



RESEARCH ARTICLE

Vowel speech recognition from rat electroencephalography using long short-term memory neural network

Jinsil Ham¹ , Hyun-Joon Yoo² , Jongin Kim³, Boreom Lee¹ *

1 Department of Biomedical Science and Engineering (BMSE), Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea, **2** Department of Physical Medicine and Rehabilitation, Korea University Anam Hospital, Korea University College of Medicine, Seoul, South Korea, **3** Deepmedi Research Institute of Technology, Deepmedi Inc., Seoul, South Korea

 These authors contributed equally to this work.

* leebr@gist.ac.kr

 OPEN ACCESS

Citation: Ham J, Yoo H-J, Kim J, Lee B (2022) Vowel speech recognition from rat electroencephalography using long short-term memory neural network. PLoS ONE 17(6): e0270405. <https://doi.org/10.1371/journal.pone.0270405>

Editor: Sidarta Ribeiro, Federal University of Rio Grande: Universidade Federal do Rio Grande, BRAZIL

Received: December 29, 2021

Accepted: June 9, 2022

Published: June 23, 2022

Copyright: © 2022 Ham et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting Information](#) files.

Funding: This work was supported by Basic Science Research program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (No. 2020R1A2B5B01002297). This work was also supported by GIST Research Institute (GRI) IIBR grant funded by the GIST in 2022. The funders had

Abstract

Over the years, considerable research has been conducted to investigate the mechanisms of speech perception and recognition. Electroencephalography (EEG) is a powerful tool for identifying brain activity; therefore, it has been widely used to determine the neural basis of speech recognition. In particular, for the classification of speech recognition, deep learning-based approaches are in the spotlight because they can automatically learn and extract representative features through end-to-end learning. This study aimed to identify particular components that are potentially related to phoneme representation in the rat brain and to discriminate brain activity for each vowel stimulus on a single-trial basis using a bidirectional long short-term memory (BiLSTM) network and classical machine learning methods. Nineteen male Sprague-Dawley rats subjected to microelectrode implantation surgery to record EEG signals from the bilateral anterior auditory fields were used. Five different vowel speech stimuli were chosen, /a/, /e/, /i/, /o/, and /u/, which have highly different formant frequencies. EEG recorded under randomly given vowel stimuli was minimally preprocessed and normalized by a z-score transformation to be used as input for the classification of speech recognition. The BiLSTM network showed the best performance among the classifiers by achieving an overall accuracy, f1-score, and Cohen's κ values of 75.18%, 0.75, and 0.68, respectively, using a 10-fold cross-validation approach. These results indicate that LSTM layers can effectively model sequential data, such as EEG; hence, informative features can be derived through BiLSTM trained with end-to-end learning without any additional hand-crafted feature extraction methods.

Introduction

Speech carries vast amounts of information to the brain, and it is one of the typical features of the brain to recognize and categorize the sounds of behaving animals. Given its importance, attempts to investigate the mechanisms of speech sound recognition have been conducted for

no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

over 100 years. One of the first neurolinguistic study of speech recognition was conducted through an observational study in the 1870s by a German neuropsychiatrist who found the crucial role of the superior temporal gyrus in speech perception, deducing that deficits in speech recognition were associated with damage to the left superior temporal gyrus [1]. It is now known that speech recognition relies predominantly on the dorsolateral temporal lobes, including the superior temporal gyrus, which contains the primary auditory cortex (A1) and anterior auditory field (AAF) [2]. Although the manner phonemes are encoded and interpreted in the brain remains controversial, it has been widely accepted that the recognition of sound is categorical. That is, discrimination is better for stimuli belonging to different phonetic categories than for stimuli belonging to the same category, even if the acoustic differences are equivalent [3, 4]. Not only humans, but also animals' perceptual systems sort continuously varying sound stimuli into a set of discrete categories [5].

With the advances in neurophysiological studies, electroencephalography (EEG) has been widely used in research involving neuroscience and neural engineering [6]. The high temporal resolution and sensitivity to different functional brain states make EEG a powerful tool to investigate real-time brain activity, and there has been increasing interest in illuminating the neural basis for categorical perception. Traditionally, EEG signals are recorded non-invasively from scalp in human study. At the level of sound or speech perception, mismatch negativity (MMN), a component of auditory evoked potential (AEP), which is elicited by oddball sounds, is widely used to study neural correlates of categorical perception [7, 8]. Naatanen et al. found evidence for language-dependent vowel representations in the human brain [9]. Another study examined the categorical perception of lexical tones and found that across-category contrast elicited a larger MMN than within-category distinction [10]. In animal experiments, more accurate EEG signals were obtained through invasive procedures. For instance, neural correlates of categorical perception and neural representations of various sounds have been studied using extra-cellular recording of action potential. Striatum-projecting neurons of song birds display categorical auditory responses and are highly sensitive to changes in note duration [11]. In addition, Kilgard et al. studied distinct neural representations of consonant and vowel sounds using intraparenchymal recording in the rat brain. Recording the multi- and single-unit responses from the inferior colliculus and A1, they suggested that the spike count encodes vowel sounds, while spike timing encodes consonant sounds [12, 13]. The effects of sound discrimination training in a rat model of autism were also investigated based on previous findings correlate neural responses to sound stimuli with sound perception ability [14]. Moreover, a recent study demonstrated that electrocorticography recorded with multi-channel array correlates with a passive exposure to a specific sound even in the auditory cortex of anesthetized rats [15].

Machine learning approaches have been used to make practical use of EEG in a wide variety of studies. Utilizing machine learning methods enables the investigation of rich information that is inherent and difficult to uncover from EEG signals [6]. Therefore, EEG-based classification can be performed in the following fields through conventional machine learning algorithms (e.g., support vector machine (SVM), k-nearest neighbors (KNN), and naive Bayes (NB)): motor imagery, emotion recognition, mental illness detection, event-related potential (ERP) detection, and so on [16, 17]. Furthermore, in recent years, owing to the increasing advances in graphic processing units and the availability of large dataset, it has become possible to conduct EEG-based classification using various deep learning networks [6, 18, 19]. Compared with conventional machine learning methods, deep learning networks are able to automatically detect and extract appropriate representations from input data [20, 21]. Hence, even with insufficient prior expert knowledge, promising results can be obtained through deep learning algorithms that do not require an additional handcrafted feature extraction process

[22, 23]. For example, in the field of speech, images, and video, the results were significantly improved by applying deep learning algorithms [24–26]. However, it is not clear whether such outperforming results always accompany the EEG-based classification domain when utilizing deep learning approaches instead of traditional machine learning methods [27]. Roy et al. showed that in most of the studies (excluding four out of 102 studies), the deep learning approach led to a higher performance than the traditional machine learning approach, and the highest improvement in accuracy was 35.3% [18, 28].

Furthermore, among the various fields of EEG-based classification studies, ERP classification studies are actively conducted by applying both conventional machine learning and deep learning methods. In an early study, the traditional grand averaging method was utilized to improve the low signal-to-noise ratio (SNR), one of the limitations of EEG signals, and to obtain ERP signals. In these studies, several ERP components were treated as feature sets for classification [29, 30]. In animal studies, the ERP features such as peak amplitude and latency are also used to discriminate ERP signals [31, 32]. However, single-trial EEG-based classification has also received much attention, since it is known that EEG data at the single-trial level possess more functional and rich information than the ERP signals obtained through the traditional grand averaging method [33, 34]. Therefore, in subsequent studies, features extracted by various algorithms such as wavelet-based algorithms [35], Gaussian mixture models [36], and spatial filtering [37] for classification using conventional machine learning methods [38, 39]. However, extracting the optimal hand-crafted features from the single-trial EEG is time-consuming and labor-intensive because additional processing steps must be executed. In this context, deep learning methods can alleviate this problem by allowing end-to-end learning. The most prevalent deep learning architecture is convolutional neural network (CNN), followed by recurrent neural network (RNN). The CNN is a special type of deep learning architecture widely used for single-trial EEG-based classification [6]. The CNN inputs are derived from raw or preprocessed EEG data, primarily in the following form: number of channels \times number of time points in a single trial. Moreover, considerable classification results have been demonstrated and it has been known to perform best when using spectrogram images as inputs [40–44]. In contrast to CNN, RNN is a highly preferred architecture, especially when handling sequential data (as in natural language processing applications) because the recurrent connection of RNN learning architecture makes it possible to utilize the previous information of the network recursively as the current input data [45]. Long short-term memory (LSTM) is a kind of RNN architecture proposed by Hochreiter and Schmidhuber to overcome the exploding and vanishing gradient problems of RNN [46]. Bidirectional LSTM (BiLSTM) is a further development of LSTM that combines the forward and backward hidden layers to access both the preceding and succeeding information. Although BiLSTM model is much complex and might need additional computational power, it is expected to solve the sequential modelling and classification task better than LSTM [47].

Previously we tried to classify EEG signals on a single-trial basis for three vowel sounds, /a/, /o/, and /u/, using machine learning techniques for the human brain. After the application of appropriate signal processing algorithms, including multivariate empirical mode decomposition (MEMD), the EEG responses were effectively classified according to each vowel sound using a linear discriminant analysis (LDA) classifier. From the time-frequency representation (TFR) of the EEG signals, it was also determined that the alpha band components were the most related neural responses of vowel sound perception [48]. However, due to the low SNR of human EEG signals, phoneme representation in the brain needs to be further assessed with a more invasive recording technique, allowing the acquisition of more reliable EEG signals. In addition, it is necessary to conduct further studies on the classification performance of each machine learning algorithm in classifying EEG responses to different phonemes.

The primary purpose of this study was to determine specific EEG components that might be related to speech representation in the rat brain to further illuminate brain responses to speech sound recognition. To acquire more accurate EEG signals, epidural EEG signals in response to auditory stimuli were recorded in AAF, which has been known to play an essential role in auditory perception and categorization [2]. In addition, this study tried to discriminate different brain responses for each speech sound on a single-trial basis using LSTM networks and other conventional machine learning techniques. It was hypothesized that the BiLSTM network would be appropriate for classifying EEG responses to vowel stimuli and would outperform other classical classifiers, because the network can perform robustly in modeling long-term dependencies of sequential data such as EEG. To the authors knowledge, LSTM networks have not been applied to the classification of EEG responses to auditory stimuli, and this is the first study to use a deep learning algorithm to analyze epidural EEG signals from AAF. Moreover, using the deep learning algorithm, EEG responses were classified to auditory stimuli using end-to-end learning with minimally preprocessed EEG signals with no additional feature extraction methods.

Materials and methods

Animals

The minimum required sample size was calculated to be 11 to 19, referring to previous animal studies that characterized neural responses to different human syllables [12, 13, 49]. Considering both scientific validity and animal ethics, a total of 19 male Sprague-Dawley rats (325–400 g, 11–13 weeks of age at the time of the experiment, Orient Bio Inc., Seongnam, Korea) were enrolled in the study. Only male rats were included in this study to avoid the potential effects of estrogen on EEG [50]. The animals were individually housed in standard plastic cages with free access to food and water and were maintained at a constant temperature ($21 \pm 1^\circ\text{C}$) with a 12 h light/dark cycle. All experimental protocols and procedures were approved by the Institutional Animal Care and Use Committee (IACUC) of the Gwangju Institute of Science and Technology (GIST). According to the committee, the study belonged to United States Department of Agriculture Category D; pain or distress was appropriately relieved with anesthetics, analgesics and/or tranquilizer drugs or other methods of relieving pain and distress. Therefore, all the surgical procedures and animal care were carried out in accordance with their guidelines to ensure minimal discomfort to the animals (approval number: GIST-2019-047).

Surgical procedures

All rats underwent microelectrode implantation surgery to acquire EEG signals in response to the speech sound stimuli. Before the surgery, the rats were anesthetized with isoflurane (5%) mixed with oxygen gas (0.6 L/min flow rate) in an induction chamber. Once the rats lost the righting reflex, they were moved into a stereotactic frame and applied an anesthetic nosecone. Isoflurane gas (maintenance dose of 1.5%) mixed with oxygen was redirected to the nosecone. Next, ear bars were inserted into the ear canals to fix the head. We then shaved the fur from the ears to just between the eyes. A line block with 2% lidocaine was performed on the scalp, and an incision was made to expose the skull. Next, the bilateral temporalis muscles were partly removed and durotomy was performed on each AAF with a dental drill to insert the epidural EEG electrodes. The electrode was a single micro-electrode that was custom-made using a micro-screw, silver wire, and a connector. The coordinates of the AAF were as follows: 4 mm posterior, 7.6 mm lateral, and 4 mm ventral to the bregma [51]. Finally, the implanted electrodes were connected to a multi-pin connector and fixed to the skull using bone cement. After completing all the surgical procedures, the rats were injected with an antibiotic (ceftazol

20 mg/kg, Guju Pharma Co, Korea) and an analgesic agent (ketoprofen 2.5 mg/kg, Uni Biotech, Korea) intramuscularly for three consecutive days. All animals were allowed to recover for a week and closely observed for any signs of pain such as reduced appetite, hunched posture, or piloerection.

Speech stimulation

Frequency information of speech sounds is known to be essential for categorical perception and recognition of different vowels [52]. In addition, components of AEP vary according to sounds with varying frequencies, and these different brain responses can be used to study sound recognition mechanisms [9]. Therefore, five different vowel speech sounds, /a/, /e/, /i/, /o/, and /u/, which have very distinct formant frequencies for each speech stimulus were chosen [53].

All speech stimuli were generated using a text-to-speech program provided by Google and the sound pitch was increased by one octave using the shiftPitch function in MATLAB 2017b (Mathworks, Inc., MA, USA) to accommodate the rat hearing range and applied root mean square normalization. The stimuli were delivered by a speaker (SRS-X88, SONY Co., Japan), which was located above one side of the cage, approximately 15 cm from the rat's head and the maximal intensity of the sound was calibrated to 60 dB SPL. The vowel speech sound was analyzed according to the time course, linear predictive coefficient (LPC) spectra, and spectrograms to verify that each stimulus has its own sound property (see Fig 1). Though the rat auditory system is not optimized for human vowel sound perception, we assumed that it is able to detect most of the sound stimuli since the frequency of sound belongs to the rat hearing range, that is, from 0.5 kHz to 64 kHz at 60 dB SPL [54].

Data acquisition

EEG signal responses to each vowel sound stimulus were acquired from the bilateral AAF after the one-week recovery period. First, the rats were anesthetized with isoflurane (5%) mixed

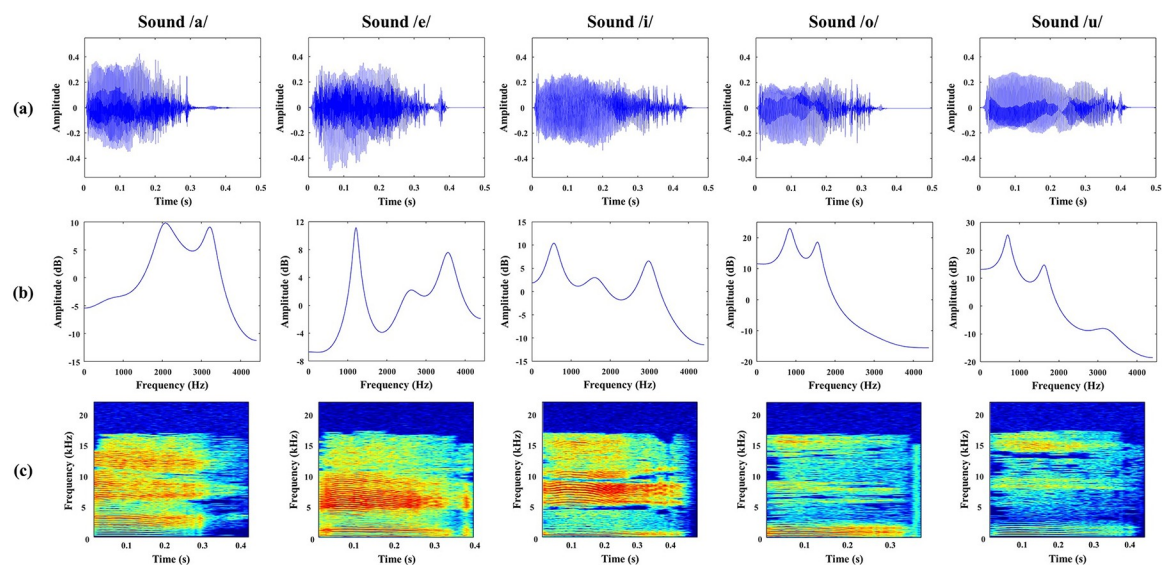


Fig 1. Characteristics of each vowel speech sound. (a) Time course, (b) linear predictive coefficient (LPC) spectra, and (c) the spectrogram of five vowel sounds used in this experiment. The peaks of LPC spectra refer to the formant frequencies of the sound stimuli. Vowel /a/ shows peaks at F1 = 651 Hz, F2 = 2034 Hz, F3 = 3234 Hz; vowel /e/ at F1 = 1211 Hz, F2 = 2559 Hz, F3 = 3570 Hz; vowel /i/ at F1 = 559 Hz, F2 = 1630 Hz, F3 = 2988 Hz; vowel /o/ at F1 = 845 Hz, F2 = 1564 Hz, F3 = 2921 Hz; vowel /u/ at F1 = 699 Hz, F2 = 1636 Hz, F3 = 3299 Hz.

<https://doi.org/10.1371/journal.pone.0270405.g001>

with oxygen gas (0.6 L/min flow rate) for the induction. After the rats lost the righting reflex, the anesthesia was maintained with isoflurane (1.5%) via nosecone during recording to prevent contamination of EEG signals from motion artifacts. Next, a multi-pin connector was connected to a recording device (g.USBamp and g.HEADstage, g.tec medical engineering GmbH, Graz, Austria), which acquired signals at a 1200 Hz sampling frequency. The epidural EEG recording was performed for 1500 s per session, during which the five vowel speech sounds were randomly presented to each rat through the experimental speaker. Each speech stimulus appeared 130–150 times per stimulus in one session. To obtain sufficient EEG data, the recording session was repeated for five consecutive days. All recordings were performed in a soundproof booth to maximize SNR. A schematic diagram of the experiment is shown in Fig 2.

EEG signal preprocessing and analysis

The acquired EEG signals were analyzed in response to each vowel sound using the FieldTrip toolbox [55] in MATLAB 2017b (Mathworks, Inc., MA, USA). In the first step, the raw EEG data were down sampled from 1200 to 250 Hz and band-pass filtered in the frequency range of 1 to 60 Hz. Then, the continuous EEG data were segmented into stimulus-specific trials with a 500 ms pre-stimulus period and 1500 ms post-stimulus period. Baseline correction was conducted based on the pre-stimulus period. To discard residual artifacts, contaminated trials were manually rejected using visual inspection methods.

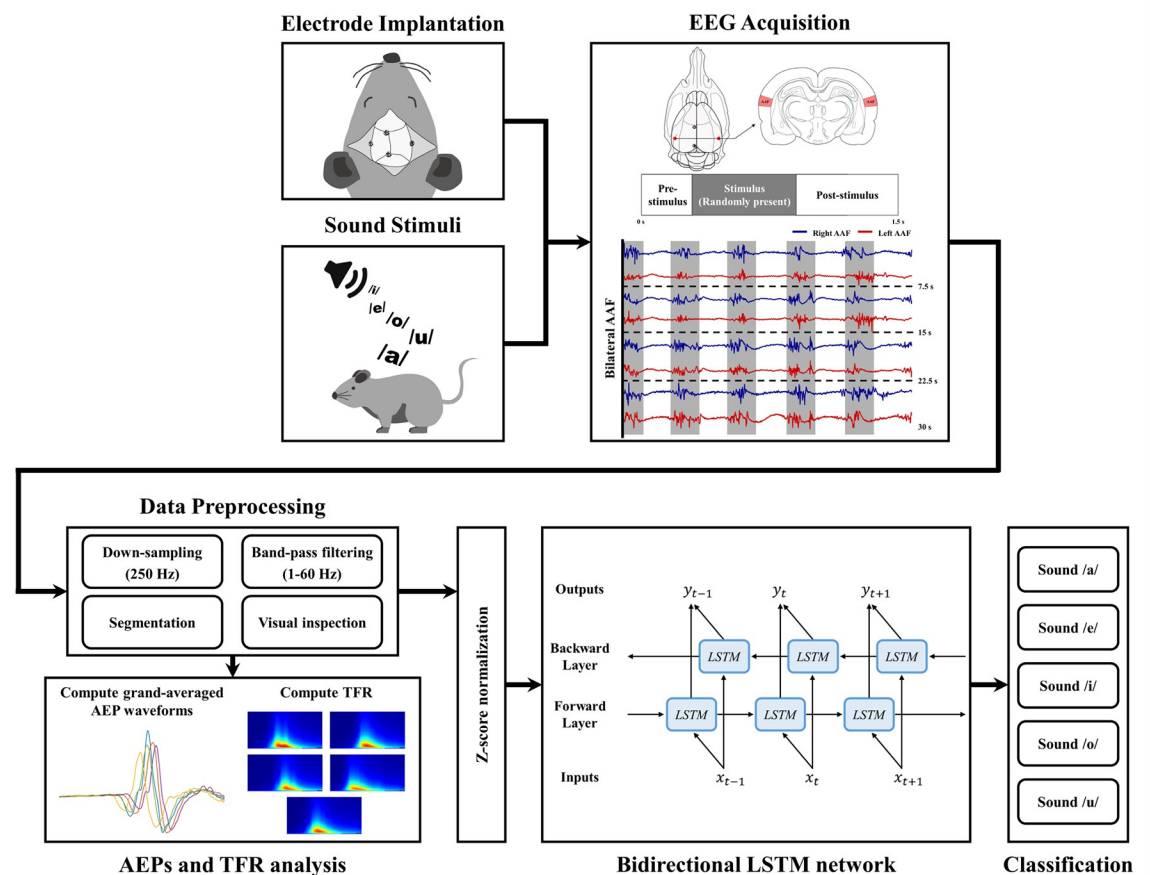


Fig 2. Schematic description of the experimental setup.

<https://doi.org/10.1371/journal.pone.0270405.g002>

After the pre-processing, the artefact-free EEG data were averaged for each speech stimulus to create the AEP waveforms and the TFR of the grand-averaged waveforms was calculated. TFR analysis was conducted based on Morlet wavelets to assess dynamic changes in spectral power over time for each speech stimulus. Utilizing grand-averaged AEP waveforms and their TFR, the time or frequency range that mainly reflected the brain response to speech stimulus was determined. In the case of TFR analysis, the analysis of variance (ANOVA) test and Bonferroni correction were performed to identify the statistical significance between the TFRs of EEG signals for each speech stimulus [48]. Through these results, the pre-processed EEG data was reorganized for later use for the purpose of classification. To ensure that the reconstructed data is meaningful, the time and frequency ranges of all EEG trials were restricted. The time range was set to 0.2–0.8 s and the frequency range was set to 1–60 Hz. After redefining the time and frequency ranges, all EEG trials were normalized using z-score normalization, a commonly used method to reduce variability among trials while maintaining a similar tendency within the trials [56, 57]. It is well known that the overall classification performance is improved following z-score normalization [56]. After this, the z-score normalized dataset was randomly shuffled and separated into a training set (90%) and test set (10%) to be used as inputs for deep learning and machine learning classifiers for speech recognition classification.

Bidirectional long short-term memory networks

LSTM is a special recurrent neural network (RNN) architecture that overcomes the vanishing/exploding gradient problem by incorporating gate structures that control the state of memory cells [46, 58]. For this reason, LSTM has shown stable and powerful performance for modeling long-term dependencies in a variety of temporal or sequential tasks [46, 58–61]. The structure of the LSTM is shown in Fig 3A. The main difference between conventional RNN and LSTM is the memory cell, c_t , which can preserve the state information which is modulated by three kinds of self-parameterized gates: the input gate i_t , forget gate f_t , and output gate o_t . The input gate i_t decides whether a new input will be accumulated in the memory cell; the forget gate f_t can discard the past status of the memory cell, c_{t-1} ; and the output gate o_t regulates the

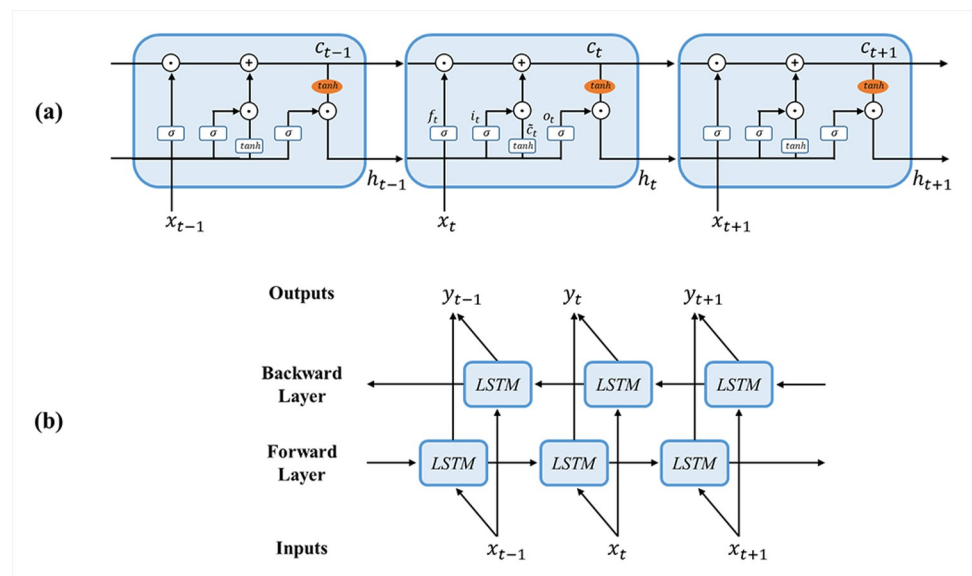


Fig 3. Structure of the BiLSTM network. (a) The structure of a long short-term memory (LSTM) cell and (b) architecture of bidirectional LSTM (BiLSTM) network.

<https://doi.org/10.1371/journal.pone.0270405.g003>

propagation of the output from the current memory cell c_t into the output response h_t . The key processing of LSTM is described by the following equations:

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

$$f_t = \sigma(W_f x_t + R_f h_{t-1} + b_f) \quad (2)$$

$$o_t = \sigma(W_o x_t + R_o h_{t-1} + b_o) \quad (3)$$

$$\tilde{c}_t = \tanh(W_c x_t + R_c h_{t-1} + b_c) \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (5)$$

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

where σ and \tanh are nonlinear activation functions. The logistic sigmoid function, defined as $\sigma(x) = 1/(1+e^{-x})$ is utilized as the gate activation function, and the hyperbolic tangent function, $\tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$, is used as the block input and output activation function. Element-wise multiplication of two vectors is denoted by \odot ; W , R represent the weight matrices, and b denotes the bias vector, which are learnable parameters that control each gate.

LSTM is attested as a powerful structure for handling sequential data [59]; however, the standard LSTM captures only the past information from the sequence in the forward direction. BiLSTM was implemented to improve the structure. BiLSTM is a type of LSTM version of a bidirectional RNN [47, 62]. It has two layers of LSTM, as shown in Fig 3B; one processes information in the forward direction while another processes it in the backward direction. By accessing both past and future information, these structures can capture rich information from a sequence. Hence, the existing literatures shows that BiLSTM performs better than the standard LSTM in classifying EEG signals according to each task [63–67].

In this study, the BiLSTM network was used to classify five different vowel speech sounds using single-trial basis EEG signals. A BiLSTM layer containing 600 LSTM units was set and to avoid overfitting, the dropout ratio on the LSTM layers was set to 0.3 [68]. After the LSTM layers, the hidden states were concatenated into the fully connected layer with a softmax activation function, used for multiclass classification. Categorical cross entropy was adopted as the loss function with the ADAM optimizer [69] and the initial learning rate and learning rate decay were set to 1e-3 and 1e-6, respectively. Furthermore, the model was trained with 500 epochs and a batch size of 64. The learning curve reached a stable plateau within 500 epochs.

These hyper-parameters were adjusted to best fit the model to the data. A stratified 10-fold cross-validation (10-CV) was used to evaluate model performance. The k-fold cross-validation is an effective method to test the success rate of models used for classification and $k = 10$ is generally considered as the most reasonable parameter in applied machine learning [70].

The model was implemented using the Keras library [71] with TensorFlow backend [72] and the Scikit-Learn library [73] in Python.

Machine learning classifiers

The performance of BiLSTM was compared with conventional machine learning classifiers: SVM with linear kernel (SVM_lin), SVM with radial basis function kernel (SVM_rbf), random forests (RF), NB, and KNN. SVM [74] aims to determine the optimally separated hyperplane by maximizing the margin, which is the distance between the support vectors. By using the

kernel trick, SVM is capable of mapping feature space from low to high dimensions; therefore, it can efficiently perform linear classification and non-linear classification. RF [75] operates by constructing multiple decision trees during the training phase and generating the final class that combines the results of each decision tree. NB [76, 77] is a probabilistic classifier based on Bayes' theorem and conditional probability which usually assume that all features are independent of each other. KNN [78] is a non-parametric approach that classifies the input based on the majority class of its k -nearest neighbors in the feature space. Usually, the k value is selected as an odd number to avoid tied classes. To train and evaluate the above machine learning models, the same 10-CV was used as in BiLSTM. All machine learning models were implemented using the Scikit-Learn library [73] in Python.

Statistical analyses

All statistical analyses were performed using SPSS software (SPSS version 20.0, SPSS Inc., Armonk, NY, USA) and MATLAB software version 2017b (Mathworks, Inc., MA, USA). The data was analyzed with parametric statistics since all the data in the study showed a normal distribution in the Shapiro–Wilk test ($p > 0.05$). ANOVA was used to analyze the statistical significance of the TFRs according to the different vowel stimuli. In addition, a repeated-measures ANOVA was conducted to compare the performance of each classifier. Subsequently, pairwise comparisons using paired t -tests were performed between the BiLSTM network and other classical machine-learning classifiers and a Bonferroni correction was performed to adjust for the type I error rate inflation. The statistical significance of the p -value was set at 0.01, when comparing the TFR of EEG responses, while the significance level of the p -value was set at 0.05, when comparing the performance between the BiLSTM network and other machine learning classifiers.

Results

Auditory evoked potentials in response to vowel sounds

A total of 19 Sprague-Dawley rats underwent epidural electrode implantation surgery, and all rats survived the surgical procedure. As a result, EEG responses to five English vowel sounds were recorded from 19 isoflurane-anesthetized rats. To extract the mean AEP waveforms, all the neural responses were averaged over the subjects for each stimulus. Fig 4 presents the averaged AEP waveforms for each vowel sound from bilateral AAF.

As expected, each categorical vowel sound evoked distinct neural activities in the bilateral AAF with varying peak amplitudes and latencies. The peak amplitude of AEPs, defined as the highest recorded voltage after the vowel stimuli, was smallest for /i/ (61.74 μV in left AAF and 61.27 μV in right AAF), while AEPs in response to /a/ showed the largest peak amplitudes (92.12 μV in left AAF and 90.18 μV in right AAF). The peak latency, defined as the duration from stimulus onset to the peak amplitude was approximately 0.39 s to 0.5 s, shortest in /i/ (0.39 s in left and right AAFs), and longest in the /o/ sound (0.51 s in left and right AAFs). As shown in Fig 4, similar AEP waveforms were observed from the left and right AAFs.

Time-frequency analysis of the EEG signals

Time-frequency analysis is a powerful method for analyzing nonstationary EEG signals over a time-frequency plane and is used to provide qualitative information for the classification of EEG [79, 80]. Therefore, the TFR of the grand-averaged EEG was calculated for each sound to identify vowel recognition-related changes in the magnitude and phase of EEG oscillations at specific frequencies (Fig 5A). From the TFR analysis, high power activation was observed

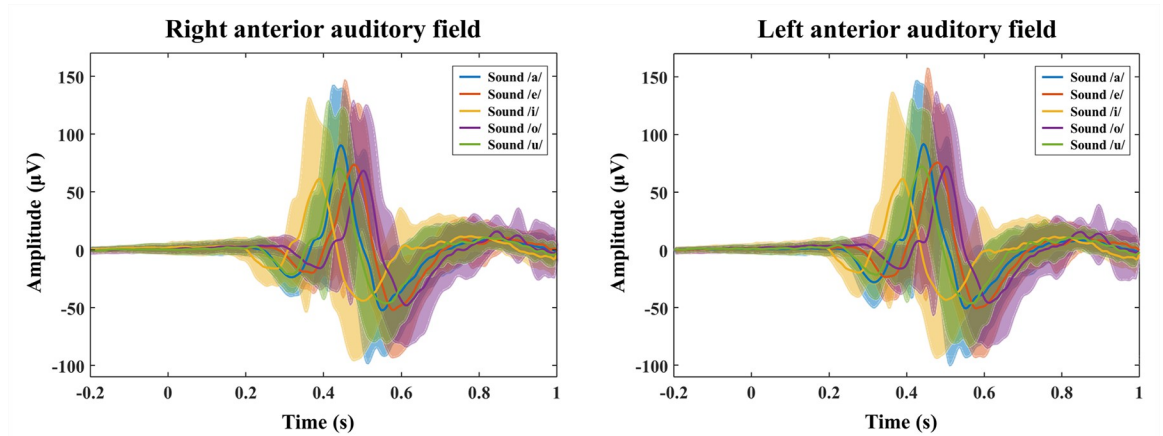


Fig 4. Averaged auditory evoked potentials (AEP) waveforms over the subjects to each vowel sound. AEPs were recorded on the right and left anterior auditory fields (AAFs). Overall, the neural responses were elicited 0.2–0.4 s after the sounds stimulus onset and showed different peak latencies and amplitudes depending on the vowel stimulus. The AEPs recorded on both AAFs were generally similar. The bold lines represent the averaged AEP waveforms, and the shaded areas represent the standard deviation.

<https://doi.org/10.1371/journal.pone.0270405.g004>

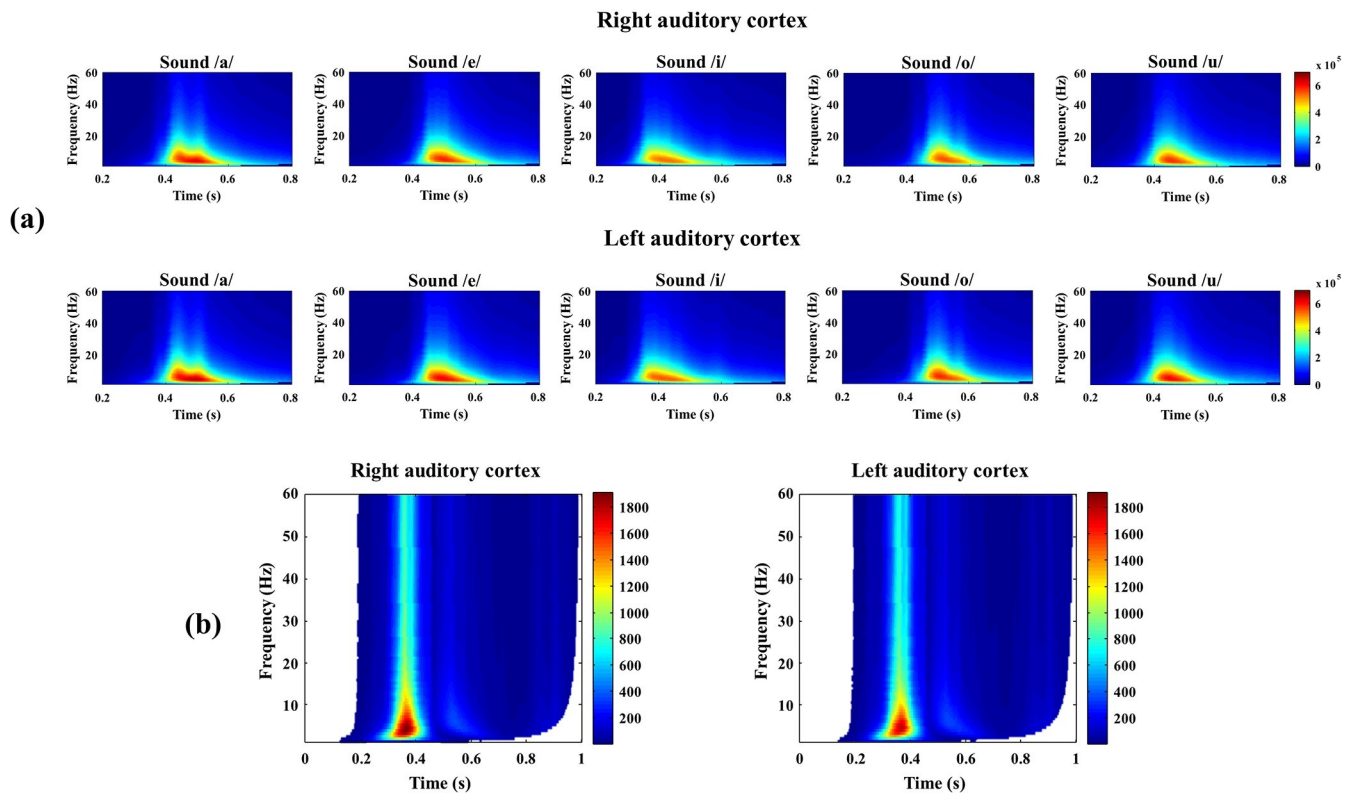


Fig 5. Subject-averaged time-frequency power results. (a) The time-frequency representation (TFR) of the event-related EEG signals (speech stimulation /a/, /e/, /i/, /o/, or /u/) over all subjects on the right and left anterior auditory fields. TFRs were plotted for the frequency range of 4 to 40 Hz, 0.2 to 0.8 s after each stimulus. One can observe high power activation in the low frequency band (especially in the delta, theta, and alpha band) between 0.3 and 0.6 s regardless of the sound stimulation. (b) Time-frequency regions with significant differences after the Bonferroni correction ($p < 0.01$ on the ANOVA test) are plotted. The color scale represents the F-values and non-significant regions are colored in white. Note that most of the brain responses between 0.2 and 0.8 s after the stimuli showed distinct neural responses across each vowel sound.

<https://doi.org/10.1371/journal.pone.0270405.g005>

around the delta (1–4 Hz), theta (4–8 Hz), and alpha (8–12 Hz) band at 0.3–0.6 s from the stimulus onset, regardless of the speech sound stimulation.

In addition, an ANOVA test with a Bonferroni correction was conducted to analyze the statistically significant TFR components according to each vowel stimulus. Subsequently, the power of statistically significant areas ($p < 0.01$) was represented by the F-value (Fig 5B). In the analysis, most of the EEG frequency band from 0.2–0.8 s was significantly different according to the vowel stimuli. In addition, part of the TFR from 0.8–1 s was also statistically different for each stimulus. Considering the AEP waveforms and the results of the ANOVA tests, it was inferred that the AEPs from 0.2–0.8 s after the vowel stimulus were the most informative neural responses and were related with the vowel sound recognition.

Model training and evaluation of the BiLSTM networks

Based on the results of Fig 5B, EEG data that were band-pass filtered between 1–60 Hz with a time window of 0.2–0.8 s were selected. Then, the z-scores of the selected EEG data were used as the input to the BiLSTM network. All EEG data were divided into 10 folds within each subject to evaluate the BiLSTM networks. Therefore, the test performance was obtained per fold using the trained model with the remaining folds in a 10-CV scheme. The performance of the network was evaluated using metrics of accuracy, f1-score, and Cohen's kappa statistic κ (Fig 6 and Table 1). The average five-class EEG discrimination accuracy of the BiLSTM network was $75.18 \pm 7.06\%$ and the f1-score was 0.74 ± 0.08 . Cohen's κ was 0.68 ± 0.09 , which was interpreted as a moderate agreement [81].

To analyze the performance of the BiLSTM network in more detail, the confusion matrix in Fig 7 was plotted. This indicated that many of the errors were due to the misclassification of the EEG responses to /u/ as /a/ and /e/ as /o/. However, the BiLSTM network classified most of

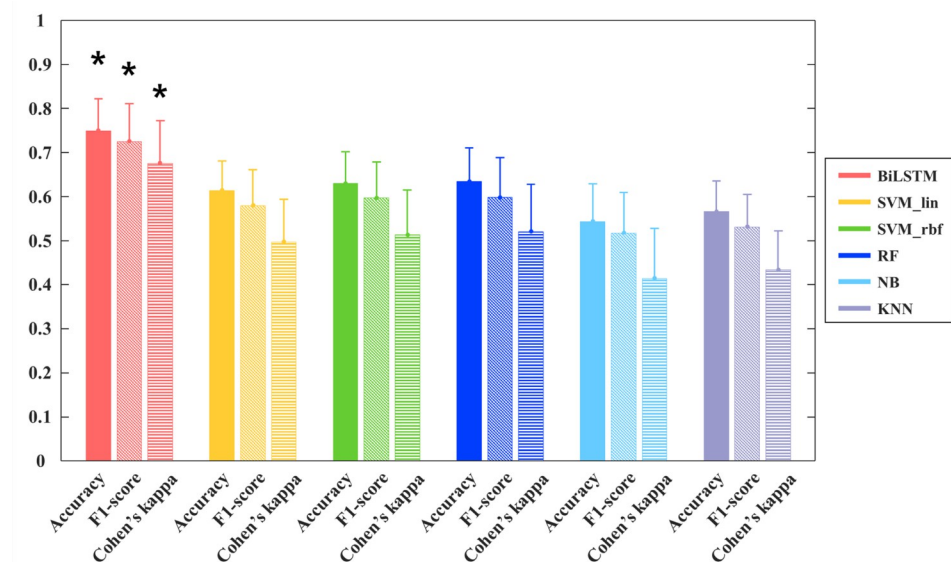


Fig 6. Comparison of the performance of the BiLSTM network and other conventional machine learning methods. The bar plots with standard deviation were drawn using the results of 10-fold cross-validation of each subject. Each bar represents accuracy (entire), f1-score (diagonal), and Cohen's kappa statistic κ (horizon) of each classifier. Asterisk (*) above the bar plot indicates the significant differences ($p < 0.01$) between the performance of the BiLSTM and all other conventional machine learning methods. BiLSTM, bidirectional long short-term memory; SVM_lin, support vector machine with linear kernel; SVM_rbf, support vector machine with radial basis function kernel; RF, random forests; NB, naïve Bayes; KNN, k-nearest neighbors.

<https://doi.org/10.1371/journal.pone.0270405.g006>

Table 1. Overall performance of the BiLSTM network and other conventional machine learning methods.

Classifier	Accuracy (%)	F1-score	Cohen's kappa (κ)
BiLSTM	75.18 \pm 7.06	0.74 \pm 0.08	0.68 \pm 0.09
SVM_lin	61.47 \pm 6.52	0.60 \pm 0.08	0.50 \pm 0.09
SVM_rbf	63.11 \pm 7.04	0.62 \pm 0.08	0.51 \pm 0.1
RF	63.21 \pm 7.41	0.62 \pm 0.09	0.52 \pm 0.1
NB	53.39 \pm 8.40	0.52 \pm 0.09	0.41 \pm 0.11
KNN	56.80 \pm 6.76	0.55 \pm 0.07	0.43 \pm 0.09

Data are presented as the mean \pm standard deviation. BiLSTM, bidirectional long short-term memory; SVM_lin, support vector machine with linear kernel; SVM_rbf, support vector machine with radial basis function kernel; RF, random forests; NB, naïve Bayes; KNN, k-nearest neighbors.

<https://doi.org/10.1371/journal.pone.0270405.t001>

the EEG responses with more than 50% accuracy, a high accuracy in the five-class EEG classification.

Comparison of the BiLSTM network with other machine learning methods

To validate the effectiveness of the BiLSTM networks in classifying EEG for vowel sound recognition, the results were compared with those of other conventional machine learning methods. Fig 6 and Table 1 show the performance of the machine learning classifiers. The RF demonstrated the highest classification accuracy among the conventional machine learning algorithms (accuracy: 63.21 \pm 7.41%, f1-score: 0.62 \pm 0.09, and Cohen's: 0.52 \pm 0.1). In the statistical analysis, the classification performance of RF was not significantly higher than that of SVM_lin and SVM_rbf, while it showed higher performance when compared with those of NB and KNN. However, when the performance of conventional machine learning algorithms, including RF, was compared with BiLSTM, it was obvious that the BiLSTM network was superior for all the metrics used in the study ($p < 0.01$).

In the confusion matrix, conventional machine-learning algorithms cannot discriminate certain EEG responses well. In particular, all the conventional machine learning algorithms had difficulty distinguishing the sound /u/. It was noted that the algorithms showed a tendency to misclassify sound /u/ as /a/ on average 30% of the time (25.96% in NB to 36.97% in KNN), resulting in a decrease in the overall classification performance (Fig 7).

Discussion

In this study, rat epidural EEG responses to five categorical vowel sounds (/a/, /e/, /i/, /o/, and /u/) were discriminated using the BiLSTM network. Five-class classifications of epidural EEG signals were performed on a single-trial basis, which is known to be challenging. To maximize learning performance, this study tried to determine specific EEG components that might be related to the recognition of speech sounds in the rat brain and utilized these EEG components as input features. As a result, a relatively high performance in classifying AEPs to five different vowel sounds was achieved using BiLSTM. A comparison of the classification performance of the BiLSTM network with other machine learning algorithms showed that the BiLSTM network outperformed other classical classifiers. These results indicate that the BiLSTM network trained with speech recognition-related EEG components reliably classifies AEPs to each categorical vowel sound with a high degree of accuracy. To our knowledge, LSTM networks have not been applied to the classification of EEG responses to auditory stimuli, and this is the first study to use a deep learning algorithm to analyze EEG signals from rat AAF.

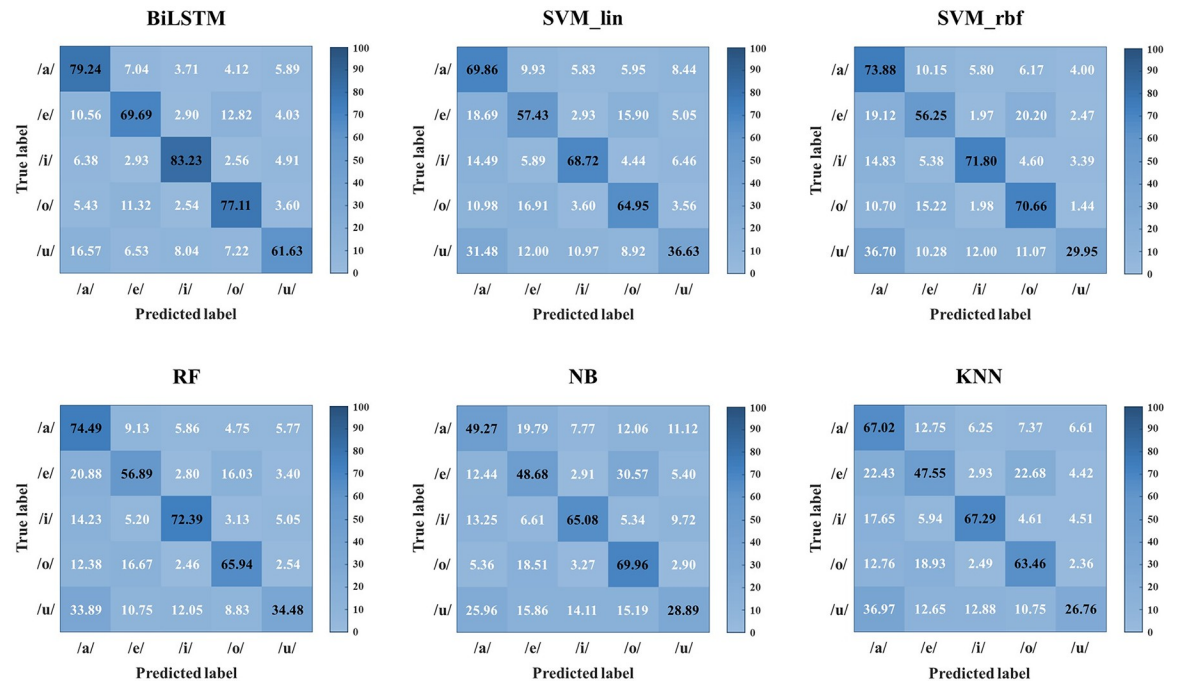


Fig 7. Confusion matrix for the BiLSTM network and other conventional machine learning classifiers. BiLSTM, bidirectional long short-term memory; SVM_lin, support vector machine with linear kernel; SVM_rbf, support vector machine with radial basis function kernel; RF, random forests; NB, naïve Bayes; KNN, k-nearest neighbors.

<https://doi.org/10.1371/journal.pone.0270405.g007>

Currently, only a few studies have used LSTM architecture to achieve state-of-the-art results in EEG-based classification. The LSTM architecture is suitable for EEG-based classification because its chain-like structure can capture the temporal sequence of EEG data [82]. In the beginning, research focused on improving the classification results through various LSTM architectures; however, the input features were still extracted manually, as in conventional machine learning methods [83, 84]. Tsiouris et al. evaluated the performance of diverse combinations of LSTM network elements in order to find the most efficient LSTM architectures for detecting epileptic seizures, thus obtaining near-perfect results in seizure prediction (100% sensitivity and 99.86% specificity) [83]. Because LSTM is a powerful structure for processing sequential data, there are also studies that use raw EEG data as input features with minimal preprocessing. As the LSTM network directly learns features from raw EEG data, the performance in emotion recognition studies improved by at least 12% [85], and the results of motor imagery classification studies also improved [86], when compared with other traditional feature extraction techniques. Moreover, the BiLSTM architecture was utilized for EEG-based classification because it can access information from both past and future states. Therefore, in detecting various brain states reflected in EEG data, such as seizure, sleep, etc. [63–67], the BiLSTM network generally outperformed the LSTM network that only captures past information from the sequence in the forward direction. For this reason, high performance has been reported in recent EEG-based classification using BiLSTM networks. Sharma et al. achieved 82.01% classification accuracy for four types of emotions based on the BiLSTM algorithm and higher-order statistics [87]. In addition, the BiLSTM networks successfully classified epilepsy types and sleep stages [88, 89].

Similar to previous studies, this study achieved comparatively good results using BiLSTM networks. The proposed algorithm successfully discriminated the EEG responses to five vowel sounds with high values of accuracy, f1-score, and Cohen’s κ of 75.18%, 74.43%, and 0.68,

respectively. The value of Cohen's κ for five-class classification is higher than that seen in most current studies [90]. As illustrated in Fig 6, the BiLSTM method produced the highest value for all metrics compared with the other machine learning methods. In addition, to determine the statistical difference in classification performance, repeated-measured ANOVA results were analyzed between BiLSTM and other classical machine learning methods using all the metric values. Through statistical analysis, it was determined that the classification performance of the BiLSTM network was significantly higher than that of other classical machine learning methods ($p < 0.01$). This result was also consistent with the confusion matrix. As shown in Fig 7, the BiLSTM network predicted the true labels of the five vowel sounds well, whereas classical machine learning methods did not. The prediction acquired through the conventional machine learning classifier was especially poor at classifying the /u/ sound; the /u/ sound was mainly misinterpreted as /a/. Even RF, which showed the best performance among the five conventional machine learning classifiers, had a classification rate of 34.48% for the /u/ sound, with a 33.89% misclassification rate of the /u/ sound as an /a/ sound. As can be seen in Fig 4, the /a/ and /u/ sounds had a similar peak latency, which is one of the main characteristics of AEP waveforms (peak latency of sound /a/: 0.448, peak latency of sound /u/: 0.444). When classification was performed based on minimally pre-processed single-trial EEG signals, it seems that such similarities could not be distinguished by conventional machine learning algorithms, whereas the BiLSTM network could distinguish them. Given that the BiLSTM network can simultaneously access all past and future contexts, rich information can be learned through this network. In addition, even though the features reflecting the characteristics of EEG responses to each vowel sound were extracted directly from the forward and backward directions of the LSTM layer, the classification performance was improved. In this study, we can derive good classification results using a simple BiLSTM architecture without an additional handcrafted feature extraction process.

Classifying the ERP responses to speech stimuli in a single trial is very challenging owing to the characteristics of the low SNR of EEG. Although one of the key advantages of the deep learning method is its ability to learn high-level features without hard-core feature extraction, we attempted to select the most relevant EEG signals related to speech recognition to achieve better performance. In this study, distinct AEP waveforms corresponding to each speech sound stimulus were observed with the high-power activation of the low-frequency band, including the delta, theta, and alpha bands, in the TFR analyses. Neural oscillations in the alpha band have been widely recognized to play an important role in auditory processing. Mazaheri et al. reported that the attenuation of alpha activity is closely related to the discrimination of auditory targets [91]. Staru et al. proved that cortical alpha oscillations are a pivotal mechanism for selectively inhibiting the processing of noise to improve the auditory selective attention toward target signals [92]. Previously, we also found that alpha power was highly activated in bilateral temporal areas after specific sound stimuli that were statistically different in terms of the type of sound [48]. In addition, the delta and theta bands are known to be associated with shaping the segmentation and perceptual influence of acoustic information [93]. Although this study is based on animal experimental data, similar speech-related components, as compared to the previous studies on human subjects, were observed in the TFR analyses. Moreover, in the statistical analysis, all the EEG bands were found to be significant within 1 s after the stimuli and represented the EEG components related to sound perception. These results were somewhat different from those of previous studies, suggesting that only specific EEG bands, such as the alpha band, were related to sound perception. It is expected that even subtle changes across all the EEG band activities are recorded through the epidural EEG recording, because it provides a higher SNR by reducing volume conduction and eliminating the artifacts that are inherent to extracranial EEG recordings.

In this study, the speech sound recognition related EEG components in rats were determined and the AEP components were successfully classified using the BiLSTM network. However, this study had some limitations. First, the number of subjects included was too small, especially for deep learning. Moreover, this study did not evaluate each classifier's performance with external validation, but instead used 10-CV to overcome the limited sample sizes. Besides, we cannot rule out the possibility that the rat's auditory system responds continuously to sound, since only a single utterance of each vowel sound was used in this study. In addition, the acquired EEG responses were affected by the anesthetic effects. Although minimal anesthetic dose was used, frequency slowing with increase in delta power is a typical finding of EEG changes after isoflurane inhalation [94]. Therefore, the vowel recognition EEG components suggested in this study may be different from EEG signals acquired from rats that are awake. However, we believe that the quality of the EEG signal is good enough since we recorded EEG through epidural electrode implantation, and it was not contaminated by motion artifacts.

Conclusions

In conclusion, this study extracted meaningful neural components related to categorical speech perception. Furthermore, based on the characteristics of the LSTM networks, it was proved that the BiLSTM network was suitable for classifying EEG responses with minimally pre-processed AEPs. Since this study is pioneer research with animal data, it may not be directly transferable to other practical applications such as brain-computer interfaces or alternative communication aids for humans. Therefore, future studies with human EEG data are required to verify the effectiveness of the BiLSTM network in classifying auditory EEG-based speech recognition. Additionally, it needs to be re-evaluated for optimal parameter tuning and feature extraction. It is expected that this study will provide a novel approach for analyzing EEG signals and as well as valuable information regarding the mechanisms of speech perception and recognition in the brain.

Supporting information

S1 Dataset.

(TXT)

S1 File.

(ZIP)

Author Contributions

Conceptualization: Jinsil Ham, Hyun-Joon Yoo, Jongin Kim, Boreom Lee.

Data curation: Jinsil Ham, Hyun-Joon Yoo, Jongin Kim.

Formal analysis: Jinsil Ham, Hyun-Joon Yoo.

Funding acquisition: Jinsil Ham, Hyun-Joon Yoo, Boreom Lee.

Investigation: Jinsil Ham, Hyun-Joon Yoo.

Methodology: Jinsil Ham, Hyun-Joon Yoo, Jongin Kim.

Project administration: Jinsil Ham, Hyun-Joon Yoo, Jongin Kim, Boreom Lee.

Resources: Jinsil Ham, Hyun-Joon Yoo.

Software: Jinsil Ham, Jongin Kim.

Supervision: Boreom Lee.

Validation: Jinsil Ham, Hyun-Joon Yoo, Boreom Lee.

Visualization: Jinsil Ham, Hyun-Joon Yoo.

Writing – original draft: Jinsil Ham, Hyun-Joon Yoo.

Writing – review & editing: Jinsil Ham, Hyun-Joon Yoo, Boreom Lee.

References

1. Wernicke C. The symptom complex of aphasia. In: Cohen RS, Wartofsky MW, editors. Proceedings of the Boston Colloquium for the Philosophy of Science 1966/1968. Dordrecht: Springer Netherlands; 1969. pp. 34–97. https://doi.org/10.1007/978-94-010-3378-7_2
2. Shi Z, Yan S, Ding Y, Zhou C, Qian S, Wang Z, et al. Anterior auditory field is needed for sound categorization in fear conditioning task of adult rat. *Front Neurosci.* 2019; 13: 1374. <https://doi.org/10.3389/fnins.2019.01374> PMID: 31920524
3. Liberman AM, Harris KS, Hoffman HS, Griffith BC. The discrimination of speech sounds within and across phoneme boundaries. *J Exp Psychol.* 1957; 54: 358–368. <https://doi.org/10.1037/h0044417> PMID: 13481283
4. Johnson K. Acoustic and auditory phonetics. Chichester: Wiley-Blackwell; 2012. Available from: <http://site.ebrary.com/id/10483227>
5. Green PA, Brandley NC, Nowicki S. Categorical perception in animal communication and decision-making. *Behav Ecol.* 2020; 31: 859–867. <https://doi.org/10.1093/BEHECO/ARAA004>
6. Craik A, He Y, Contreras-Vidal JL. Deep learning for electroencephalogram (EEG) classification tasks: A review. *J Neural Eng.* 2019; 16: 28. <https://doi.org/10.1088/1741-2552/ab0ab5> PMID: 30808014
7. Näätänen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology.* Elsevier; 2007. pp. 2544–2590. <https://doi.org/10.1016/j.clinph.2007.04.026> PMID: 17931964
8. Garrido MI, Kilner JM, Stephan KE, Friston KJ. The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology.* Elsevier; 2009. pp. 453–463. <https://doi.org/10.1016/j.clinph.2008.11.029> PMID: 19181570
9. Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, et al. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature.* 1997; 385: 432–434. <https://doi.org/10.1038/385432a0> PMID: 9009189
10. Xi J, Zhang L, Shu H, Zhang Y, Li P. Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience.* 2010; 170: 223–231. <https://doi.org/10.1016/j.neuroscience.2010.06.077> PMID: 20633613
11. Prather JF, Nowicki S, Anderson RC, Peters S, Mooney R. Neural correlates of categorical perception in learned vocal communication. *Nat Neurosci.* 2009; 12: 221–228. <https://doi.org/10.1038/nn.2246> PMID: 19136972
12. Perez CA, Engineer CT, Jakkamsetti V, Carraway RS, Perry MS, Kilgard MP. Different timescales for the neural coding of consonant and vowel sounds. *Cereb Cortex.* 2013; 23: 670–683. <https://doi.org/10.1093/cercor/bhs045> PMID: 22426334
13. Engineer CT, Perez CA, Chen YH, Carraway RS, Reed AC, Shetake JA, et al. Cortical activity patterns predict speech discrimination ability. *Nat Neurosci.* 2008; 11: 1842–1845. <https://doi.org/10.1038/nn.2109> PMID: 18425123
14. Engineer CT, Centanni TM, Im KW, Kilgard MP. Speech sound discrimination training improves auditory cortex responses in a rat model of autism. *Front Syst Neurosci.* 2014; 0: 137. <https://doi.org/10.3389/FNSYS.2014.00137> PMID: 25140133
15. Kang H, Auksztulewicz R, An H, Abi Chacra N, Sutter ML, Schnupp JWH. Neural correlates of auditory pattern learning in the auditory cortex. *Front Neurosci.* 2021; 0: 261. <https://doi.org/10.3389/fnins.2021.610978> PMID: 33790730
16. Hosseini MP, Hosseini A, Ahi K. A review on machine learning for EEG signal processing in bioengineering. *IEEE Rev Biomed Eng.* 2021; 14: 204–218. <https://doi.org/10.1109/RBME.2020.2969915> PMID: 32011262

17. Khosla A, Khandnor P, Chand T. A comparative analysis of signal processing and classification methods for different applications based on EEG signals. *Biocybern Biomed Eng*. 2020; 40: 649–690. <https://doi.org/10.1016/j.bbe.2020.02.002>
18. Roy Y, Banville H, Albuquerque I, Gramfort A, Falk TH, Faubert J. Deep learning-based electroencephalography analysis: A systematic review. *J Neural Eng*. 2019; 16: 37. <https://doi.org/10.1088/1741-2552/ab260c> PMID: 31151119
19. Rim B, Sung NJ, Min S, Hong M. Deep learning in physiological signal data: A survey. *Sensors (Switzerland)*. 2020; 20: 969. <https://doi.org/10.3390/s20040969> PMID: 32054042
20. Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Comput Biol Med*. 2018; 100: 270–278. <https://doi.org/10.1016/j.compbimed.2017.09.017> PMID: 28974302
21. Plis SM, Hjelm DR, Salakhutdinov R, Allen EA, Bockholt HJ, Long JD, et al. Deep learning for neuroimaging: a validation study. *Front Neurosci*. 2014; 8: 229. <https://doi.org/10.3389/fnins.2014.00229> PMID: 25191215
22. An X, Kuang D, Guo X, Zhao Y, He L. A deep learning method for classification of EEG data based on motor imagery. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Verlag; 2014. pp. 203–210. https://doi.org/10.1007/978-3-319-09330-7_25
23. Ay B, Yildirim O, Talo M, Baloglu UB, Aydin G, Puthankattil SD, et al. Automated depression detection using deep representation and sequence learning with EEG signals. *J Med Syst*. 2019; 43: 1–12. <https://doi.org/10.1007/s10916-019-1345-y> PMID: 31139932
24. Graves A, Mohamed AR, Hinton G. Speech recognition with deep recurrent neural networks. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing—Proceedings*. 2013. pp. 6645–6649. <https://doi.org/10.1109/ICASSP.2013.6638947>
25. Guerra E, de Lara J, Malizia A, Díaz P. Supporting user-oriented analysis for multi-view domain-specific visual languages. *Inf Softw Technol*. 2009; 51: 769–784. <https://doi.org/10.1016/j.infsof.2008.09.005>
26. Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L. Large-scale video classification with convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014. pp. 1725–1732.
27. Lotte F, Bougrain L, Clerc M. *Electroencephalography (EEG)-based brain-computer interfaces*. Wiley Encycl Electr Electron Eng. 2015; 1–20. <https://doi.org/10.1002/047134608X.W8278>
28. Yin Z, Zhang J. Cross-subject recognition of operator functional states via EEG and switching deep belief networks with adaptive weights. *Neurocomputing*. 2017; 260: 349–366. <https://doi.org/10.1016/j.neucom.2017.05.002>
29. Blankertz B, Lemm S, Treder M, Haufe S, Müller KR. Single-trial analysis and classification of ERP components—A tutorial. *Neuroimage*. 2011; 56: 814–825. <https://doi.org/10.1016/j.neuroimage.2010.06.048> PMID: 20600976
30. Wang C, Xiong S, Hu X, Yao L, Zhang J. Combining features from ERP components in single-trial EEG for discriminating four-category visual objects. *J Neural Eng*. 2012; 9: 056013. <https://doi.org/10.1088/1741-2560/9/5/056013> PMID: 22983495
31. Onoda K, Sakata S. An ERP study of temporal discrimination in rats. *Behav Processes*. 2006; 71: 235–240. <https://doi.org/10.1016/J.BEPROC.2005.12.006> PMID: 16427215
32. Richard N, Laursen B, Grupe M, Drewes AM, Graversen C, Sørensen HBD, et al. Adapted wavelet transform improves time-frequency representations: a study of auditory elicited P300-like event-related potentials in rats. *J Neural Eng*. 2017; 14: 026012. <https://doi.org/10.1088/1741-2552/aa536e> PMID: 28177924
33. Makeig S, Westerfield M, Jung TP, Enghoff S, Townsend J, Courchesne E, et al. Dynamic brain sources of visual evoked responses. *Science (80-)*. 2002; 295: 690–694. <https://doi.org/10.1126/science.1066168> PMID: 11809976
34. Quian Quiroga R, Garcia H. Single-trial event-related potentials with wavelet denoising. *Clin Neurophysiol*. 2003; 114: 376–390. [https://doi.org/10.1016/s1388-2457\(02\)00365-6](https://doi.org/10.1016/s1388-2457(02)00365-6) PMID: 12559247
35. Mustafa M, Guthe S, Magnor M. Single-trial EEG classification of artifacts in videos. *ACM Trans Appl Percept*. 2012; 9: 12. <https://doi.org/10.1145/2325722.2325725>
36. Tzovara A, Murray MM, Plomp G, Herzog MH, Michel CM, De Lucia M. Decoding stimulus-related information from single-trial EEG responses based on voltage topographies. *Pattern Recognit*. 2012; 45: 2109–2122. <https://doi.org/10.1016/j.patcog.2011.04.007>
37. DaSalla CS, Kambara H, Sato M, Koike Y. Spatial filtering and single-trial classification of EEG during vowel speech imagery. *i-CREATE 2009—International Convention on Rehabilitation Engineering and*

- Assistive Technology. Association for Computing Machinery; 2009. pp. 1–4. <https://doi.org/10.1145/1592700.1592731>
38. Yi HG, Xie Z, Reetzke R, Dimakis AG, Chandrasekaran B. Vowel decoding from single-trial speech-evoked electrophysiological responses: A feature-based machine learning approach. *Brain Behav.* 2017; 7: e00665. <https://doi.org/10.1002/brb3.665> PMID: 28638700
 39. Treder MS, Purwins H, Miklody D, Sturm I, Blankertz B. Decoding auditory attention to instruments in polyphonic music using single-trial EEG classification. *J Neural Eng.* 2014; 11: 026009. <https://doi.org/10.1088/1741-2560/11/2/026009> PMID: 24608228
 40. Liu M, Wu W, Gu Z, Yu Z, Qi FF, Li Y. Deep learning based on Batch Normalization for P300 signal detection. *Neurocomputing.* 2018; 275: 288–297. <https://doi.org/10.1016/j.neucom.2017.08.039>
 41. Carabez E, Sugi M, Nambu I, Wada Y. Convolutional neural networks with 3D input for P300 identification in auditory brain-computer interfaces. *Comput Intell Neurosci.* 2017; 2017. <https://doi.org/10.1155/2017/8163949> PMID: 29250108
 42. Pereira A, Padden D, Jantz J, Lin K, Alcaide-Aguirre R, Pereira AE, et al. Cross-subject EEG event-related potential classification for brain-computer interfaces using residual networks. 2018 Sep. <https://doi.org/10.13140/RG.2.2.16257.10086>
 43. Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces. *J Neural Eng.* 2018; 15: 056013. <https://doi.org/10.1088/1741-2552/aace8c> PMID: 29932424
 44. Dithaporn A, Banluesombatkul N, Ketrat S, Chuangsuwanich E, Wilaiprasitporn T. Universal joint feature extraction for P300 EEG classification using multi-task autoencoder. *IEEE Access.* 2019; 7: 68415–68428. <https://doi.org/10.1109/ACCESS.2019.2919143>
 45. Yildirim Ö. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. *Comput Biol Med.* 2018; 96: 189–202. <https://doi.org/10.1016/j.combiomed.2018.03.016> PMID: 29614430
 46. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* 1997; 9: 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735> PMID: 9377276
 47. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks.* Pergamon; 2005. pp. 602–610. <https://doi.org/10.1016/j.neunet.2005.06.042> PMID: 16112549
 48. Kim J, Lee SK, Lee B. EEG classification in a single-trial basis for vowel speech perception using multivariate empirical mode decomposition. *J Neural Eng.* 2014; 11: 036010. <https://doi.org/10.1088/1741-2560/11/3/036010> PMID: 24809722
 49. Mahmoudzadeh M, Dehaene-Lambertz G, Wallois F. Electrophysiological and hemodynamic mismatch responses in rats listening to human speech syllables. Astikainen PS, editor. *PLoS One.* 2017; 12: e0173801. <https://doi.org/10.1371/journal.pone.0173801> PMID: 28291832
 50. Swift KM, Keus K, Echeverria CG, Cabrera Y, Jimenez J, Holloway J, et al. Sex differences within sleep in gonadally intact rats. *Sleep.* 2020; 43: 1–14. <https://doi.org/10.1093/sleep/zsz289> PMID: 31784755
 51. Polley DB, Read HL, Storace DA, Merzenich MM. Multiparametric auditory receptive field organization across five cortical fields in the albino rat. *J Neurophysiol.* 2007; 97: 3621–3638. <https://doi.org/10.1152/JN.01298.2006/ASSET/IMAGES/LARGE/Z9K0050782110014.JPEG>
 52. Moore BCJ. Perceptual consequences of cochlear damage. Perceptual consequences of cochlear damage. New York, NY, US: Oxford University Press; 1995. <https://doi.org/10.1093/acprof:oso/9780198523307.001.0001>
 53. Peterson GE, Barney HL. Control methods used in a study of the vowels. *J Acoust Soc Am.* 1952; 24: 175–184. <https://doi.org/10.1121/1.1906875>
 54. Heffner HE, Heffner RS. Hearing ranges of laboratory animals. *J Am Assoc Lab Anim Sci.* 2007 [cited 2 May 2022]. Available from: <http://www.nrel.gov/docs/fy02osti/30844.pdf#search=%22avi>
 55. Oostenveld R, Fries P, Maris E, Schoffelen JM. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci.* 2011. <https://doi.org/10.1155/2011/156869> PMID: 21253357
 56. Zhang X, Yao L, Zhang D, Wang X, Sheng QZ, Gu T. Multi-person brain activity recognition via comprehensive EEG signal analysis. *ACM Int Conf Proceeding Ser.* 2017; 28–37. Available from: <http://arxiv.org/abs/1709.09077>
 57. Qiu Y, Zhou W, Yu N, Du P. Denoising sparse autoencoder-based ictal EEG classification. *IEEE Trans Neural Syst Rehabil Eng.* 2018; 26: 1717–1726. <https://doi.org/10.1109/TNSRE.2018.2864306> PMID: 30106681
 58. Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016. Available from: <https://www.deeplearningbook.org/>

59. Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks. *Adv Neural Inf Process Syst*. 2014; 4: 3104–3112. Available from: <http://arxiv.org/abs/1409.3215>
60. Graves A. Generating sequences with recurrent neural networks. 2013; 1–43.
61. Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks. 30th International Conference on Machine Learning, ICML 2013. 2013.
62. Schuster M, Paliwal KK. Bidirectional recurrent neural networks. *IEEE Trans Signal Process*. 1997; 45: 2673–2681. <https://doi.org/10.1109/78.650093>
63. Ni Z, Yuksel AC, Ni X, Mandel MI, Xie L. Confused or not confused?: Disentangling brain activity from EEG data using bidirectional LSTM recurrent neural networks. *ACM-BCB 2017—Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. Association for Computing Machinery, Inc; 2017*. pp. 241–246. <https://doi.org/10.1145/3107411.3107513>
64. Ogawa T, Sasaka Y, Maeda K, Haseyama M. Favorite video classification based on multimodal bidirectional LSTM. *IEEE Access*. 2018; 6: 61401–61409. <https://doi.org/10.1109/ACCESS.2018.2876710>
65. Geng M, Zhou W, Liu G, Li C, Zhang Y. Epileptic seizure detection based on stockwell transform and bidirectional long short-term memory. *IEEE Trans Neural Syst Rehabil Eng*. 2020; 28: 573–580. <https://doi.org/10.1109/TNSRE.2020.2966290> PMID: 31940545
66. Hu X, Yuan S, Xu F, Leng Y, Yuan K, Yuan Q. Scalp EEG classification using deep Bi-LSTM network for seizure detection. *Comput Biol Med*. 2020; 124: 103919. <https://doi.org/10.1016/j.combiomed.2020.103919> PMID: 32771673
67. Fraiwan L, Alkhodari M. Investigating the use of uni-directional and bi-directional long short-term memory models for automatic sleep stage scoring. *Informatics Med Unlocked*. 2020; 20: 100370. <https://doi.org/10.1016/j.imu.2020.100370>
68. Srivastava N, Hinton G, Krizhevsky A, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014. <https://doi.org/10.5555/2627435.2670313>
69. Kingma DP, Ba JL. Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations, ICLR; 2015. Available from: <https://arxiv.org/abs/1412.6980v9>
70. Marcot BG, Hanea AM. What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis? *Comput Stat*. 2021; 36: 2009–2031. <https://doi.org/10.1007/S00180-020-00999-9/TABLES/5>
71. Chollet F. Keras: Deep learning library for Theano and Tensorflow. Available from: <https://keras.io/> 2015; 7: T1.
72. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: A system for large-scale machine learning. 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16). Savannah, GA: {USENIX} Association; 2016. pp. 265–283. Available from: <https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi>
73. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res*. 2011; 12: 2825–2830.
74. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995; 20: 273–297. <https://doi.org/10.1007/bf00994018>
75. Breiman L. Random forests. *Mach Learn*. 2001; 45: 5–32.
76. Langley P, Iba W, Thompson K. An analysis of Bayesian classifiers. *AAAI*. 1992. pp. 223–228.
77. Duda RO, Hart PE, Stork DG. *Pattern classification and scene analysis*. Wiley New York; 1973.
78. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. *Am Stat*. 1992; 46: 175–185. <https://doi.org/10.1080/00031305.1992.10475879>
79. Boashash B, Azemi G, Ali Khan N. Principles of time-frequency feature extraction for change detection in non-stationary signals: Applications to newborn EEG abnormality detection. *Pattern Recognit*. 2015; 48: 616–627. <https://doi.org/10.1016/j.patcog.2014.08.016>
80. Harpale VK, Bairagi VK. Time and frequency domain analysis of EEG signals for seizure detection: A review. *International Conference on Microelectronics, Computing and Communication, MicroCom 2016. Institute of Electrical and Electronics Engineers Inc.; 2016*. <https://doi.org/10.1109/MicroCom.2016.7522581>
81. McHugh ML. Interrater reliability: The kappa statistic. *Biochem Medica*. 2012; 22: 276–282. <https://doi.org/10.11613/bm.2012.031> PMID: 23092060
82. Jozefowicz R, Zaremba W. An Empirical exploration of recurrent network architectures. *PMLR*; 2015. pp. 2342–2350. Available from: <http://proceedings.mlr.press/v37/jozefowicz15.html>

83. Tsiouris K, Pezoulas VC, Zervakis M, Konitsiotis S, Koutsouris DD, Fotiadis DI. A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals. *Comput Biol Med.* 2018; 99: 24–37. <https://doi.org/10.1016/j.compbiomed.2018.05.019> PMID: 29807250
84. Michielli N, Acharya UR, Molinari F. Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals. *Comput Biol Med.* 2019; 106: 71–81. <https://doi.org/10.1016/j.compbiomed.2019.01.013> PMID: 30685634
85. Alhagry S, Fahmy AA, El-Khoribi RA. Emotion recognition based on EEG using LSTM recurrent neural network. *International Journal of Advanced Computer Science and Applications (IJACSA).* 2017. Available from: www.ijacsa.thesai.org
86. Wang P, Jiang A, Liu X, Shang J, Zhang L. LSTM-based EEG classification in motor imagery tasks. *IEEE Trans Neural Syst Rehabil Eng.* 2018; 26: 2086–2095. <https://doi.org/10.1109/TNSRE.2018.2876129> PMID: 30334800
87. Sharma R, Pachori RB, Sircar P. Automated emotion recognition based on higher order statistics and deep learning algorithm. *Biomed Signal Process Control.* 2020; 58: 101867. <https://doi.org/10.1016/j.bspc.2020.101867>
88. Fraiwan L, Alkhodari M. Classification of focal and non-focal epileptic patients using single channel EEG and long short-term memory learning system. *IEEE Access.* 2020; 8: 77255–77262. <https://doi.org/10.1109/ACCESS.2020.2989442>
89. Fraiwan L, Alkhodari M. Neonatal sleep stage identification using long short-term memory learning system. *Med Biol Eng Comput.* 2020; 58: 1383–1391. <https://doi.org/10.1007/s11517-020-02169-x> PMID: 32281071
90. Wei Y, Qi X, Wang H, Liu Z, Wang G, Yan X. A multi-class automatic sleep staging method based on long short-term memory network using single-lead electrocardiogram signals. *IEEE Access.* 2019; 7: 85959–85970. <https://doi.org/10.1109/ACCESS.2019.2924980>
91. Mazaheri A, Picton TW. EEG spectral dynamics during discrimination of auditory and visual targets. *Cogn Brain Res.* 2005; 24: 81–96. <https://doi.org/10.1016/j.cogbrainres.2004.12.013> PMID: 15922161
92. Strauß A, Wöstmann M, Obleser J. Cortical alpha oscillations as a tool for auditory selective inhibition. *Front Hum Neurosci.* 2014; 8: 350. <https://doi.org/10.3389/fnhum.2014.00350> PMID: 24904385
93. Kubetschek C, Kayser C. Delta/Theta band EEG activity shapes the rhythmic perceptual sampling of auditory scenes. *Sci Rep.* 2021; 11: 2370. <https://doi.org/10.1038/s41598-021-82008-7> PMID: 33504860
94. Maclver MB, Bland BH. Chaos analysis of EEG during isoflurane-induced loss of righting in rats. *Front Syst Neurosci.* 2014; 8: 203. <https://doi.org/10.3389/fnsys.2014.00203> PMID: 25360091