between 0 and 1 [14]. The linear transformation of the characteristics is maintained by min-max normalization. The Eq. (2) represents the min-max normalization. The $\min_i$ is the attribute's smallest value. The $\max_i$ is its maximum value.

$$N = \frac{n_i - \min_i}{\max_i - \min_i}. \quad (2)$$

Table 1 Description of the dataset.

Sl. No	Task	Description
1	B-1	Eyes open
2	B-2	Eyes closed
3	Task 1	Make a left or right fist, and then shut it
4	Task 2	Imagine making a fist with your left hand and then closing it as well
5	Task 3	The two fists or both feet may be opened and closed
6	Task 4	Imagine opening and closing both fists or both feet
7	Task 1	Perform alternating open and close movements with either the left or right fist
8	Task 2	Mentally visualize the act of opening and closing either the left or right fist
9	Task 3	Execute alternating open and close movements with both fists or both feet
10	Task 4	Visualise opening and closing both fists or feet

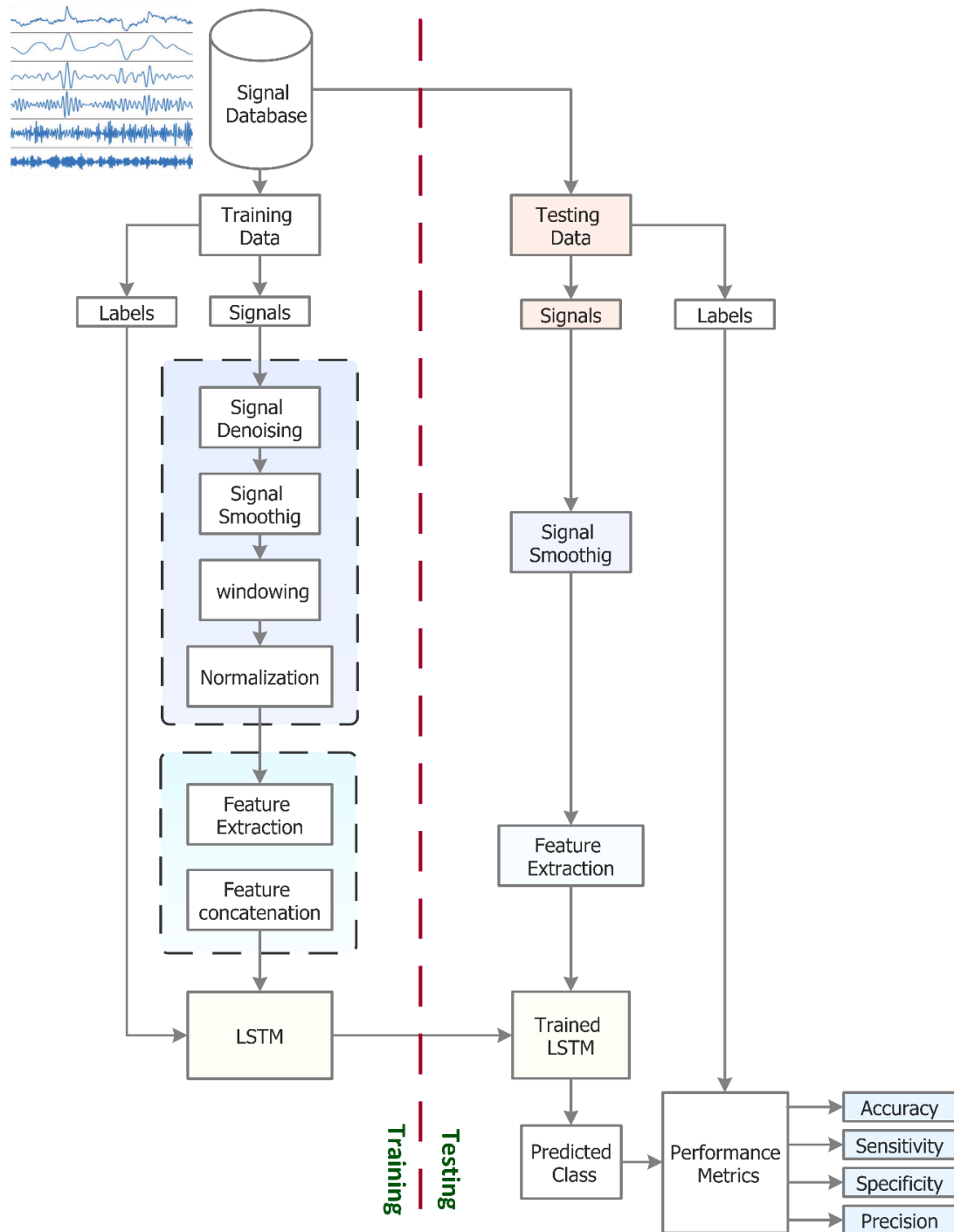


Fig. 1 Block diagram of the proposed method.

2.3 Feature Extraction

Mean

The mean value is the sum divided by the signal values. A relative measure of signal intensity could be how bright the signal appears compared to another [15].

$$\bar{x} = \frac{1}{n} (\sum_{i=1}^n x_i) = \frac{x_1 + x_2 + \dots + x_n}{n}. \quad (3)$$

Standard Deviation (SD)

The degree of dispersion of the data from the mean is expressed as a standard deviation. Using the following Eq. (3), the EEG signal's standard deviation can be determined as following [16]:

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}. \quad (4)$$

Kurtosis

Kurtosis data sets with a high skewness value are more likely to have large tails or severe outliers [17]. Exceptions are usually absent, or the tails are light in data sets with moderate skewness values (Ieracitano).

$$\text{Kurtosis}[X] = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{E[(X-\mu)^4]}{(E[(X-\mu)^2])^2} = \frac{\mu_4}{\sigma^4}. \quad (5)$$

Skewness

A measure reflecting a distribution's imbalance is skewness [18]. A distribution is said to be asymmetric when its left and right sides are devoid of signal [19].

$$\tilde{\mu}_3 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3} = \frac{E[(X-\mu)^3]}{(E[(X-\mu)^2])^{3/2}} = \frac{k_3}{k_2^{3/2}}. \quad (6)$$

Moment

The Moment invariant feature extraction approach is applied to extract the global features for form recognition and identification analysis [20]. Since its introduction, several varieties of the moment invariants approach have existed [21].

$$\mu_n = \int_{-\infty}^{\infty} (x - c)^n f(x) dx. \quad (7)$$

Energy

The energy refers to the likelihood (normalized histogram) of the brightness at the point [22]:

$$EG = \sum_i^{N_G} \sum_j^{N_G} \{p(i, j)\}^2. \quad (8)$$

Entropy (EN)

The Entropy (EN), a metric of unpredictableness, has the lowest value for an erratic matrix [23]. The following Eq. (9) defines this coefficient:

$$EN = -\sum_i^{N_G} \sum_j^{N_G} p(i, j) \log(p(i, j)). \quad (9)$$

Inertia (IN)

The Inertia (IN) also called the Contrast feature, measures the signal's contrast or any local fluctuations to show its quality [24]. This parameter is specified by the Eq. (10) below:

$$IN = -\sum_i^{N_G} \sum_j^{N_G} (i - j)^2 p(i, j). \quad (10)$$

Correlation (CO)

The term correlation (CO) adjective assesses the linear relationships between the co-occurrence matrix's rows and columns. This parameter's definition is given by the Eq. (11) below:

$$CO = \frac{\sum_i^{N_G} \sum_j^{N_G} (i - \mu_x)(j - \mu_y) p(i, j)}{\sigma_x \sigma_y}, \quad (11)$$

where μ_x, μ_y, σ_x , and σ_y are the means and standard deviations of $p_x p_y$.

Inverse Difference Moment (IDM)

The intensity of the spread of the items in the co-occurrence matrix and the components in its diagonal is measured by homogeneity, commonly known as the inverted differential second.

$$idm = \sum_i^{N_G} \sum_j^{N_G} \frac{1}{1 + (i - j)^2} p(i, j). \quad (12)$$

Difference Entropy (DE)

Discrete entropy, used in signal processing, measures the number of bits needed to encode signal data. The signal will become more precise with the greater magnitude of entropy.

$$DE = -\sum_{k=0}^{N_G-1} P_{x-y}(k) \log_2 P_{x-y}(k). \quad (13)$$

2.4 Frequency Features

The Short-Time Fourier Transform (STFT), also referred to as the short-term Fourier transform, is a versatile and powerful tool in audio signal processing. It provides a time-frequency distribution that represents complex amplitude as a function of both time and frequency for a given signal. While it shares similarities with the Fourier transform in its use of fixed basis functions, the STFT differs by employing fixed-size time-shifted window functions to transform the signal, or $w(n)$.

$$X(k, m) = \sum_{n=0}^{N-1} x(n+m)w(n)W_N^{nk}, \quad (14)$$

$$k, m = 0, 1, \dots, N-1,$$

where m is the amount of shift. When compared to the Fourier transform, the STFT offers superior capability in terms of temporal and spectrum localised information. The produced characteristics, on the other hand, are unable to swiftly localise time and frequency because of the unending combination of temporal and frequency precision that is induced by the traditional Heisenberg's uncertainty principle. Furthermore, due to the fact that it has a constant time window and fixed basis functions, the STFT is still unable to detect events that have varying durations or when the signal is composed of quick events.

2.5 LSTM

Deep learning and intelligent systems use a Short-Term Long Memory (LSTM) deep neural network. In addition to analyzing single data bits, this type of Recurrent Neural Network (RNN) can analyze whole data sequences. Similar to how the moment-to-moment shift

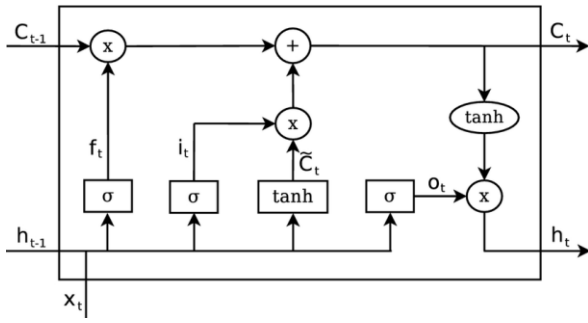


Fig. 2 LSTM architecture.

in electromagnetic discharge frequencies in the brain maintains short attention spans, the activity patterns in the system are disturbed once each time step. According to biological changes in synapse strength that are used to retain long-term memories, the connectivity weights and biases in the system disturbances once for each occurrence of training. Long-term memory is the term used to describe the short-term memory that the LSTM design aims to provide for RNN that can withstand thousands of time-steps.

Input and output gates are the only components of the LSTM design. The expressed as following:

$$i_t = \sigma(W_{ih}h_{t-1} + W_{ix}x_t + b_i), \quad (15)$$

$$c_t = \tanh(W_{ch}h_{t-1} + W_{cx}x_t + b_c), \quad (16)$$

$$c_t = c_{t-1} + i_t c_t, \quad (17)$$

$$o_t = \sigma(W_{oh}h_{t-1} + W_{ox}x_t + b_o), \quad (18)$$

$$h_t = o_t \tanh(c_t), \quad (19)$$

where c_t denotes the cell state of LSTM. W_i , W_{ch} , and W_o are the weights, and the operator ‘ $\cdot$ ’ shows two vectors multiplied together pointwise. The LSTM’s structure is seen in Fig. 2. The LSTM’s parameters are shown in Table 2. The selected loss function for the training process is Cross Entropy. The Adam optimizer is employed to optimize the model’s parameters during training. A dropout rate of 0.5 is implemented, indicating that 50% of the neurons are randomly omitted during each training iteration to prevent overfitting. The training process is set to run for a maximum of 800 epochs, defining the number of times the entire dataset is processed by the model. During each iteration of training, a batch size of 100 is utilized, determining the number of data samples processed before updating the model’s parameters. The learning process commences with an initial learning rate of 0.005, influencing the size of the steps taken during the optimization process. The Sigmoid activation function is chosen for the gates within the LSTM network, contributing to the control over the flow of information in the memory cell.

Algorithm 1 outlines the proposed method for classifying EEG signals, starting with input EEG data

denoted as S_k and aiming to produce a classified EEG signal as output. The algorithm iterates through a dataset of EEG signals, performing a sequence of processing steps on each signal.

For each iteration from $k = 0$ to N , the algorithm follows these steps: Read the EEG signal S_k . Apply preprocessing to the signal, resulting in the preprocessed signal S_R . Extract statistical features from the preprocessed signal, which characterizes its statistical properties. Further extract frequency features (F_R) from the preprocessed signal to capture its spectral characteristics. Initialize the LSTM network ($LSTM_I$) with training parameters. Train the LSTM network (LSTM) using the initialized parameters, along with the statistical and frequency features (S_k and M_R) obtained from the EEG signal. The algorithm repeats these steps for each signal in the dataset. The overall goal of this process is to classify EEG signals effectively using LSTM-based neural networks. It is important to note that the actual processing details and specific parameters used in these steps may vary depending on the implementation and goals of the EEG signal classification task.

Table 2 Training parameter of LSTM.

Parameter	Value
Loss Function	Cross Entropy
Optimizer	Adam
Dropout	0.5
Maximum Epochs	800
Batch size	100
Initial learning rate	0.005
Gate activation function	Sigmoid

Algorithm 1 Algorithm of Proposed method.

Input	: Signal(S_k)
Output	: Classified EEG Signal
For $k=0$ to N Read Signal S_k , $S_R = \text{preprocessed}(S_k)$ $M_R = \text{extract statistical feature}(S_R)$ $F_R = \text{extract frequency feature}(S_R)$ $LSTM_I =$ Initialize(Training Parameter) $LSTM = \text{Train}(UNET_I, S_k, M_R)$ EndFor	

The proposed BCI work is implemented using the Python programming language. Google Colab environment is used to compile the Python code. Various libraries are used to perform the feature extraction and ML training process. Keras library is used to perform the LSTM training operation. To improve the computation speed, the GPU of Colab is utilized. In this work, 80% of the data is used for training, and 20% is used for testing.

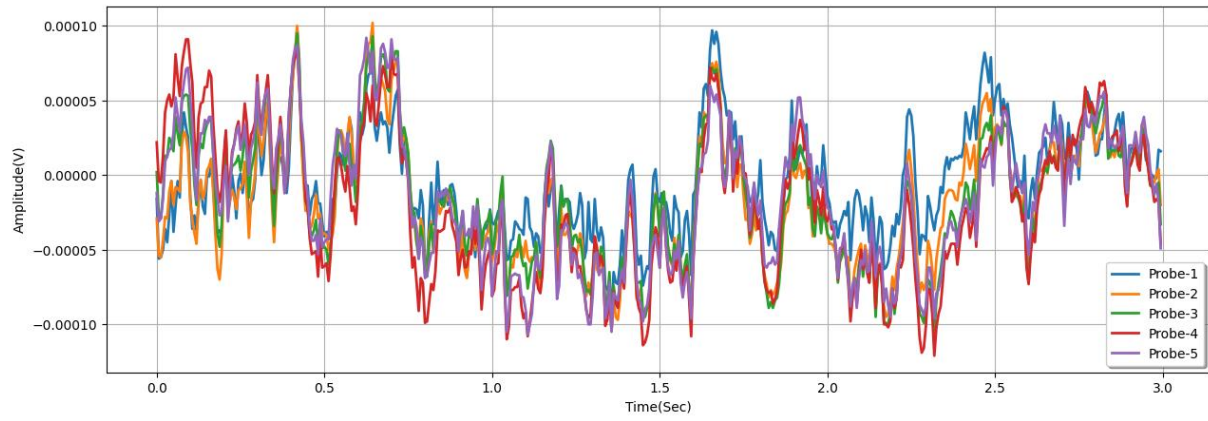


Fig. 3 Sample EEG signals from from the dataset.

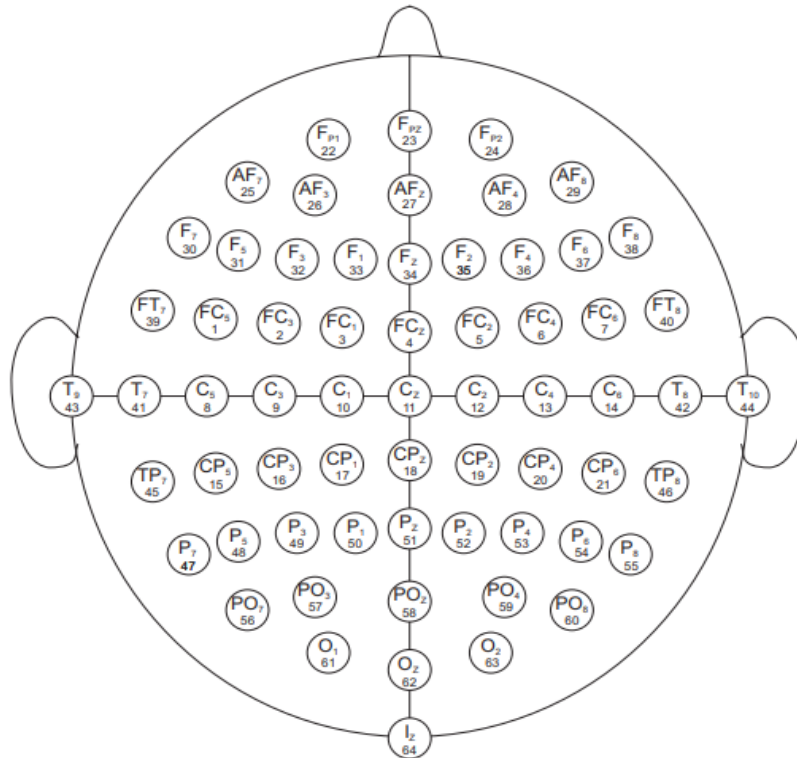


Fig. 4 Probe arrangement.

2.6 Performance Metrics

The accuracy gauges how many instances are correctly classified about the number of occurrences. The sensitivity may be stated as the proportion of correctly labelled positives. The ratio of the total of true negative and false positive results is used to measure specificity. The proportion of better conversion results in diagnostic tests and analytics is known as precision.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad (20)$$

$$Sensitivity = \frac{TP}{TP+FN}, \quad (21)$$

$$Specificity = \frac{TN}{TN+FP}, \quad (22)$$

$$Precision = \frac{TP}{TP+FP}. \quad (23)$$

3 Results and Discussion

As indicated in Fig. 4, the EEGs were collected from 64 neurons using the worldwide 10-10 standard, eliminating terminals Nz, F9, F10, FT9, FT10, A1, A2, TP9, TP10, P9, and P10. Noting that signals in the database are numbers from 0 to 63, whereas the numerals in the picture vary from 1 to 64, the numbers beneath each electrode name reflect the sequence wherein they appear in the data. Fig. 3 shows the sample EEG signals from the dataset. Fig. 4 depicts the placement of the probes.

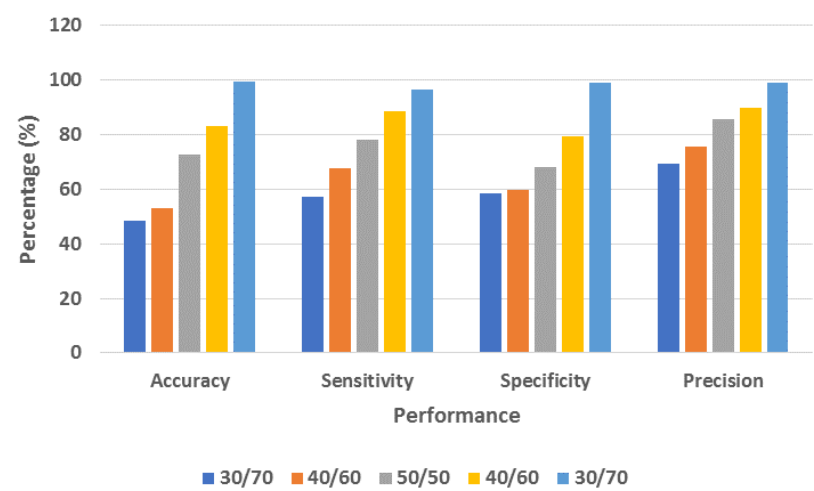


Fig. 5 Performance of the proposed method based on the training and testing ratio.

Table 3 Performance-based on training and testing ratio.

Training / Testing ratio	Accuracy	Sensitivity	Specificity	Precision
40/60	53.12%	67.49%	59.8%	75.63%
50/50	72.51%	78.17%	68.18%	85.64%
60/40	83.14%	88.39%	79.39%	89.96%
70/30	93.17%	90.21%	94.11%	94.71%
80/20	99.35%	96.38%	99.18%	99.21%

Table 4 Comparative performance of the proposed method.

Method	Accuracy	Sensitivity	Specificity	Precision
Ref. [22]	90%	–	–	–
Ref. [23]	98.31%	–	–	–
Ref. [24]	98.3%	–	–	94.7%
This work	99.35%	96.38%	99.18%	99.21%

The performance of classification process is given in Fig. 5. As shown, the accuracy, sensitivity, specificity and precision values are showed for different training ratio.

Table 3 presents the performance metrics of the model based on different training and testing data ratios. At a training/testing ratio of 40/60, the model achieved an accuracy of 53.12%, sensitivity of 67.49%, specificity of 59.8%, and precision of 75.63%. When the ratio was adjusted to 50/50, performance improved, with accuracy reaching 72.51%, sensitivity at 78.17%, specificity at 68.18%, and precision at 85.64%. Further increases in the training set proportion resulted in higher overall performance, with notable values at a ratio of 80/20, where accuracy peaked at 99.35%, sensitivity at 96.38%, specificity at 99.18%, and precision at 99.21%. These results underscore the influence of the training/testing

ratio on the model’s ability to generalize and perform effectively across different datasets.

The comparative effectiveness of the suggested strategy is displayed in Table 4. The Method [22] achieved an accuracy of 90%, while the Method [23] demonstrated an improved accuracy of 98.31%. Further, the Method [24] exhibited an accuracy of 98.3% and displayed a notable precision of 94.7%. The current work, however, outperformed these methods, attaining an impressive accuracy of 99.35%. Additionally, it achieved a high sensitivity rate of 96.38% and an impressive specificity of 99.18%. The precision for this work was also commendable at 99.21%. These results collectively underscore the advancements made in the current study compared to the previous methods. The suggested method’s confusion matrix is displayed in Fig. 6. Fig. 7 shows the performance of the proposed method based on each class.

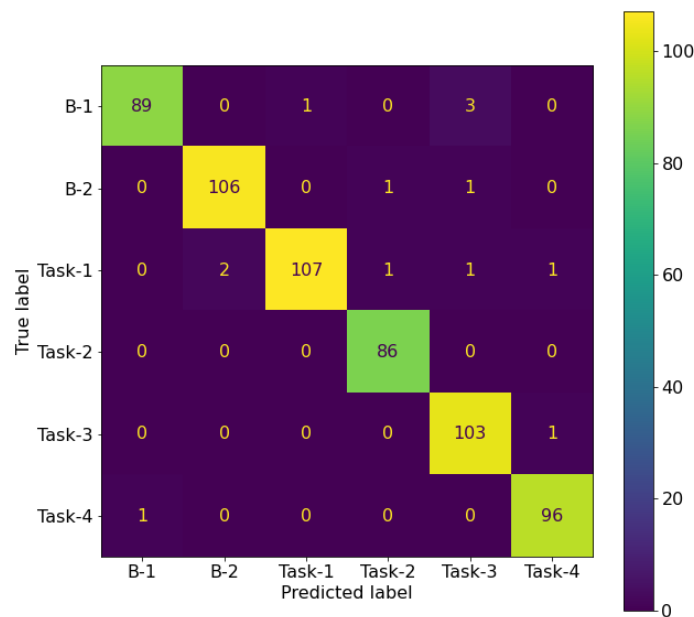


Fig. 6 Confusion matrix of the proposed method.

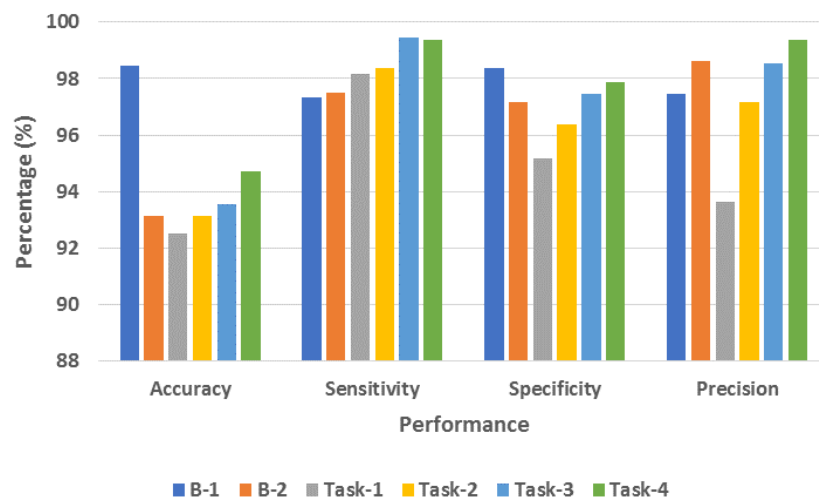


Fig. 7 Performance of the proposed method.

4 Conclusion

The proposed research focuses on enhancing EEG-based intention recognition through the implementation of LSTM networks in conjunction with signal normalization and smoothing techniques. The effectiveness of these refined algorithms is demonstrated through extensive experimentation with publicly available BCI databases, showcasing their ability to accommodate a wide range of human intentions and EEG sensitivities. Furthermore, this work includes a comprehensive examination of the various iterations of the proposed model and their interaction with dataset characteristics. This breakthrough represents a significant advancement in the precise identification of human intentions, holding great significance for the advancement of practical BCI system development. The findings of this work are supported by impressive performance metrics, including a remarkable accuracy

rate of 99.35%, a sensitivity rate of 96.38%, specificities that reach 99.18%, and an exceptional precision rate of 99.21%.

Author Contributions

This paper is the responsibility of all authors, who authorised its submission.

Disclosures

On representation of all authors, the associated author denies conflict of interest.

Data Availability

The data that support the findings of this study are openly available at <https://archive.physionet.org/pn4/eeegmmidb/> (accessed 10 January 2024).

References

1. N. Kosmyna, A. Lécuyer, “A conceptual space for EEG-based brain-computer interfaces,” *PLoS ONE* 14(1), e0210145 (2019).
2. X. Gu, Z. Cao, A. Jolfaei, P. Xu, D. Wu, T.-P. Jung, and C.-T. Lin, “EEG-Based Brain-Computer Interfaces (BCIs): A Survey of Recent Studies on Signal Sensing Technologies and Computational Intelligence Approaches and Their Applications,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 18(5), 1645–1666 (2021).
3. H. Cecotti, A. Graser, “Convolutional Neural Networks for P300 Detection with Application to Brain-Computer Interfaces,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33(3), 433–445 (2011).
4. N. Padfield, J. Zabalza, H. Zhao, V. Masero, and J. Ren, “EEG-Based Brain-Computer Interfaces Using Motor-Imagery: Techniques and Challenges,” *Sensors* 19(6), 1423 (2019).
5. C.-T. Lin, T.-T. N. Do, “Direct-Sense Brain-Computer Interfaces and Wearable Computers,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 51(1), 298–312 (2021).
6. N. Even-Chen, S. D. Stavisky, C. Pandarinath, P. Nuyujukian, C. H. Blabe, L. R. Hochberg, J. M. Henderson, and K. V. Shenoy, “Feasibility of Automatic Error Detect-and-Undo System in Human Intracortical Brain-Computer Interfaces,” *IEEE Transactions on Biomedical Engineering* 65(8), 1771–1784 (2018).
7. O.-Y. Kwon, M.-H. Lee, C. Guan, and S.-W. Lee, “Subject-Independent Brain-Computer Interfaces Based on Deep Convolutional Neural Networks,” *IEEE Transactions on Neural Networks and Learning Systems* 31(10), 3839–3852 (2020).
8. C.-H. Han, K.-R. Muller, and H.-J. Hwang, “Enhanced Performance of a Brain Switch by Simultaneous Use of EEG and NIRS Data for Asynchronous Brain-Computer Interface,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28(10), 2102–2112 (2020).
9. Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang, and A. Cichocki, “Sparse Bayesian Classification of EEG for Brain-Computer Interface,” *IEEE Transactions on Neural Networks and Learning Systems* 27(11), 2256–2267 (2016).
10. A. Gupta, R. K. Agrawal, J. S. Kirar, B. Kaur, W. Ding, C.-T. Lin, J. Andreu-Perez, and M. Prasad, “A hierarchical meta-model for multi-class mental task based brain-computer interfaces,” *Neurocomputing* 389, 207–217 (2020).
11. D. J. McFarland, J. R. Wolpaw, “EEG-based brain-computer interfaces,” *Current Opinion in Biomedical Engineering* 4, 194–200 (2017).
12. V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, “EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces,” *Journal of Neural Engineering* 15(5), 056013 (2018).
13. K. Wang, M. Xu, Y. Wang, S. Zhang, L. Chen, and D. Ming, “Enhance decoding of pre-movement EEG patterns for brain-computer interfaces,” *Journal of Neural Engineering* 17(1), 016033 (2020).
14. D. Zhang, L. Yao, X. Zhang, S. Wang, W. Chen, R. Boots, and B. Benatallah, “Cascade and Parallel Convolutional Recurrent Neural Networks on EEG-based Intention Recognition for Brain Computer Interface,” *Proceedings of the AAAI Conference on Artificial Intelligence* 32(1), (2018).
15. K.-S. Hong, M. J. Khan, and M. J. Hong, “Feature Extraction and Classification Methods for Hybrid fNIRS-EEG Brain-Computer Interfaces,” *Frontiers in Human Neuroscience* 12, 246 (2018).
16. Y. Wang, M. Nakanishi, and D. Zhang, “EEG-Based Brain-Computer Interfaces,” in *Neural Interface: Frontiers and Applications. Advances in Experimental Medicine and Biology*, 1101, X. Zheng (Ed.), Springer, Singapore, 41–65 (2019).
17. Z. Jin, G. Zhou, D. Gao, and Y. Zhang, “EEG classification using sparse Bayesian extreme learning machine for brain-computer interface,” *Neural Computing and Applications* 32(11), 6601–6609 (2020).
18. C. Ieracitano, N. Mammone, A. Hussain, and F. C. Morabito, “A novel explainable machine learning approach for EEG-based brain-computer interface systems,” *Neural Computing and Applications* 34(14), 11347–11360 (2022).
19. G. Zhang, V. Davoodnia, A. Sepas-Moghaddam, Y. Zhang, and A. Etemad, “Classification of Hand Movements From EEG Using a Deep Attention-Based LSTM Network,” *IEEE Sensors Journal* 20(6), 3113–3122 (2020).
20. E. C. Djamal, H. Fadhilah, A. Najmurokhman, A. Wulandari, and F. Renaldi, “Emotion brain-computer interface using wavelet and recurrent neural networks,” *International Journal of Advances in Intelligent Informatics* 6(1), 1 (2020).
21. K. Takahashi, Z. Sun, J. Solé-Casals, A. Cichocki, A. H. Phan, Q. Zhao, H.-H. Zhao, S. Deng, and R. Micheletto, “Data augmentation for Convolutional LSTM based brain computer interface system,” *Applied Soft Computing* 122, 108811 (2022).
22. Tan C, Sun F, Zhang W. Deep transfer learning for EEG-based brain-computer interface. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2018. p. 916-920.
23. S. Sheykhivand, Z. Mousavi, T. Y. Rezaei, and A. Farzammia, “Recognizing Emotions Evoked by Music Using CNN-LSTM Networks on EEG Signals,” *IEEE Access* 8, 139332–139345 (2020).
24. Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang, and A. Cichocki, “Sparse Bayesian Classification of EEG for Brain-Computer Interface,” *IEEE Transactions on Neural Networks and Learning Systems* 27(11), 2256–2267 (2016).