

PAPER

EEG-based classification of imagined digits using a recurrent neural network

To cite this article: Nrushsingh Charan Mahapatra and Prachet Bhuyan 2023 *J. Neural Eng.* **20** 026040

View the [article online](#) for updates and enhancements.

You may also like

- [Online classification of imagined speech using functional near-infrared spectroscopy signals](#)

Alborz Rezzazadeh Sereshkeh, Rozhin Yousefi, Andrew T Wong et al.

- [Imagined speech increases the hemodynamic response and functional connectivity of the dorsal motor cortex](#)

Xiaopeng Si, Sicheng Li, Shaoxin Xiang et al.

- [Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features](#)

Chuong H Nguyen, George K Karavas and Panagiotis Artempiadis

The Breath Biopsy® Guide
Fourth edition

BREATH BIOPSY

OWLSTONE MEDICAL

Journal of Neural Engineering



PAPER

RECEIVED
22 November 2022

REVISED
30 March 2023

ACCEPTED FOR PUBLICATION
31 March 2023

PUBLISHED
28 April 2023

EEG-based classification of imagined digits using a recurrent neural network

Nrushsingh Charan Mahapatra^{1,2,*} and Prachet Bhuyan²

¹ Intel Technology India Pvt Ltd, Bengaluru 560103, India

² School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar 751024, India

* Author to whom any correspondence should be addressed.

E-mail: 1981030@kiit.ac.in

Keywords: brain–computer interfaces, deep learning, electroencephalography (EEG), signal processing, imagined speech, bidirectional recurrent neural network

Abstract

Objective. In recent years, imagined speech brain–computer (machine) interface applications have been an important field of study that can improve the lives of patients with speech problems through alternative verbal communication. This study aims to classify the imagined speech of numerical digits from electroencephalography (EEG) signals by exploiting the past and future temporal characteristics of the signal using several deep learning models. **Approach.** This study proposes a methodological combination of EEG signal processing techniques and deep learning models for the recognition of imagined speech signals. EEG signals were filtered and preprocessed using the discrete wavelet transform to remove artifacts and retrieve feature information. To classify the preprocessed imagined speech neural signals, multiple versions of multilayer bidirectional recurrent neural networks were used. **Main results.** The method is examined by leveraging MUSE and EPOC signals from MNIST imagined digits in the MindBigData open-access database. The presented methodology's classification performance accuracy was noteworthy, with the model's multiclass overall classification accuracy reaching a maximum of 96.18% on MUSE signals and 71.60% on EPOC signals. **Significance.** This study shows that the proposed signal preprocessing approach and the stacked bidirectional recurrent network model are suitable for extracting the high temporal resolution of EEG signals in order to classify imagined digits, indicating the unique neural identity of each imagined digit class that distinguishes it from the others.

1. Introduction

Speech is a fundamental human characteristic in terms of communication, with vocal language facilitated by significant interactions between neurological processing and pronunciation. Further, speech is linked to electrical activity in the brain or biosignals, which are recorded in both the temporal and frequency domains. The non-invasive electroencephalography (EEG) modality monitors the neural signals of the brain's thoughts of words, syllables, numbers, etc by placing electrodes or sensors on the scalp [1]. In patients with speech difficulties, an EEG modality with high temporal resolution is a suitable method for imagined speech-based communication, but has drawbacks such as a low signal-to-noise ratio (SNR) and typically being not very wearable.

The inner pronunciation of words without the release of sounds is known as imagined speech [2]. In an attempt to address the symptoms of certain speech disabilities and enhance the ability of those afflicted to communicate and socialize effectively, recent explorations in human–computer interface technology, known as brain–computer interface communication systems, are primarily based on neuroimaging signals and therefore do not rely on the individual brain's normal output channels of peripheral nervous system and muscle activity. The purpose of the imagined speech brain–computer (machine) interface is to transform neural signals based on imagined speech into written representations [3]; furthermore, research demonstrates that neural speech prostheses can translate electrocorticography signals into meaningful synthesized speech [4]. Researchers

from the fields of computer science and neuroscience from all over the world are trying to decode speech from brain waves. The research studies of imagined speech decoding using EEG signals are distributed into the classification of the speech prompts, such as syllables [5, 6], vowels [7–9], characters [10], digits [11–14] and words [15, 16].

Recent studies have employed deep learning techniques for EEG-based imagined speech decoding for a variety of speech prompts and recent research in this realm has been investigated to understand these advancements. Sereshkheh *et al* [17] experimented with the imagined speech of two English words using the multilayer perceptron classifier and obtained an average accuracy of 63.2%. Cooney *et al* [18] explored the decoding of five imagined vowels and attained a classification accuracy of 35.68% using the pre-processing independent component analysis (ICA) method and classifier convolutional neural networks (CNNs) as well as transfer learning techniques. Berg *et al* [19] used ICA for signal preprocessing and a CNN to classify four words, with an average accuracy of 29.67% for the classification of imagery speech. Parhi and Tewfik [20] used two different deep learning models, one CNN-based and one long short-term memory (LSTM)-based, to classify three vowels and achieved an accuracy of more than 85% for vowel decoding. Rusnac and Grigore [21] combined Mel cepstral coefficients and the CNN model to classify 10 different phonemes and words, yielding a mean accuracy of 39% for imagined speech recognition. Sarmiento *et al* [22] used a CNN-based model for the classification of imagined vowels, obtaining accuracy of 65.62% and 85.66% for two imagined speech vowel databases, respectively. Lee *et al* [23] used a deep neural network model with mixed attention layers to classify four words of imagined speech. Singh and Gumaste [24] used the covariance matrix for signal preprocessing and the classifier artificial neural network for word decoding, reaching an average classification accuracy of 85% when classifying one long word versus one short word. Vorontsova *et al* [25] used the transfer learning classification model ResNet18 in conjunction with a gated recurrent unit (GRU) neural network for nine words and attained 85% average accuracy while performing multiclass classification. Lee *et al* [26] used a deep metric learning CNN-based classification model for decoding words and obtained a six-class classification accuracy of 45%. Rusnac and Grigore [27] decoded imagined speech from multiple subjects employing a feature extraction method based on cross-covariance in the frequency domain and used the CNN method for multiclass classification, achieving 37% accuracy. An observation has been made that the majority of research has concentrated on employing CNN-based models to decode imagined speech, whereas a couple

of studies have employed unidirectional recurrent neural networks (RNNs) to achieve a certain level of accuracy.

Furthermore, studies on imagined digit decoding utilizing deep learning and machine learning approaches have been reviewed. Mishra *et al* [12] investigated imagined digits signal from MindBigData (EPOC), which were preprocessed with a correlation approach with 14-channel data refinement and classified using a CNN, yielding a classification accuracy of 32.4%. Bird *et al* [14] classified the imagined digits signal from MindBigData (MUSE) with the help of an optimized multilayer perceptron (MLP) and obtained an accuracy of 27.07%, whereas the unidirectional LSTM model only managed to classify them with an accuracy of 10.77% and there were no statistically significant findings. Jolly *et al* [13] used the unidirectional GRU model and the CNN model to classify the imagined digits signal from MindBigData (MUSE) and obtained a multiclass classification performance of 33.8% accuracy. In another experiment by Pratama *et al* [28], the imagined digit signals from MindBigData were classified using feature extraction applying principal component analysis and the classifier k -nearest neighbors, with an average accuracy of 12.3%. Additionally, the investigation by Kumar *et al* [11] classified imagined speech signals consisting of digits, characters and objects using the random forest machine learning technique and achieved a classification accuracy of 67.03%.

Despite several attempts, decoding imagined speech EEG signals remains challenging due to low SNR and a deficit in understanding the relevant neural processes. In some previous studies, unidirectional RNN techniques were employed to utilize temporal information to decode imagined speech [14, 20, 25], whereas the majority of the studies involved the CNN model in the decoding process [18, 19, 21, 22, 26, 27]. The CNN frameworks could not retrieve the long-term temporal feature correlations of the signals as the convolution layer extracts the spatial information. The framework of a unidirectional RNN can be used to extract temporal features. However, the bidirectional RNN frameworks further enhance the temporal correlation of feature extraction as the layer receives input from both future and past instances simultaneously. As the subject in the signal acquisition trial imagined the digit for a period of seconds, the past–future temporal information of the EEG signal could have contained distinguishing feature information of an imagined digit, which should be included in the investigation. There is indeed potential for further investigation into the decoding of imagined digits by leveraging a bidirectional RNN architecture.

In many disciplines of research, variants of RNNs are effective at decoding sequence information, such

as in automatic speech recognition [29, 30], statistical machine translation [31], sequence tagging [32] and text classification [33]. The simple RNN framework is hindered by a limited attention span, which is enhanced in the GRU and LSTM variants by the gate structure, which is employed to regulate the information flow throughout the sequence and to reduce the influence of the short attention span. In this study, multilayer bidirectional-RNN, multilayer bidirectional-GRU and multilayer bidirectional-LSTM models were investigated for incorporating temporal information from multichannel EEG signals of imagined digits. The experiment was carried out using open-access imagined speech data. The raw signals were preprocessed using discrete wavelet transform (DWT) techniques before being used to train and evaluate the model. The deep learning model effectively decoded or classified the multiclass imagined digits neural signals. The contributions of this experiment are summarized as follows:

- This study explores deep learning models (multilayer bidirectional RNNs) that can learn feature representations of future and past activation temporal features in imagined digit EEG signals and effectively decode the imagined digit.
- Furthermore, leveraging MindBigData's EPOC and MUSE databases [34], it was asserted that the framework can be applied to different low-density EEG signal modalities.

2. Methods

The imagined speech-based brain–computer interface is fundamentally a combination of hardware and software components. The system structure depicted in figure 1 is composed of several components, including EEG signal acquisition, signal preprocessing and signal classification. During the signal acquisition phase, the EEG device was utilized to acquire imagined digit EEG signals from subjects. The raw EEG signals contain artifacts, which are removed and signal features extracted during the signal preprocessing phase. Subsequently, multiclass classifiers, such as multilayer bidirectional RNNs, are used in the signal classification phases.

2.1. Experimental EEG data

The EEG signal sets used in the experiments were EPOC and MUSE from the open-access database MindBigData [34]. The EPOC headset was used to record the multichannel EEG signals. EEG signals from each trial were recorded for 2 s. The signals were collected at a 128 Hz sampling rate and have 14 channels (AF4, F4, F8, FC6, O2, P8, T8, T7, P7, O1, FC5, F7, F3 and AF3), with the electrode placement according to the 10–20 international system [35]. The

MUSE headset was used to record the multichannel EEG signals. The four-channel (TP9, FP1, FP2 and TP10) electrode placements were used to record EEG data for 2 s at a sampling rate of 220 Hz in each trial. Table 1 summarizes the EEG signal attribute information from the EPOC and MUSE open-access databases.

2.2. EEG signal preprocessing

Signal preprocessing is a complex technique that is used to eliminate any undesirable information contained in the EEG signal and an improvement in signal quality from effective preprocessing boosts feature separability and the performance of the classifier. The raw EEG signals were preprocessed in the experiment with an infinite impulse response Butterworth high-pass filter of order 5 at 0.1 Hz to erase the low-frequency noise, followed by a notch filter to remove 60 Hz electrical environmental noise. The feature extraction technique used in this study was based on the wavelet transformation [36]. The wavelet transform splits or decomposes the EEG signals into a set of orthogonal basis function wavelets and reconstructs the feature information accurately to form the feature information vector, with part of the decomposed wavelet coefficients cancelled out before the inverse transformation [37]. The wavelet coefficients have been created by dilation, translation and scaling of the mother wavelet. The decomposition was computed by repeatedly filtering the discrete signals up to a predefined level. Earlier studies have demonstrated the effectiveness of utilizing Daubechies-4 (order 4) wavelet feature extraction for the classification of different types of EEG signals [38–41]. In this study, we applied the DWT using the Daubechies-4 wavelet with two-level decomposition on EPOC and the three-level decomposition on MUSE for denoising and information extraction. At the first level, each channel signal was decomposed into approximation coefficients (which were in the lower frequency sub-band) and detail coefficients (which were in the higher frequency sub-band). At the second level, the first-level approximation coefficients (the lower frequency sub-band) were decomposed into approximation coefficients and detail coefficients. Similarly, the lower frequency sub-band approximation coefficients were decomposed at each level. All coefficients were used for the synthesized feature information vector after acquiring all approximation and detail coefficients at the final level of decomposition. The thresholding technique was used to eliminate small components from the wavelet representation, assuming they were primarily noise [36]. Inverse reconstruction of the original EEG waveforms was performed using the remaining DWT components, and the reconstructed signals or feature information vectors were used as inputs to neural networks.

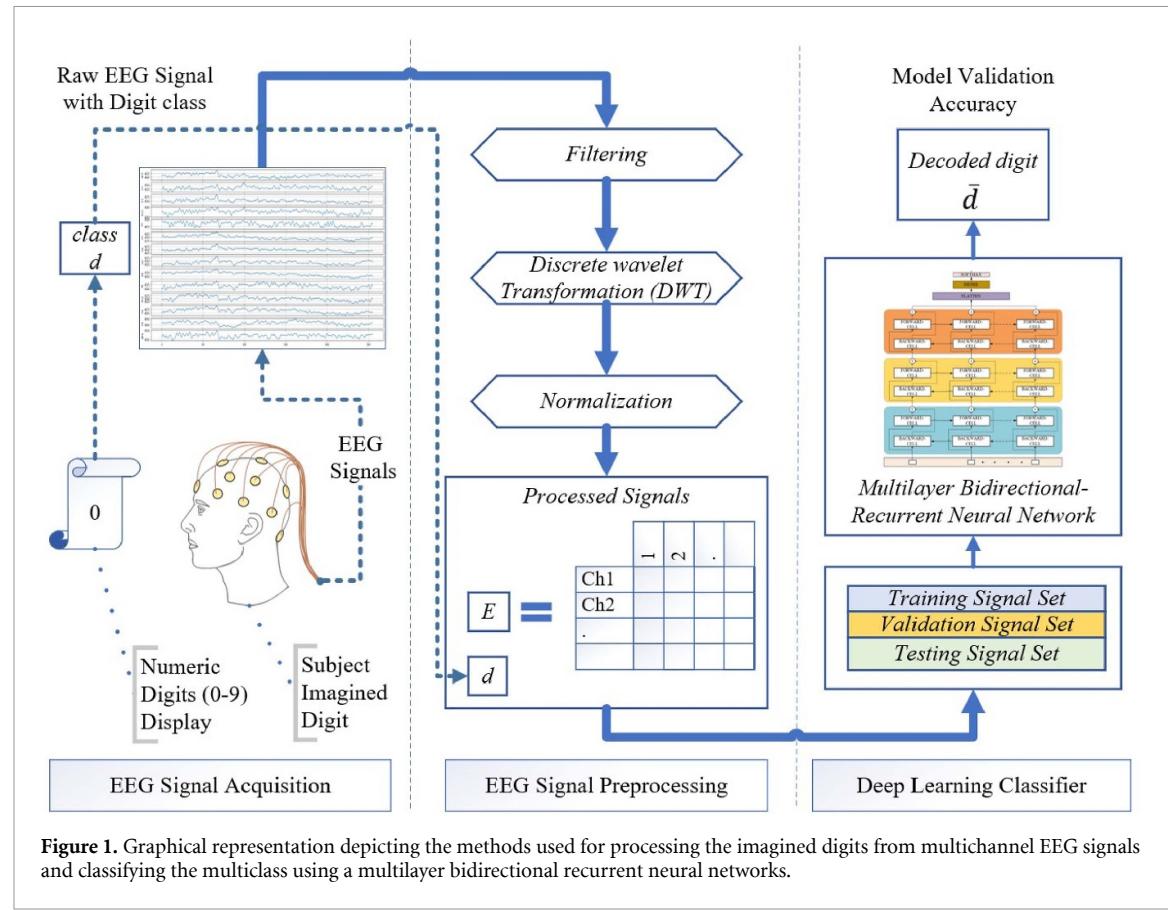


Figure 1. Graphical representation depicting the methods used for processing the imagined digits from multichannel EEG signals and classifying the multiclass using a multilayer bidirectional recurrent neural networks.

Table 1. Information from the MindBigData EEG signal open-access database.

Signals attributes	EPOC	MUSE
EEG device	Emotiv EPOC	Interaxon Muse
No. of channels	14	4
Sampling rate	128 Hz	220 Hz
No. of signals	910 476 ^a	163 932 ^a
Speech class	10 digits ^b	10 digits ^b

^a EEG signals were obtained over the course of about 2 years on a single subject performing multiple trials and the subject was the same in both datasets.

^b Imagined speech digits: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

Figures 2 and 3 depict the DWT signal decomposition and reconstruction using the signals EPOC and MUSE, respectively.

The input signal vectors were standardized using the Z-score normalization process before being fed into the model for training and validation. The Z-score specified in equation (1) was applied to each record in the dataset. The standardized input signals were normalized using the min–max scaling method, yielding input signal vectors ranging from 0 to 1.

$$Z \text{ score}(E) = (E - \mu)/\sigma, \quad (1)$$

where μ and σ are mean as well as the standard deviation, respectively, of all the EEG signal records.

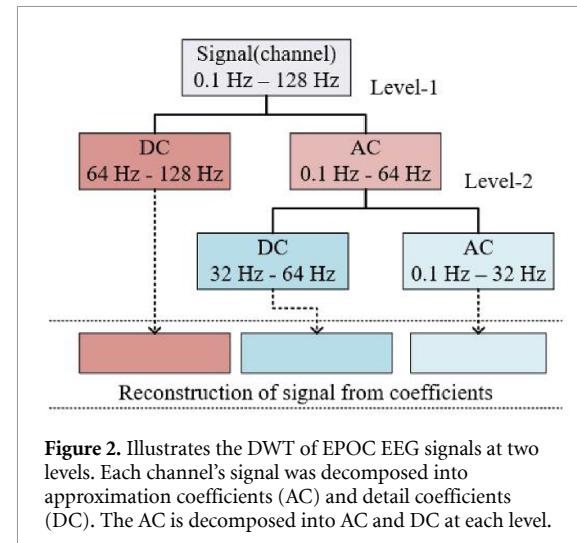


Figure 2. Illustrates the DWT of EPOC EEG signals at two levels. Each channel's signal was decomposed into approximation coefficients (AC) and detail coefficients (DC). The AC is decomposed into AC and DC at each level.

2.3. Deep learning model architecture

The purpose of this study is to assess RNN algorithms for imagined digit decoding from the EEG signals of the subject. The input data is interpreted as an EEG signal (E) from a multichannel time sampling point vector, denoted by equation (2). The output data is imagined digits $d \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ class. The proposed supervised classification model is defined as in equation (3).

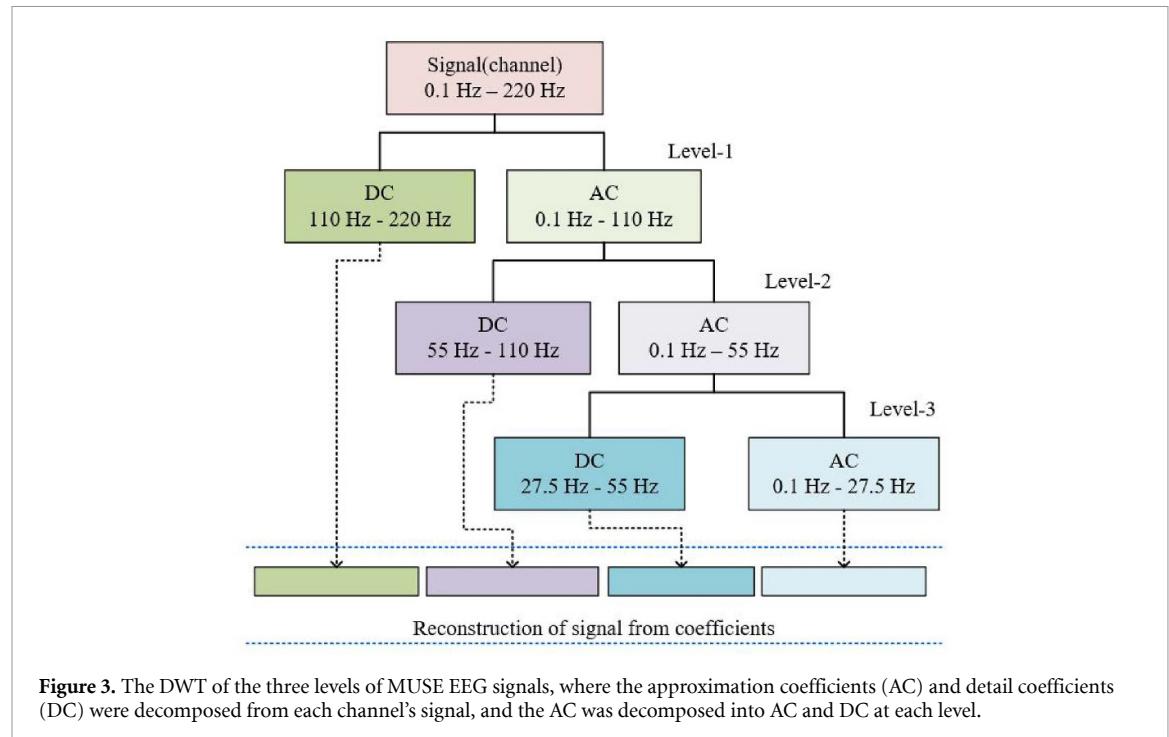


Figure 3. The DWT of the three levels of MUSE EEG signals, where the approximation coefficients (AC) and detail coefficients (DC) were decomposed from each channel's signal, and the AC was decomposed into AC and DC at each level.

$$\text{signal } (E) = \begin{pmatrix} e_{1,1} & e_{1,2} & \dots & e_{1,i} & \dots & e_{1,m} \\ e_{2,1} & e_{2,2} & \dots & e_{2,i} & \dots & e_{2,m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ e_{j,1} & e_{j,2} & \dots & e_{j,i} & \dots & e_{j,m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ e_{n,1} & e_{n,2} & \dots & e_{n,i} & \dots & e_{n,m} \end{pmatrix}, \quad (2)$$

where $e_{(j,i)} \in R$ is the sampling point of the channel j at the time i and (n, m) represent the input dimensions defined as the total number of channel signal features as well as the total number of sampling points per channel, respectively.

$$f(E) = \bar{d} + Err, \quad (3)$$

where E is the input EEG signal, \bar{d} represents the one-hot encoding of the digit and Err is the classification error.

The proposed deep learning model was designed to decode the conditional probability $P(d|E)$ derived from the multivariate input EEG signal vector E into a class of probability distribution d digit class. Figure 4 illustrates the deep learning model architecture of the multilayer bidirectional RNNs. In the experiments, three variants of RNNs were employed, such as multilayer bidirectional-RNN, multilayer bidirectional-GRU, or multilayer bidirectional-LSTM frameworks. The architecture includes an input layer of preprocessed EEG signals, multiple layers of bidirectional-cell, a flattening layer, a fully connected or dense layer, and an output layer with softmax classification.

2.3.1. RNN cell

The simple RNN cell output is obtained by using the input signal as well as the previous hidden state and

running that through the hyperbolic tangent activation function (\tanh) without any gates. The non-linear activation function \tanh has a range of -1 to 1 and is defined as $\tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$. Simple RNN frameworks have short-term memory issues that impede the learning of long data sequences due to gradients exploding or vanishing. Figure 5 illustrates the simple RNN cell. Given an input signal $e_t = [e_{(1,t)}, e_{(2,t)}, \dots, e_{(m,t)}]$ at time t , at a time stamp $(t-1)$ the hidden state of the prior layer is s_{t-1} . The computation of the output at time stamp t is o_t , and the hidden state at time stamp t is s_t of a simple RNN cell, as defined in equation (4). As the output o_t is determined based on the hidden state s_{t-1} at time step $(t-1)$ and input signal e_t at time step t , hence input e_{t-1} at time step $(t-1)$ can affect the output o_t at time step t through the recurrent linking.

$$\text{output}(o_t) = s_t = \tanh(w_s * s_{t-1} + w_e * e_t + b_t), \quad (4)$$

where w_s is the weight matrix of the hidden state, w_e is the input weight matrix and b_t is the bias matrix at time stamp t .

2.3.2. GRU cell

A GRU is an RNN variation that includes a gated procedure to govern information flow between cells in the neural network [31]. The GRU extracts dependencies from large volumes of sequential input signals while preserving prior information and avoiding gradient vanishing or exploding problems by using combination components such as (a) the reset gate, which determines the amount of past information from the previous time step that could be ignored

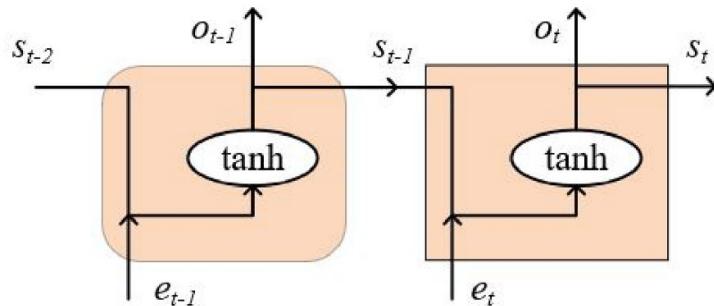
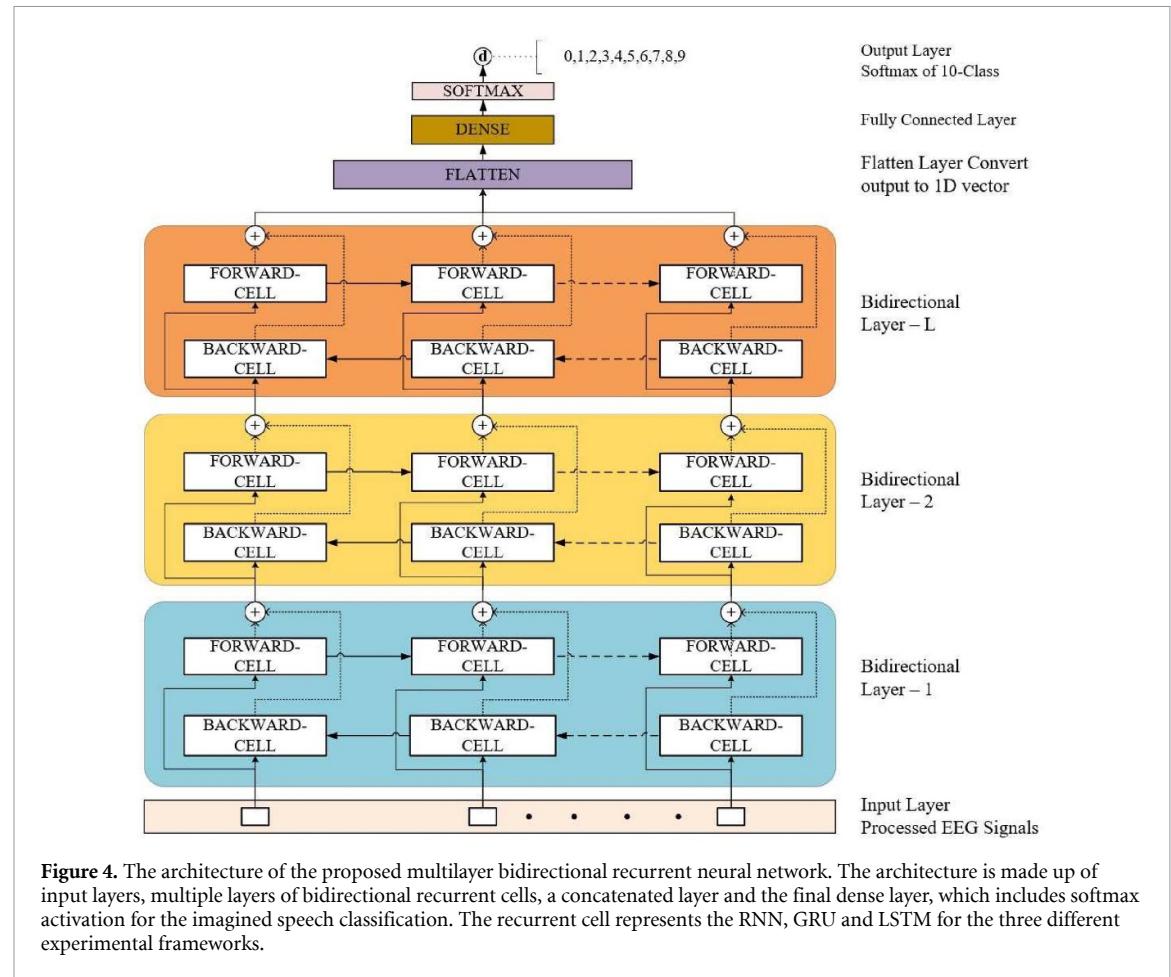


Figure 5. The recurrent units of a simple RNN cell at time stamp ($t-1$) and t .

and stored into the current memory content, (b) the update gate, which determines the amount of past information from the previous time step that should be passed into the current time step, and (c) the final memory at the current time step computed based on the update gate information and the current memory content. Figure 6 depicts the GRU memory cell.

$e_t = [e_{(1,t)}, e_{(1,t)}, \dots, e_{(m,t)}]$ is the input signal at time stamp t , s_{t-1} is the prior layer's hidden state at time stamp $(t-1)$ and c_{t-1} is the cell state at time stamp $(t-1)$. Equations for reset gates, update gates and output computation are defined in (5a)–(5d).

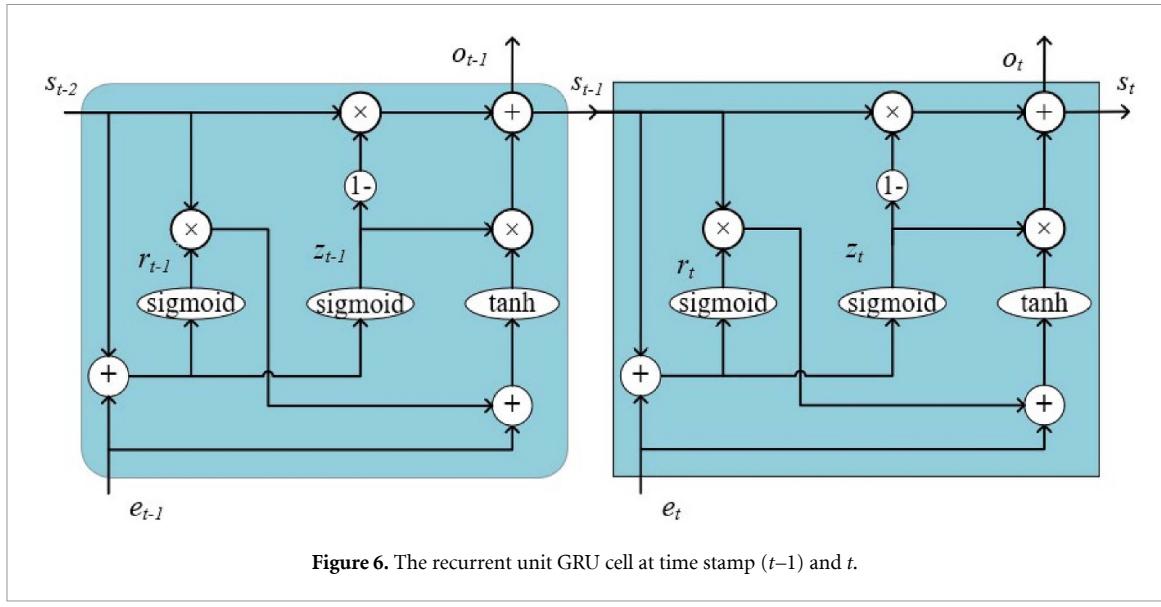
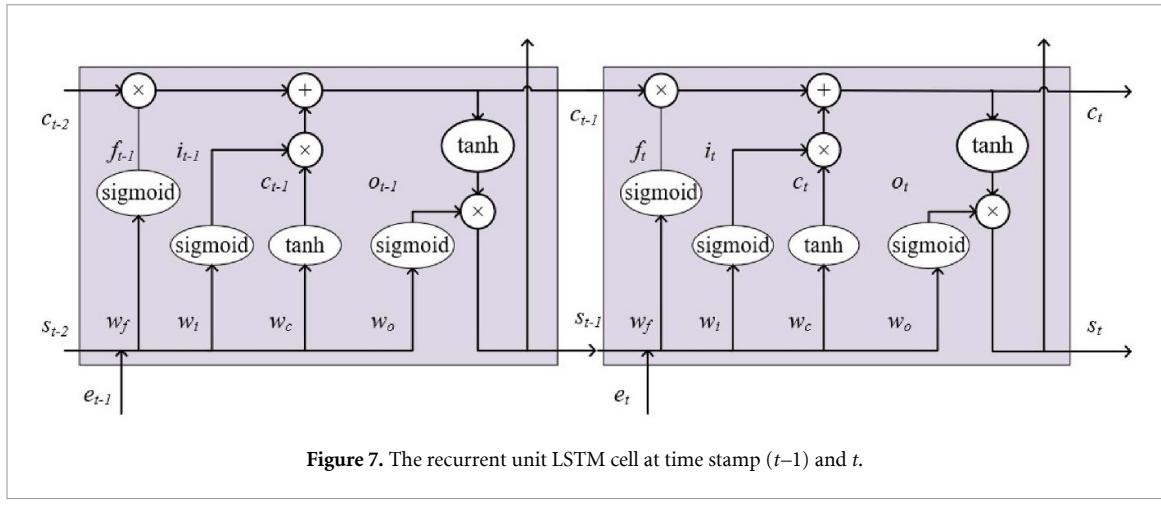
$$\text{reset gate } (r_t) = \text{sigmoid}(w_r \cdot [s_{t-1}, e_t] + b_r), \quad (5a)$$

$$\text{update gate } (z_t) = \text{sigmoid}(w_z \cdot [s_{t-1}, e_t] + b_z), \quad (5b)$$

$$\begin{aligned} \text{hidden state } (s_t) = & (1 - z_t) * s_{t-1} + z_t * \tanh \\ & \times (w_h \cdot [e_t, r_t * s_{t-1}] + b_h), \end{aligned} \quad (5c)$$

$$\text{output } (o_t) = s_t, \quad (5d)$$

where s_t is the hidden state at time stamp t , o_t is the output at time stamp t . Sigmoid is nonlinear

Figure 6. The recurrent unit GRU cell at time stamp ($t-1$) and t .Figure 7. The recurrent unit LSTM cell at time stamp ($t-1$) and t .

activation function is defined as $\text{sigmoid}(x) = (1/(1 + e^{-x}))$. The weight matrices are w_r, w_z, w_h . The bias matrices are b_r, b_z, b_h .

2.3.3. LSTM cell

The LSTM is a special form as well as an enhanced RNN that can solve both vanishing and exploding gradient issues, making it perfect for time-sequence applications with long-term dependencies [42, 43]. The LSTM cell comprises four fundamental components: (1) the input gate, which controls the information that enters the memory cell by using the nonlinear sigmoid activation function; (2) the memory cell, which remembers values over discrete time periods; (c) the forget gate, which is in charge of managing prior memory through the activation tanh; and (d) the output gate, which transfers the memory cell's output to the outside via sigmoid activation. Figure 7 depicts the LSTM cell.

$e_t = [e_{(1,t)}, e_{(1,t)}, \dots, e_{(m,t)}]$ is the input signal at time stamp t , s_{t-1} is the prior layer's hidden state at time stamp $(t-1)$ and c_{t-1} is the cell state at time

stamp $(t-1)$. Equations for gate and output computation are defined in (6a)–(6e).

$$\text{input gate } (i_t) = \text{sigmoid}(w_i \cdot [s_{t-1}, e_t] + b_i), \quad (6a)$$

$$\text{forget gate } (f_t) = \text{sigmoid}(w_f \cdot [s_{t-1}, e_t] + b_f), \quad (6b)$$

$$\begin{aligned} \text{final memory cell } (c_t) &= f_t * c_{t-1} \\ &\quad + i_t * \tanh(w_c \cdot [s_{t-1}, e_t] + b_c), \end{aligned} \quad (6c)$$

$$\text{output gate } (o_t) = \text{sigmoid}(w_o \cdot [s_{t-1}, e_t] + b_o), \quad (6d)$$

$$\text{hidden state } (s_t) = o_t * \tanh(c_t), \quad (6e)$$

where s_t is the hidden state at time stamp t , c_t is the memory cell state at time stamp t and o_t is the output at time stamp t . The nonlinear activation functions are sigmoid and tanh. The weight matrices are w_i, w_f, w_c, w_o . The bias matrices are b_i, b_f, b_c, b_o .

2.3.4. Bidirectional recurrent layers

The bidirectional recurrent layer is made up of forward and backward networks like FORWARD-CELL and BACKWARD-CELL, with the layer output being the concatenation of the outputs of both forward and backward networks, where CELL represents the RNN, LSTM or GRU in the three experimental model frameworks [44]. The bidirectional recurrent layer was able to aggregate the information from both the past and future of the signal sampling points by taking the input signal vectors in two directions, such as FORWARD-CELL feeding the signal sampling points from the past to the future and BACKWARD-CELL feeding the signal sampling points from the future to the past.

One popular method for preventing overfitting in neural networks is dropout [45]. In the proposed architecture, dropout is employed as a mask for each recurrent CELL output.

2.3.5. Input layer

The preprocessed EEG signals were fed into the input layer as a batch. The input signal has a two-dimensional representation in terms of the number of channels or features and the number of sampling points per channel. The number of units and latent feature dimensions in the memory cells of the respective layers was implemented in accordance with the number of EEG sampling points per channel in the neural signal. The signals were gathered using m sampling points per channel, and the experiments were run with the ratio of m sampling points as the number of cell units.

2.3.6. Output layer

The final bidirectional recurrent output was flattened and fed into the fully connected layer, which included an activation exponential linear unit (ELU). The ELU is described in equation (7), which has a hyperparameter $\alpha > 0$ that limits the output to being saturated by negative inputs and minimizes the issue of vanishing gradients [46]. The model's final output vector, which was generated via softmax nonlinearity activation, was a well-formed probability distribution across the output digit classes. The model output class digits represent the class with the maximum probability value.

$$\text{ELU}(x) = \begin{cases} x, & \text{if } x \geq 0 \\ \alpha(e^x - 1), & \text{otherwise} \end{cases} . \quad (7)$$

2.4. Model training and validation

The proposed deep learning model was built with the Keras [47] and Tensorflow [48] frameworks. The Adam optimization algorithm [49] and cost function cross-entropy were used to train the model. The cross-entropy loss function over individual EEG signal input E and speech digits class d is defined as in equation (8).

$$\text{cross entropy}(E, d) = -\sum_{i=0}^L 1.(i=d) \log \bar{d}_i , \quad (8)$$

where L is the total number of imagined speech digits or classes, d is the true digit class and \bar{d}_i is the model classified as a one-hot encoded digit class of signal E . If the true digit d is i , the indicator function $1.(i=d)$ equals 1, otherwise it equals 0.

The EEG datasets were balanced in relation to the intended digit class. The process involved dividing the signal set into training (80%), validation (10%) and testing (10%) in order to assess the generalization performance of a classification model and ensure that the model is not overfitting. The validation dataset was used to assess tuning or hyperparameter selection, whereas the testing dataset was used to evaluate the final model's performance.

The hyperparameters of the bidirectional RNNs include the parameters that define the structure of the model along with the parameters that influence the training of the model. Several experiments have been conducted to find optimum hyperparameters such as batch size, learning rate, regularization dropout rate, the number of recurrent CELL units (or latent feature dimensions), multilayer depth, etc. Random search and defining the search boundaries were used in this experiment to optimize the model's hyperparameters, where random combinations of values of the hyperparameters from the search space were used to improve the accuracy on the validation set [50]. The model's structural hyperparameter multilayer depth was optimized using discrete distributions in the search space that ranged from 2 to 4. The dropout rate hyperparameters of the model were optimized using a discrete distribution space with a range of 0.0–0.25 with steps of 0.05. The learning rate hyperparameters of the Adam algorithm were tuned to have values distributed uniformly between 0.0001 and 0.01. The batch size hyperparameter was tuned using discrete distributions with steps of 128 between the minimum value of 256 and the maximum value of 1024. The hyperparameters of the model were determined based on the classification performance accuracy of the imagined digit, and using the random search technique with an optimizer method called Adam to minimize cross-entropy loss, the tuned hyperparameters had a dropout rate of 0.1, multilayer depth of 3, batch size of 1024 and learning rate of 0.001. Furthermore, the proposed model's performance in decoding imagined digits has improved significantly with respect to the latent feature dimension unit size in a recurrent CELL. Table 2 summarizes the parameters of the different models for both the EPOC and MUSE open-access datasets. We kept the model framework uniform with respect to the specific dataset for the three different models, such as the multilayer bidirectional-RNN, multilayer bidirectional-GRU and multilayer bidirectional-LSTM.

Table 2. Summary of the parameters of the experimental models.

	Parameter	Values	
		EPOC	MUSE
Input layer	dimension	(14 256)	(4440)
Bidirection layer	Number of layers	3	3
Layer-1	Number of units	256	440
Layer-2	Number of units	128	220
Layer-3	Number of units	64	110
Classification	Dense layer	10	
	Activation	Softmax	
Training	Optimization	Adam	
	Loss function	Cross-entropy	
	Batch size	1024	
	Learning rate	0.001	

3. Results

In this study, the performance metrics of the models were computed as classification accuracy, precision (the percentage of correctly classified signals across all classified signals), recall (the percentage of relevant signals that were successfully classified) and f_1 -score (the precision-recall harmonic mean).

3.1. Model performance on the EPOC signals

The experimental findings of the evaluation performance measures of accuracy, precision, recall and f_1 -score for three multilayer models (bidirectional-RNN, bidirectional-GRU, and bidirectional-LSTM) with EPOC signals are shown in table 3. In this case, the overall accuracy of the multiclass classification in the multilayer bidirectional-RNN model was around 24.18% and it was about 65.90% for the multilayer bidirectional-GRU model. The bidirectional LSTM model has the highest average accuracy of all the models at 71.60%, with precision ranging from 0.66 to 0.78, recall ranging from 0.69 to 0.75, and the f_1 -score ranging from 0.70 to 0.74.

The confusion matrices of three multilayer models, bidirectional-RNN, bidirectional-GRU and bidirectional-LSTM, are shown in figures 8–10, respectively, for each of the 10 digits of imagined speech EEG signals from the EPOC. These matrices comprise information about the actual digit class and predicted digit class obtained with the proposed model, with the top-to-bottom diagonal indicating the reliable classification and the other indicating classification errors.

3.2. Model performance on the MUSE signals

The performance metrics, such as overall accuracy, precision, recall and f_1 -score, of the multilayer bidirectional recurrent models (bidirectional-RNN, bidirectional-GRU and bidirectional-LSTM)

on MUSE signals are summarized in detail in table 4. The evaluation performance accuracy of these models is generally between 94.14% and 96.18%, and the bidirectional-LSTM and the bidirectional-GRU models yield similar results and, in comparison, better results than the bidirectional-RNN model.

Figures 11–13 depict the confusion matrices of the multilayer bidirectional-RNN, bidirectional-GRU and bidirectional-LSTM of the multiclass classification tasks for each of the 10 imagined speech digit signals of MUSE. Based on the confusion matrices, the models bidirectional-GRU and bidirectional-LSTM accelerated the classification ability of all 10 imagined speech digits.

4. Discussion

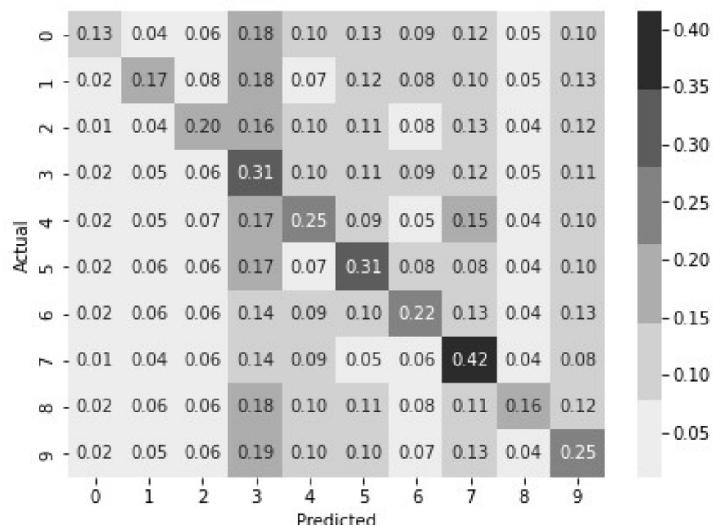
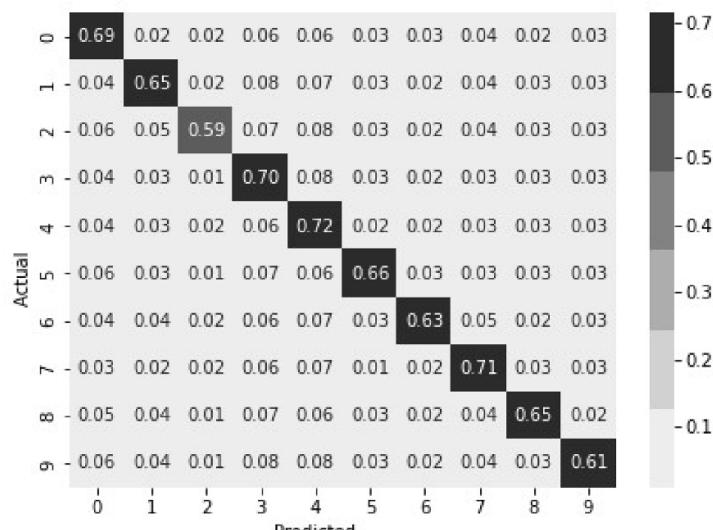
In this experiment, the methods used were the EEG feature extraction technique DWT and the deep learning model's bidirectional RNN. In order to extract the characteristics of the low-frequency signals from the decomposed sub-bands, the wavelet transformation is performed using the Daubechies-4 mother wavelet. The DWT extracted time-frequency features of the imagined digit EEG signals, which were fed into the deep learning models. The bidirectional RNN's effective classification accuracy in differentiating between imagined digits suggested that the DWT coefficient obtained the appropriate features for describing the EEG signals.

Our experiments indicated that the bidirectional-LSTM performs better than the bidirectional-RNNs and bidirectional-GRU models on the imagined digit classification task on EPOC and MUSE. All models were trained using the same processed signals and the same optimization approach, with all other model parameters remaining constant. Based on our findings, the benefits appear to be related to the architectural advantages and the capacity of bidirectional-LSTM to bridge lengthy time delays, as well as the use of bidirectional training to interpret reverse temporal relationships. Furthermore, the advantages of LSTM outweigh those of RNN and GRU, owing to RNN's and GRU's limited range of time dependencies, which preclude the use of the signal's long sequence information.

Since neural signals are time-series signals, RNNs could exploit the temporal aspects of EEG signal feature representation. Moreover, the stacked frameworks have enhanced decoding performance because bidirectional recurrent layers aggregate high temporal information regarding both past and future signal sampling points. In addition, the number of higher layers used in stacked bidirectional RNN topologies gives extensive differentiating feature information about the signals, which contributes to effective classification results.

Table 3. Evaluation of classification results of the EPOC for all the models.

Speech class	Bidirectional-RNN			Bidirectional-GRU			Bidirectional-LSTM		
	Pr.	Re.	f1	Pr.	Re.	f1	Pr.	Re.	f1
0	0.44	0.13	0.20	0.62	0.69	0.65	0.69	0.73	0.71
1	0.27	0.17	0.21	0.69	0.65	0.67	0.71	0.71	0.71
2	0.27	0.20	0.23	0.81	0.59	0.68	0.71	0.72	0.72
3	0.18	0.31	0.22	0.54	0.70	0.61	0.75	0.69	0.72
4	0.24	0.25	0.25	0.53	0.72	0.61	0.70	0.71	0.70
5	0.25	0.31	0.28	0.73	0.66	0.69	0.78	0.71	0.74
6	0.24	0.22	0.23	0.75	0.63	0.68	0.66	0.75	0.70
7	0.26	0.42	0.32	0.66	0.71	0.68	0.74	0.73	0.74
8	0.30	0.16	0.21	0.73	0.65	0.69	0.74	0.70	0.72
9	0.20	0.25	0.22	0.70	0.61	0.65	0.70	0.71	0.70
Overall accuracy	24.18%			65.90%			71.60%		

**Figure 8.** The confusion matrix report for the multilayer bidirectional-RNN model of the EPOC dataset.**Figure 9.** The confusion matrix report for the multilayer bidirectional-GRU model of the EPOC dataset.

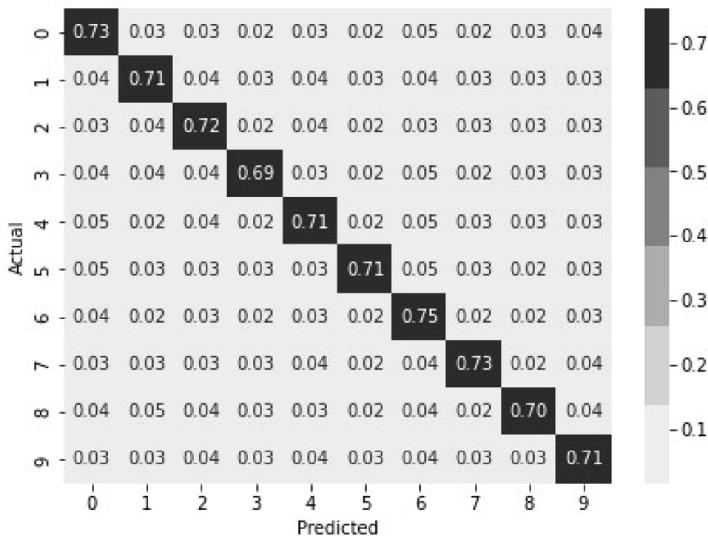


Figure 10. The confusion matrix report for the multilayer bidirectional-LSTM model of the EPOC dataset.

Table 4. Summary of the classification results of all models for the MUSE signals.

Speech class	Bidirectional-RNN			Bidirectional-GRU			Bidirectional-LSTM		
	Pr.	Re.	f1	Pr.	Re.	f1	Pr.	Re.	f1
0	0.96	0.93	0.94	0.96	0.96	0.96	0.97	0.96	0.97
1	0.94	0.94	0.94	0.95	0.96	0.95	0.97	0.95	0.96
2	0.93	0.95	0.94	0.94	0.97	0.95	0.96	0.97	0.97
3	0.95	0.93	0.94	0.97	0.95	0.96	0.97	0.97	0.97
4	0.91	0.95	0.93	0.94	0.96	0.95	0.95	0.96	0.96
5	0.95	0.94	0.94	0.96	0.95	0.95	0.95	0.95	0.95
6	0.93	0.96	0.95	0.96	0.96	0.96	0.97	0.96	0.97
7	0.94	0.95	0.94	0.96	0.96	0.96	0.96	0.96	0.96
8	0.93	0.94	0.93	0.96	0.95	0.95	0.96	0.98	0.97
9	0.97	0.94	0.95	0.97	0.97	0.97	0.95	0.96	0.96
Overall accuracy	94.14%			95.77%			96.18%		

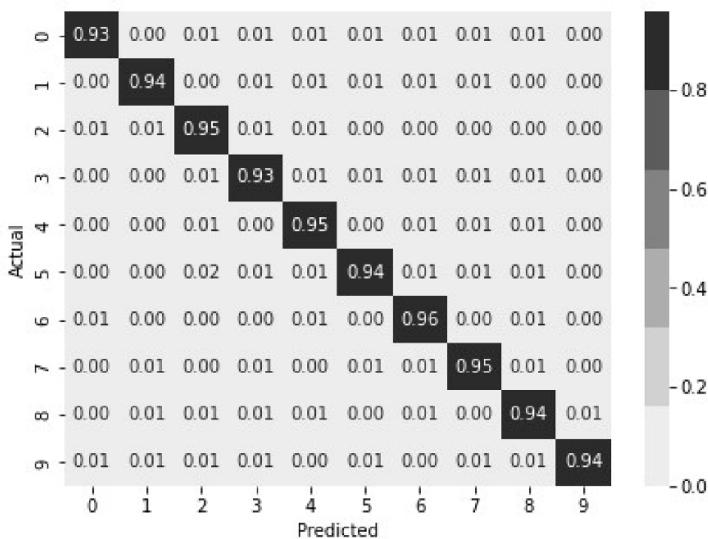


Figure 11. The confusion matrix report for the multilayer bidirectional-RNN model of the MUSE dataset.

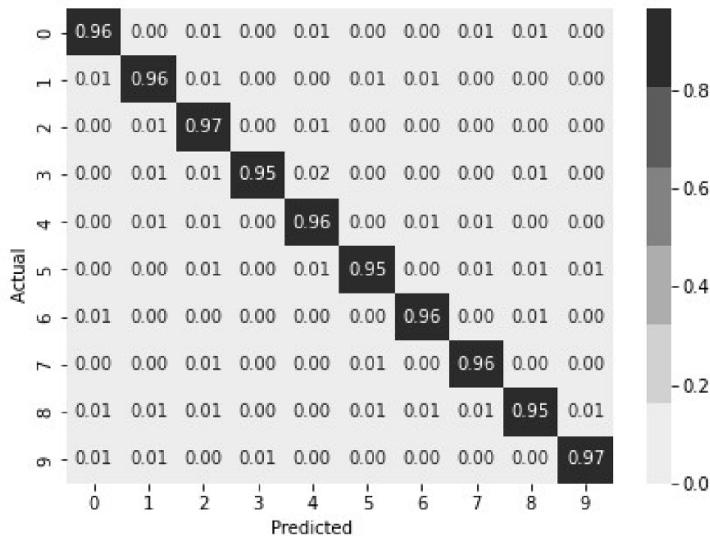


Figure 12. The confusion matrix report for the multilayer bidirectional-GRU model of the MUSE dataset.

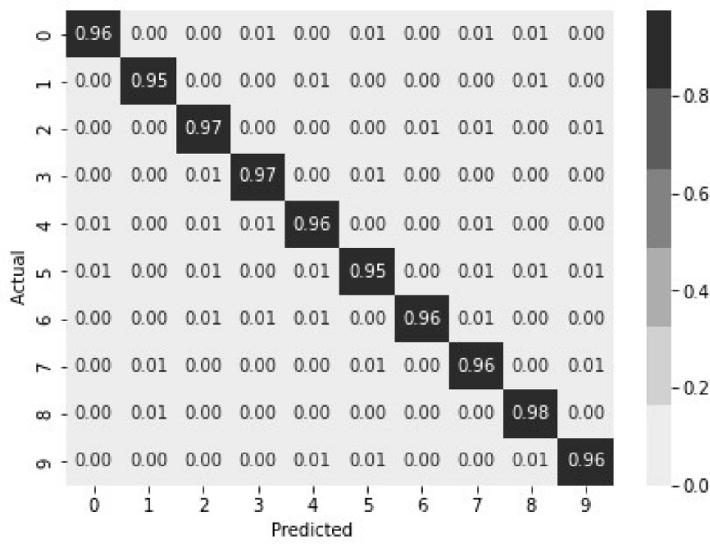


Figure 13. The confusion matrix report for the multilayer bidirectional-LSTM model of the MUSE dataset.

4.1. Comparison of performance on the EPOC and MUSE

A comparison of the imagined digit multiclass classification performance of the three models on the EPOC and MUSE signals in terms of evaluation accuracy is shown in figure 14. In contrast to the similarities in imagined digit multiclass classification performance that are shown by the three models on the four-channel MUSE findings, the results of the three models on the 14-channel EPOC yield a different performance. In terms of classification performance, the MUSE outperforms the EPOC. The MUSE could have higher-quality recordings than the EPOC or EEG channels could be employed in more advantageous locations. Moreover, this could be the consequence of differences in EEG signals due to

research instruments, such as EPOC, which has 14 channels, and MUSE, which has four, as well as higher sampling rates. The multilayer bidirectional-GRU and multilayer bidirectional-LSTM performed better on MUSE than the multilayer bidirectional-RNN, which has the shortcoming of disappearing and exploding gradients. Furthermore, the comparison is nonetheless constrained by the heterogeneity of the two datasets in terms of experimental settings and methods, such as the different number and placement of EEG channels.

4.2. Comparison with state-of-the-art performance

The results were compared to previous research in the field of imagined digits using MindBigData datasets to assess the efficacy of the proposed methods. Table 5

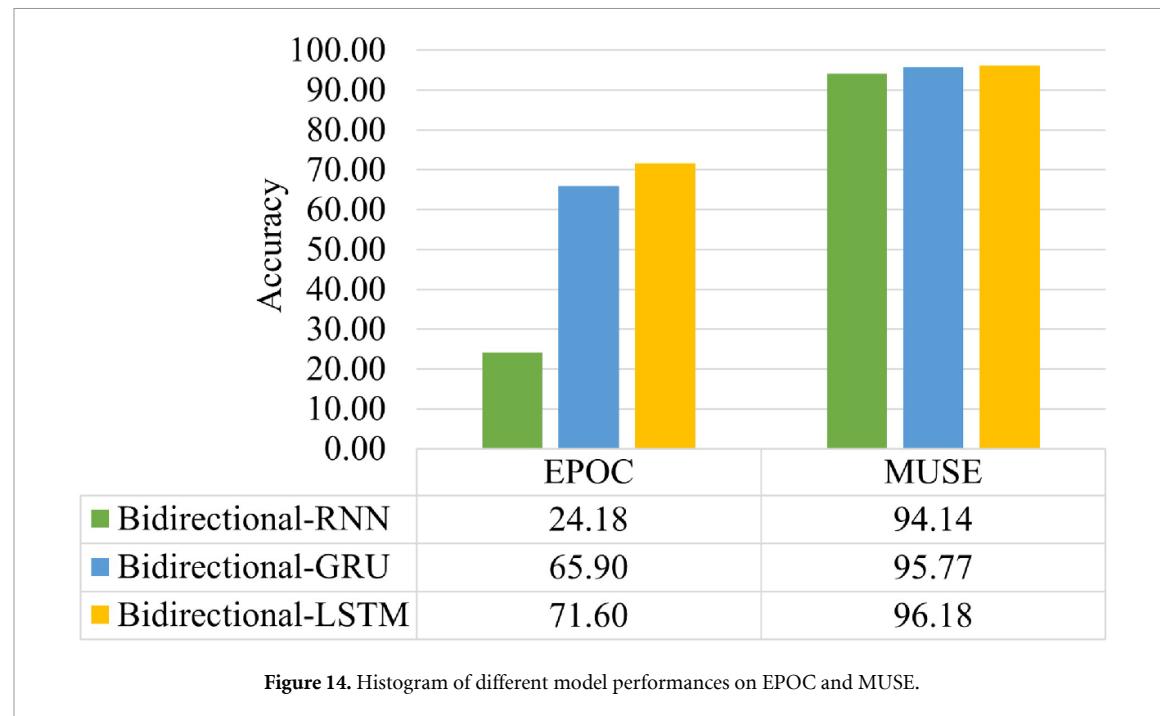


Figure 14. Histogram of different model performances on EPOC and MUSE.

Table 5. Comparison of the classification accuracy of the proposed and state-of-the-art methods.

Authors	EEG data	Methods	Accuracy (%)
Bird <i>et al</i> [14]	MUSE (4 channel)	Attribute selection + MLP	30
Jolly <i>et al</i> [13]	MUSE (4 channel)	Unidirectional-GRU	33.8
Pratama <i>et al</i> [28]	MUSE (4 channel)	FFT + k-NN	31
Mishra <i>et al</i> [12]	EPOC (14 channel)	Data refinement + 1D CNN	32.4
Mishra <i>et al</i> [12]	EPOC (4 channel)	Channel selection + 1D CNN	70.1
Proposed model#1	EPOC (14 channel)	DWT + bidirectional-LSTM	71.60
Proposed model#2	MUSE (4 channel)	DWT + bidirectional-LSTM	96.18

shows the performance comparison of state-of-the-art methods for the decoding of imagined digits from EEG signals. The comparison was challenging since each investigation used a different strategy for EEG signal processing prior to the classification model, although all the studies were based on MindBigData. Mishra *et al* [12] looked at multiple experiments based on data refinement with or without channel selection in the signal processing, whereas Bird *et al* [14] used attribute selection based on the covariance matrix. Pratama *et al* [28] used the fast Fourier transform technique for feature extraction. Jolly *et al* [13] used no signal preprocessing methods, and this experiment employed wavelet-based signal processing. In the classification of imagined digits for subject-dependent EEG signals, the proposed DWT method and multilayer bidirectional-LSTM model performed better.

4.3. Limitation of the study

Although the experiments show evidence of effectiveness in the decoding of imagined speech prompts focusing on numeric digits, which are represented as single words of vocabulary with a limited length of sequence, the model frameworks require

experience learning from multiword sequence sentences as imagined digits (e.g. 31-thirty one, 33-thirty three, 101-hundred one, 121-hundred twenty one, etc). As the length of the phrase sequence increases, a multilayer RNN architecture must be assessed for past-future temporal feature information extraction.

5. Conclusion

This paper proposes a methodology for decoding imagined speech from EEG signals of digits (0, 1, 2, 3, 4, 5, 6, 7, 8 and 9) that incorporates wavelet signal processing techniques as well as three different multilayer bidirectional RNNs. The EPOC and MUSE datasets from open-access MindBigData of EEG signals of imagined digits were utilized to train and evaluate the models. Among the models, the multilayer bidirectional-LSTM model improves decoding efficiency by increasing the retrieved information from EEG signals while being computationally expensive. Furthermore, the results highlight that the multilayer bidirectional-LSTM model learns the past-future temporal characteristics of the EEG signals. The experimental results suggest that the proposed technique is efficient and dependable.

Future research should concentrate on multiword imagined digits (such as short phrases of two to three words) from multiple subjects. EEG signal decoding is one of the most challenging problems to solve owing to the diversity of EEG equipment, the subject's health and cross-subject imagined speech. To further understand the influence of past–future temporal features on signal decoding, future research should emphasize cross-subject EEG signal decoding of short phrases employing effective feature selection approaches and RNNs.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <http://www.mindbigdata.com/opendb/index.html>.

ORCID iD

Nrushsingh Charan Mahapatra 
<https://orcid.org/0000-0001-9166-5451>

References

- [1] Suppes P, Lu Z L and Han B 1997 Brain wave recognition of words *Proc. Natl Acad. Sci.* **94** 14965–9
- [2] Schultz T, Wand M, Hueber T, Krusienski D J, Herff C and Brumberg J S 2017 Biosignal-based spoken communication: a survey *IEEE/ACM Trans. Audio Speech Lang. Process.* **25** 2257–71
- [3] Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain–computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–91
- [4] Anumanchipalli G K, Chartier J and Chang E F 2019 Speech synthesis from neural decoding of spoken sentences *Nature* **568** 493–8
- [5] Brigham K and Kumar B V K V 2010 Imagined speech classification with EEG signals for silent communication: a preliminary investigation into synthetic telepathy *2010 4th Int. Conf. on Bioinformatics and Biomedical Engineering* (Chengdu: IEEE) pp 1–4
- [6] Deng S, Srinivasan R, Lappas T and D'Zmura M 2010 EEG classification of imagined syllable rhythm using Hilbert spectrum methods *J. Neural Eng.* **7** 046006
- [7] DaSalla C S, Kambara H, Koike Y and Sato M 2009 Spatial filtering and single-trial classification of EEG during vowel speech imagery *Proc. 3rd Int. Convention on Rehabilitation Engineering and Assistive Technology - ICRAETE '09* (Singapore: ACM Press) p 1
- [8] Kim J, Lee S K and Lee B 2014 EEG classification in a single-trial basis for vowel speech perception using multivariate empirical mode decomposition *J. Neural Eng.* **11** 036010
- [9] Min B, Kim J, Park H j and Lee B 2016 Vowel imagery decoding toward silent speech BCI using extreme learning machine with electroencephalogram *BioMed Res. Int.* **2016** 1–11
- [10] Wang L, Zhang X, Zhong X and Zhang Y 2013 Analysis and classification of speech imagery EEG for BCI *Biomed. Signal Process. Control* **8** 901–8
- [11] Kumar P, Saini R, Roy P P, Sahu P K and Dogra D P 2018 Envisioned speech recognition using EEG sensors *Pers. Ubiquitous Comput.* **22** 185–99
- [12] Mishra R, Sharma K and Bhavsar A 2021 Visual brain decoding for short duration EEG signals *2021 29th European Signal Processing Conf. (EUSIPCO)* (Dublin: IEEE) pp 1226–30
- [13] Jolly B L K, Aggrawal P, Nath S S, Gupta V, Grover M S and Shah R R 2019 Universal EEG encoder for learning diverse intelligent tasks *2019 IEEE 5th Int. Conf. on Multimedia Big Data (BigMM)* (Singapore: IEEE) pp 213–8
- [14] Bird J J, Faria D R, Manso L J, Ekárt A and Buckingham C D 2019 A deep evolutionary approach to bioinspired classifier optimisation for brain-machine interaction *Complexity* **2019** 1–14
- [15] Qureshi M N I, Min B, Park H J, Cho D, Choi W and Lee B 2018 Multiclass classification of word imagination speech with hybrid connectivity features *IEEE Trans. Bio-Med. Eng.* **65** 2168–77
- [16] González-Castañeda E F, Torres-García A A, Reyes-García C A and Villaseñor-Pineda L 2017 Sonification and textification: proposing methods for classifying unspoken words from EEG signals *Biomed. Signal Process. Control* **37** 82–91
- [17] Sereshkeh A R, Trott R, Bricout A and Chau T 2017 EEG classification of covert speech using regularized neural networks *IEEE/ACM Trans. Audio Speech Lang. Process.* **25** 2292–300
- [18] Cooney C, Folli R and Coyle D 2019 Optimizing layers improves CNN generalization and transfer learning for imagined speech decoding from EEG *2019 IEEE Int. Conf. on Systems, Man and Cybernetics (SMC)* (Bari: IEEE) pp 1311–6
- [19] Berg B v d, Donkelaar S v and Alimardani M 2021 Inner speech classification using EEG signals: a deep learning approach *2021 IEEE 2nd Int. Conf. on Human-Machine Systems (ICHMS)* (Magdeburg: IEEE) pp 1–4
- [20] Parhi M and Tewfik A H 2021 Classifying imaginary vowels from frontal lobe EEG via deep learning *2020 28th European Signal Processing Conf. (EUSIPCO)* (Amsterdam: IEEE) pp 1195–9
- [21] Rusnac A L and Grigore O 2021 Convolutional neural network applied in EEG imagined phoneme recognition system *2021 12th Int. Symp. on Advanced Topics in Electrical Engineering (ATEE)* (Bucharest: IEEE) pp 1–4
- [22] Sarmiento L C, Villamizar S, López O, Collazos A C, Sarmiento J and Rodríguez J B 2021 Recognition of EEG signals from imagined vowels using deep learning methods *Sensors* **21** 6503
- [23] Lee D H, Kim S J and Lee K W 2021 Decoding high-level imagined speech using attention-based deep neural networks (arXiv:2112.06922)
- [24] Singh A and Gumaste A 2021 Decoding imagined speech and computer control using brain waves *J. Neurosci. Methods* **358** 109196
- [25] Vorontsova D *et al* 2021 Silent EEG-speech recognition using convolutional and recurrent neural network with 85% accuracy of 9 words classification *Sensors* **21** 6744
- [26] Lee D Y, Lee M and Lee S W 2021 Decoding imagined speech based on deep metric learning for intuitive BCI communication *IEEE Trans. Neural Syst. Rehabil. Eng.* **29** 1363–74
- [27] Rusnac A L and Grigore O 2022 CNN architectures and feature extraction methods for EEG imaginary speech recognition *Sensors* **22** 4679
- [28] Pratama I W P, Kesiman M W A and Gunadi I G A 2021 Frequency band and PCA feature comparison for EEG signal classification *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi* **12** 1
- [29] Graves A, Mohamed A r and Hinton G 2013 Speech recognition with deep recurrent neural networks *2013 IEEE Int. Conf. on Acoustics, Speech and Signal Processing* (Vancouver, BC: IEEE) pp 6645–9
- [30] Miao Y, Gowayyed M and Metze F 2015 EESEN: end-to-end speech recognition using deep RNN models and WFST-based decoding *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)* (Scottsdale, AZ: IEEE) pp 167–74

- [31] Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H and Bengio Y 2014 Learning phrase representations using RNN encoder-decoder for statistical machine translation (arXiv:[1406.1078](#))
- [32] Huang Z, Xu W and Yu K 2015 Bidirectional LSTM-CRF models for sequence tagging (arXiv:[1508.01991](#))
- [33] Zhou C, Sun C, Liu Z and Lau F C M 2015 A C-LSTM neural network for text classification (arXiv:[1511.08630](#))
- [34] Vivancos D 2018 MindBigData the MNIST of brain digits open database place: internet (available at: www.mindbigdata.com/opendb/index.html)
- [35] Klem G H, Lüders H O, Jasper H H and Elger C 1999 The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology *Electroencephalogr. Clin. Neurophysiol. Suppl.* **52** 3–6 (available at: <https://media.journals.elsevier.com/content/files/clinph-chapter11-14082757.pdf>)
- [36] Rosso O, Martin M, Figliola A, Keller K and Plastino A 2006 EEG analysis using wavelet-based information tools *J. Neurosci. Methods* **153** 163–82
- [37] Hazarika N, Chen J Z, Tsoi A C and Sergejew A 1997 Classification of EEG signals using the wavelet transform *Signal Process.* **59** 61–72
- [38] Ji M and Dong Z 2019 EEG signals feature extraction based on DWT and EMD combined with approximate entropy *Brain Sci.* **9** 201
- [39] El-Gindy S A E, Hamad A, El-Shafai W, Khalaf A A M, El-Dolil S M, Taha T E, El-Fishawy A S, Alotaiby T N, Alshebeili S A and El-Samie F E A 2021 Efficient communication and EEG signal classification in wavelet domain for epilepsy patients *J. Ambient Intell. Humaniz. Comput.* **12** 9193–208
- [40] Deng X, Zhang B, Liu K, Wang J, Yang P and Hu C 2021 The classification of motor imagery EEG signals based on the time-frequency-spatial feature 2021 IEEE 10th Data Driven Control and Learning Systems Conf. (DDCLS) (Suzhou: IEEE) pp 6–11
- [41] Agarwal P and Kumar S 2022 Electroencephalography based imagined alphabets classification using spatial and time-domain features *Int. J. Imaging Syst. Technol.* **32** 111–22
- [42] Hochreiter S and Schmidhuber J 1997 Long short-term memory *Neural Comput.* **9** 1735–80
- [43] Sak H, Senior A and Beaufays F 2014 Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition (arXiv:[1402.1128](#))
- [44] Schuster M and Paliwal K K 1997 Bidirectional recurrent neural networks *IEEE Trans. Signal Process.* **45** 2673–81
- [45] Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R 2014 Dropout: a simple way to prevent neural networks from overfitting *J. Mach. Learn. Res.* **15** 1929–58 (available at: <https://www.jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf>)
- [46] Clevert D A, Unterthiner T and Hochreiter S 2016 Fast and accurate deep network learning by exponential linear units (ELUs) (arXiv:[1511.07289](#))
- [47] Chollet F *et al* 2015 Keras (available at: <https://keras.io>)
- [48] Abadi M *et al* 2015 TensorFlow: large-scale machine learning on heterogeneous distributed systems 19 (available at: www.tensorflow.org/)
- [49] Kingma D P and Ba J 2017 Adam: a method for stochastic optimization (arXiv:[1412.6980](#))
- [50] Bergstra J and Bengio Y 2012 Random search for hyper-parameter optimization *J. Mach. Learn. Res.* **13** 281–305 (available at: <https://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf>)