Automating synaptic plasticity analysis: A deep learning approach to segmenting hippocampal field potential signal

Sabri Altunkaya [1]

Department of Electrical and Electronics Engineering, Necmettin Erbakan University, Konya 42090, Turkey

A R T I C L E   I N F O

A B S T R A C T

Hippocampal field potentials are widely used in research on neurodegenerative diseases, epilepsy, neuropharmacology, and particularly long- and short-term synaptic plasticity. To conduct these studies, it is necessary to identify specific components within hippocampal field potential signals. However, manually marking the relevant signal points for analysis is a time-consuming, error-prone, and subjective process. Currently, there is no specialized software dedicated to automating this task. In this study, three different recurrent neural network-based deep learning architectures were examined for the automatic segmentation of hippocampal field potential signals in two separate experimental studies. In the first experimental study, 10,836 epochs of field potential signals recorded from 54 rats were used, and in the second experimental study, field potential signals with noise added to the above data at different rates were used. The best model achieved an average f-score of 98.1% on noise-free data and 97.15% on data with noise, highlighting its robustness in real-world scenarios. Furthermore, we assessed system stability using the repeated holdout method, which randomly split the data into training and testing sets 100 times, and each time trained a new version of the system. As a result, the proposed system was proven to be reliable and generalizable by showing similar average scores and low variability across all 100 iterations of the test.

1. Introduction

Extracellular field potentials (EFP) are voltage fluctuations generated by the collective electrical activity of numerous neurons. These potentials arise from the summation of tiny electrical dipoles created when ions flow in and out of cells. These ionic currents propagate through the resistive extracellular space, producing measurable voltage changes [1]. They cover a variety of signals, from invasive recordings such as electrocorticogram (ECoG) and local field potential (LFP), to noninvasive methods such as the electroencephalogram (EEG) [2,3]. One specific type of EFP is the evoked field potentials in the hippocampus (HFP), which are classified as LFP and recorded as dentate gyrus responses evoked by electrical stimulation of the perforant pathway. The evoked HFP is a complex waveform containing different signal components. These signal components are a stimulus artifacts, and a field excitatory post synaptic potentials (fEPSP), and population spikes (PS) [4]. The fEPSP reflects synaptic depolarization resulting from the efflux of positive ions from the recording area. The PS represents the electrical activity generated by action potentials traveling along the stimulated

fibers and arriving at the recording site [4–6]. The fEPSP is quantified by measuring the slope of the initial, sharp rise in the evoked HFP following the stimulation artifact [7]. The steepness of the slope is correlated with the magnitude of postsynaptic depolarization and indirectly measures the strength of synaptic connections [4]. The PS is quantified by measuring the amplitude of the fast negative transient superimposed on the fEPSP [8]. It is a measure of the overall excitability of the granule cell population [5,6]. Recording evoked field potentials in the hippocampus has been widely used to investigate both long- and short-term forms of synaptic plasticity, providing useful information about the neural mechanisms of learning and memory [9–19]. In these studies, HFP components and their associated fEPSP parameters need to be estimated. However, manual signal labeling, such as identifying HFP signal components, is prone to human error, especially in repetitive and monotonous tasks. The reliability of labels can be significantly affected by human factors, which renders the labeling process subjective [20–22]. In addition, no automatic tool exists for analysis of HFP signals without the assistance of a neuroscientist. This lack of automation arises from several factors, including temporal shifts, as noted in [9], and the
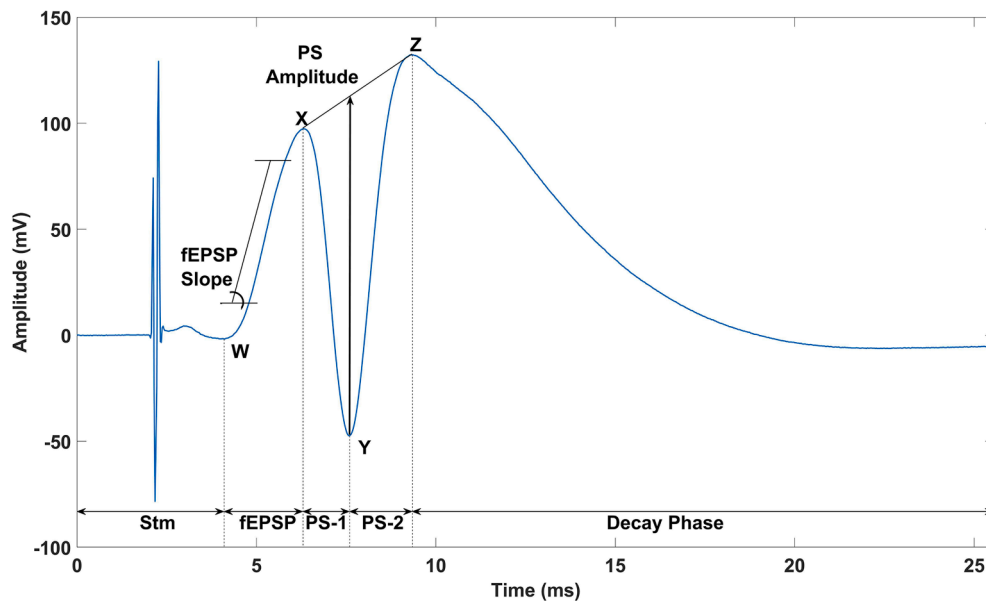
Fig. 1. A sample of the HFP signal epoch and its labels. (Stm: Stimulation, fEPSP: field excitatory postsynaptic potential, PS-1: Population spike-1, PS-2: Population spike-2).

presence of noise in the field potential signals originating from the recording system.

Deep learning-based time-series segmentation methods, which divide signals into non-overlapping segments [23], such as those used by Zhao et al. [20], can effectively identify HFP signal components. Deep learning models can learn complex patterns and features directly from raw input data without the need for manual feature engineering. These make them well-suited for tasks such as time-series segmentation [24]. Deep learning-based segmentation approaches have been used for different tasks in the literature, such as segmentation of the QRS complex, P and T waves of ECG signals [24–26], segmentation of the first and second heart sounds, systolic and diastolic periods of heart sound signals [27], segmentation of the systolic and diastolic phases in wrist pulse signals [28], segmentation of Korotkoff sounds[29], sleep staging using single-channel EEG and EOG [22,30], and human action recognition [31]. The deep learning models used for the segmentation of time-domain signals are based on recurrent neural networks (RNNs). RNNs are designed to process sequential or time-varying data [32,33]. The most popular such networks are bidirectional RNNs [34], long short-term memory (LSTM) networks [35], bidirectional LSTM [36], and gated recurrent unit (GRU) networks [37].

Anderson and Collingridge developed a stimulation, data collection, and automated analysis program for performing on experiments synaptic plasticity. When determining fEPSP parameters, they looked for maxima in a certain time interval after stimulation. In this study, no measurements were made regarding the accuracy of the positions of the maximums and the calculated fEPSP parameters [38]. Gholmieh et al. developed an FPGA-based hardware algorithm for real-time extraction of population EPSP and population spike amplitudes from in vitro hippocampal field potential recordings. [39]. Rubega et al. employed Phillips-Tikhonov regularization, which is an effective method for handling noisy data, to identify maxima in local field signals. To minimize errors in minima detection, the experimenter applied this method within a predefined time window. This approach yielded latency and amplitude features of LFP with relative errors below 3 % [19]. Zhao et al. proposed a method for calculating fEPSP parameters that uses expert systems and deep learning algorithms [20]. In Zhao et al.'s study, an expert system was developed to remove contaminated field potential signals and label them. Then an LSTM based system was developed to segment the HFP signals. The accuracy of the proposed LSTM model was

96.73 %. Although the study achieved high accuracy, it did not include other metrics, and the fEPSP parameters were not validated by comparing the automatically and manually derived fEPSP parameters. Lastly, the developed model has not been evaluated for noisy data, the pre-stimulation and final part of the decay phase in HFP signals were not included in the segmentation process, and only one deep learning model was evaluated.

While deep learning-based approaches have been successfully used in the past for the segmentation of electrophysiological recording, including of neuroplasticity-related fEPSP, they were performed in noise-free only condition, and they did not compare the fEPSP parameter obtained result of segmentation with those obtained manually. Here, we propose utilizing deep learning methods for the segmentation of data collected for the research of metaplasticity mechanisms [11,15]. All models were evaluated with standard performance metrics and compared to manually obtained fEPSP parameters. All these evaluations were repeated for hippocampal field potential signals with varying levels of noise added.

2. Methods

2.1. Animals

This study analyzed data from existing animal research to avoid further animal use. We used the hippocampal field potential records of 54 60-day-old adult male Wistar albino rats that were produced in the Erciyes University's experimental and clinical research center with the approval of the animal experiments ethics committee dated 17.07.2019 and numbered 19/133.

2.2. Stereotaxic surgery and recording of HFP

Rats were anesthetized with an intraperitoneal injection of 1.2 g/kg urethane. Rat skulls were fixed using a stereotactic frame, and a small hole was drilled on the right side of the skulls. After craniotomy, bipolar stainless-steel wire electrodes were inserted into the angular bundle of the perforant pathway for stimulation. A borosilicate glass capillary micropipette filled with 3 M NaCl solution and an Ag/AgCl wire which is in contact with the solution was used as the recording electrode. The recording electrode was lowered into the dentate gyrus for response. An
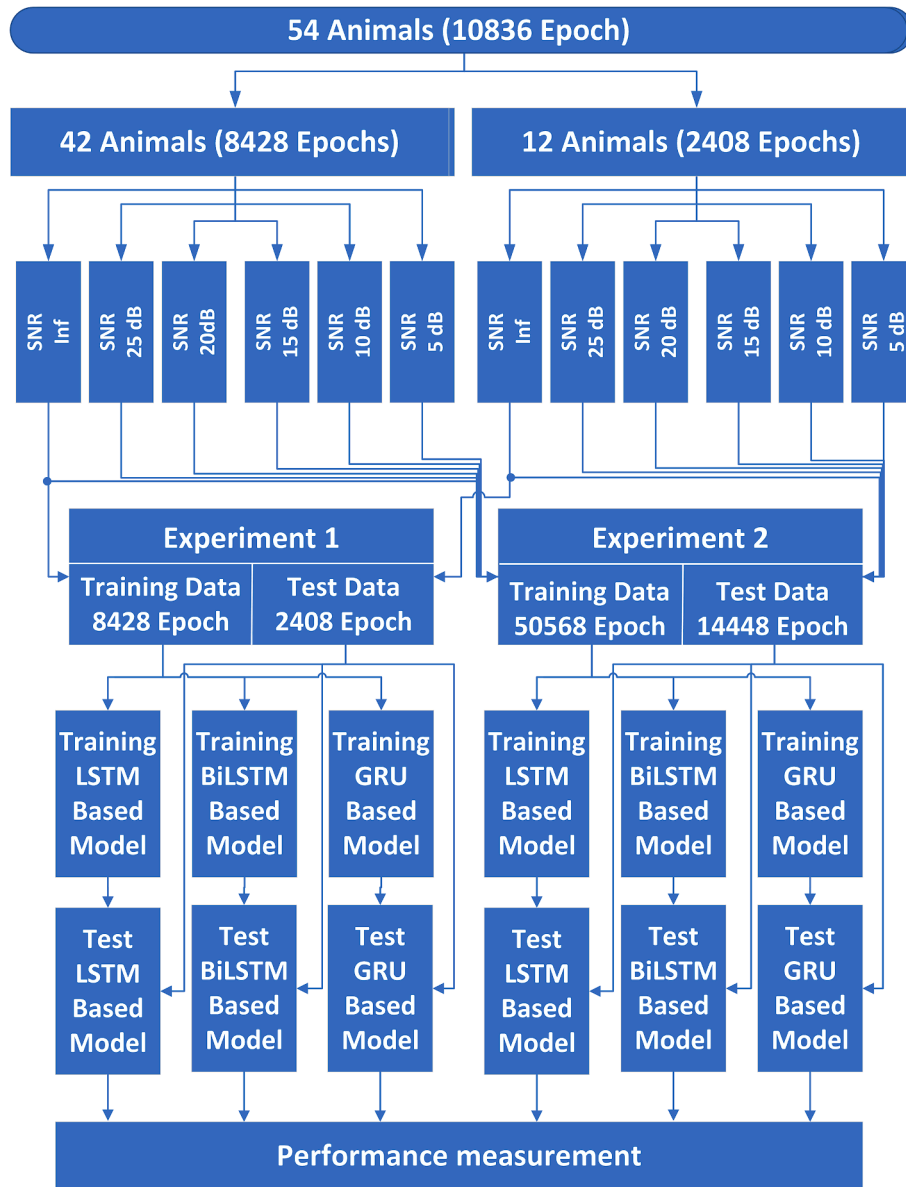
Fig. 2. Schematic of experimental setup and data structure.

Ag–AgCl disk electrode positioned under the neck skin was used for reference. The entire system was shielded using a Faraday cage. The locations of the electrodes were determined using coordinates taken from the Paxinos and Watson atlases [40], and the locations of the dentate gyrus and perforant pathway were previously confirmed [16]. Data acquisition hardware (DAQ) and software (AD Instrument, Powerlab/8SP and Scope, USA) were used to implement synaptic plasticity protocols. The stimulation produced by the DAQ was sent to the stimulation electrode via a stimulus isolator (World Precision Instruments, A385, USA). The recording electrode was connected to the analog input of the data acquisition hardware over a computer-controlled current and voltage clamp (MultiClamp 700B Axon CNS, Molecular Devices, USA), and the response to the stimulation was recorded. More detailed information about the preparation of rats, electrode placement, and recording system can be found in the literature [15,41]. All data were recorded at a sampling rate of 40 kHz. This resulted in each epoch of the HFP signal having 1024 points, corresponding to a duration of 25.575 ms.

2.3. Dataset

Fifty-four rats were divided into three groups. The first group underwent long-term potentiation (LTP), the second group underwent long term depression (LTD), and the third group experienced metaplasticity (MP) induction. A detailed description and implementation details of LTP, LTD, and MP can be found in [11,15]. A total of 206 epochs of HFP signals were obtained from a rat subjected to the LTP experiment, 180 epochs from a rat subjected to the LTD experiment, and 216 epochs from a rat subjected to the MP experiment. A total of 10,836 epochs of HFP signals were recorded, including 3708 epochs from 18 rats subjected to LTP, 3240 epochs from 18 rats subjected to LTD, and 3888 epochs from 18 rats subjected to MP. An expert neuroscientist individually labeled each resulting epoch. After labeling, basic preprocessing steps were applied to each epoch. These are removing the DC value of a signal by subtracting its mean and normalizing it to the absolute maximum of the signal.
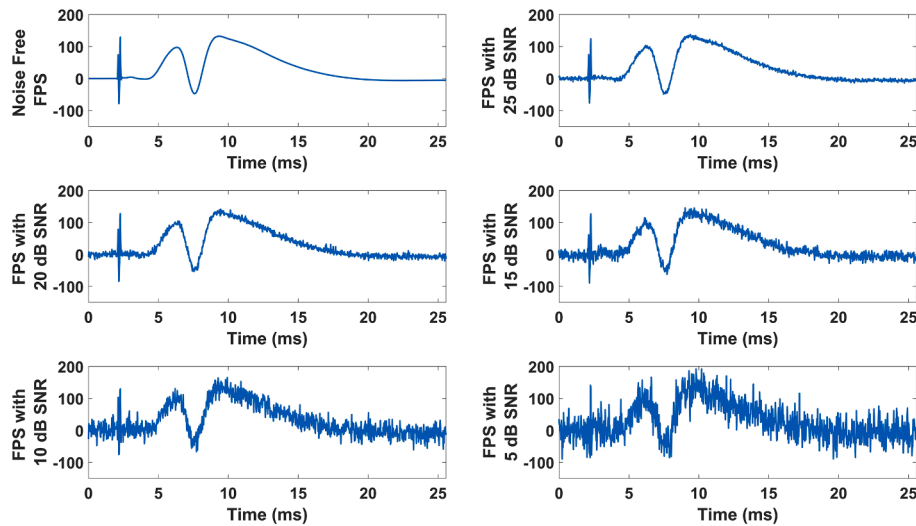
Fig. 3. An epoch of a HFP signal with five different noise levels added (FPS: Field potential signal).

2.4. Estimating the fEPSP slope and the PS amplitude

The fEPSP slope is the slope of the initial increase after the stimulation artifact in the HFP signal. The PS amplitude is the difference in voltage between the negative and positive peaks of the population spike. Fig. 1 shows a sample of the HFP signal segment with key maximum and minimum points (W , X, Y and Z). The fEPSP slope is calculated by subtracting the amplitude 20 % above the amplitude of point W (A_{W20}) from the amplitude 20 % below the amplitude of point X (A_{X20}) and dividing it by the relevant time interval (t_{X20}, t_{W20}). The PS amplitude is calculated by taking the average of the amplitudes obtained by subtracting the Y point amplitude (A_Y) from the X point amplitude (A_Z) and Z point amplitude (A_Z) [9,42,43]. The fEPSP slope and PS amplitude calculations described above are shown in Eq. (1).

$$EPSP\ Slope = \frac{A_{X20} - A_{W20}}{t_{X20} - t_{W20}} \tag{1}$$

$$PS\ Amplitude = \frac{(A_X - A_Y) + (A_Z - A_Y)}{2}$$

2.5. Experiments

Two experimental setups were used in this study. Two different models were developed: one, using raw HFP signals and the other using HFP signals containing additional noise at different SNR rates. Fig. 2 shows the experimental setups and the data structure of these experimental setups.

In the first experiment, raw HFP signals recorded from 54 rats were used. Of the 54 rats, 42 (approximately 80 %) were randomly allocated to training data and 12 (approximately 20 %) to test data. While making this distinction, care was taken to ensure that there were equal numbers of LTP-, LTD- and MP-induced rats in both the training and test groups. In other words, 14 LTP-, 14 LTD- and 14 MP-induced rats were used in training data, and 4 LTP-, 4 LTD- and 4 MP-induced rats were used in the test data. As a result, 8428 epochs were used for training and 2408 epochs were used for testing in the first experiment.

In the second experiment, we aimed to develop a model that can also handle noisy HFP signals. For this purpose, noise was added to each epoch of the HFP signal so that the signal-to-noise ratio of the HFP signal is 5, 10, 15, 20, and 25 dB. As a result of adding noise, we obtained five noisy and one noise-free HFP signal set. Fig. 3 shows an epoch of raw and HFP signals with five different noise levels. Then, these raw and 5 different noisy HFP signals were combined. As a result of this combination, a dataset containing 65,016 epochs HFP signal was obtained. Training and testing data were separated by ensuring equal numbers of LTP-, LTD- and MP-induced animals, as in Experiment 1. As a result, 65,016 epochs of HFP signals were used, 50,568 epochs were used for training, and 14,448 epochs were used for testing in the second experiment.

In addition, the model obtained in experiment two was tested separately for raw and 5 different noise-added HFPs, and the success of model two at different noise levels was demonstrated.

2.6. Deep learning model architecture

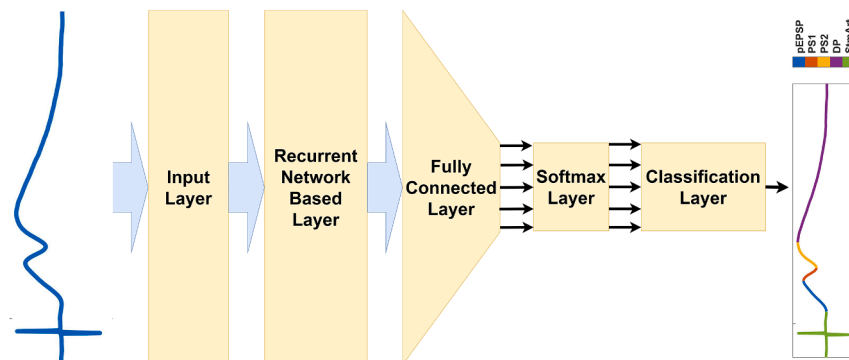In this section, we introduce a deep learning model for the semantic

Fig. 4. Architecture of the deep learning model for labeling the HFP signal.

Table 1

Overall metrics and per-segment f-scores of the best models in Experiment 1.

Model	Method	NU	Per-segment f-score (%)					Overall metrics (%)			
			Stm	fEPSP	PS1	PS2	DP	ACC	F1	PR	RE
Exp1	LSTM	600	99.00	97.60	97.60	95.80	99.50	99.5	97.9	97.6	98.2
	BiLSTM	**600**	**99.00**	**97.70**	**98.10**	**96.10**	**99.50**	**99.6**	**98.1**	**97.9**	**98.2**
	GRU	500	98.80	97.30	97.70	95.90	99.50	99.5	97.8	97.4	98.3

segmentation of HFP signals. The model includes an input layer, a recurrent neural network-based layer, a fully connected layer, a SoftMax layer, and a classification layer, as shown in Fig. 4.

The recurrent neural network-based layer is essential for modeling the temporal dynamics of signals and generating informative representations for subsequent layers. Recurrent neural networks (RNNs) are a class of neural network models specifically designed for sequential and time series data, that use hidden states as memory [44]. However, they suffer from vanishing/exploding gradients, which hinder long-term dependency learning. In this study, three different RNN-based algorithms designed to mitigate gradient vanishing and exploding problems, which hinder long-term dependency learning, were independently employed. These are LSTM networks long-term dependencies through their cell state [35,45,46], BiLSTM networks that capture information from both the past and future using two recurrent layers [47,48], and GRU networks that efficiently manage information flow with update and reset gates [37,49].

Fully connected and SoftMax layers are used to classify the output of an RNN-based layer and assign a label to each data point. A fully connected layer obtains the output of the RNN-based layer and multiplies it by a weight matrix. Then, it adds a bias vector to the results and outputs the results to the SoftMax layer. The SoftMax layer takes the output of the fully connected layer and applies a SoftMax function to it. This function converts the output to a probability distribution over all possible labels. The classification layer then takes the probability distribution and selects the label with the highest probability.

2.7. Training procedure

The experiments were implemented using the MATLAB (The MathWorks Inc., R2024a) environment. An adaptive moment estimation optimizer (ADAM) [50] with default parameters ($\beta 1 = 0.9, \beta 2 = 0.999$) was used to optimize all weights. Various initial learning rates (0.001 and 0.01), mini batch size (8 16 32), and number of units (NU) of the RNN-based layer (200 300 400 500 600) were trained and tested. Consequently, 30 different models were trained and tested for 3 different mini bath size values, 2 different learning rate values, and 5 different NU values of the RNN-based layer for each RNN-based deep learning algorithms. Learning rate of 0.001 and batch size of 8 give the best test accuracy. We conducted the experiments on a notebook with one Intel(R) Core (TM) i5-11400H CPU @ 2.70 GHz processor and 16 GB RAM and one NVIDIA GeForce RTX3050 TI4 GB GPU. All experiments were performed on Windows 10 Pro 64-Bit.

2.8. Evaluation measures

Segmentation performance was evaluated using the accuracy, recall, precision, the f-score, and confusion matrices. The accuracy was calculated using the overall time points in all five classes. Precision, recall, and the f-score were calculated for each class, and the arithmetic mean was obtained for the overall comparison. The fEPSP parameters (EPSP slope and PS amplitude) were calculated using the time points obtained for actual and predicted segmentation and compared using the mean absolute percentage error (MAPE), normalized mean square error (NMSE), and Pearson's correlation coefficient. The calculation of evaluation metrics is described in Supplementary Table 1.

3. Results and discussion

3.1. Experiment 1

Table 1 lists the models that gave the best results among the models for each RNN-based algorithm in Experiments 1. The best models for Experiment 1 gave approximately similar results, as shown in Table 1. The BiLSTM-based model was the best model found for Experiment 1 with 98.1 % f-score, 97.9 % precision, and 98.2 % recall. To better examine the evaluation parameters of the segments, the confusion matrix of the BiLSTM model is presented in Table 2. The confusion matrix was normalized on a row (true class) basis. When the confusion matrix was examined, the f-score, recall, and precision values of the PS2 segment had the lowest values among the other segments. The recall value of the PS2 segment was 96.9 %. Therefore, 3.1 % of the time points that actually should have been PS2 were predicted incorrectly. Most of these incorrect predictions (2.6 %) were due to the PS2 segment being predicted as the DP segment. The precision of the PS2 segment was 95.4 %. In other words, 4.6 % of the time points that the model predicted as PS2 were incorrect. As shown in the vertically normalized confusion matrix in Supplementary Table 2, 4.3 % of these incorrect predictions were actually DP segments.

The population spike component of the HFP reflects the number of neurons that produce action potentials. It begins with the simultaneous discharge of neurons in the neuron pool whose membrane potential raises enough to trigger the formation of an action potential. At this point, due to the change in the direction of the current, the field potential trace changes direction and makes its first negative deviation. When the amount of simultaneously firing neurons reaches its maximum, the PS component reaches its maximum value and then begins to decrease, forming the second positive peak. As the electrical activity of the neurons decreases, the trace begins to decrease again,

Table 2

Confusion matrix and per-segment evaluation metrics of the BiLSTM model in Experiment 1.

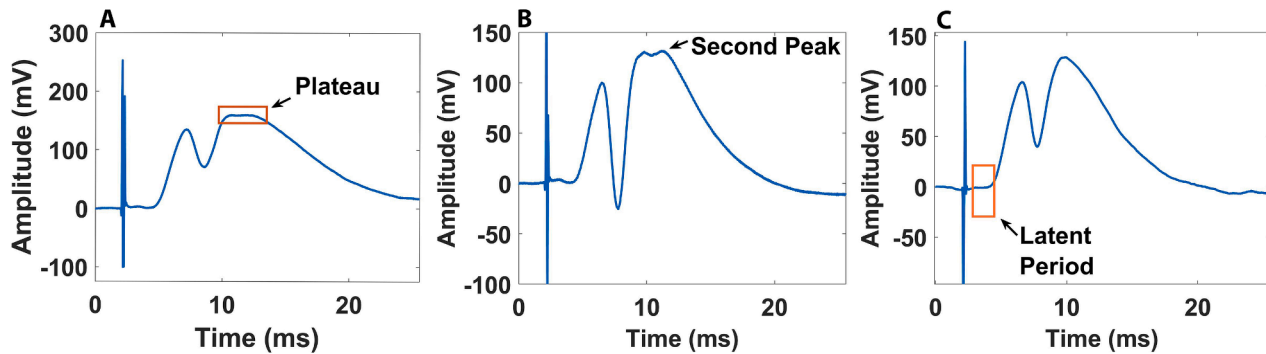|		Predicted Class (%)						Per-segment metrics (%)		
		Stm	fEPSP	PS1	PS2	DP		F1	PR	RE
True Class (%)	**Stm**	99.3	0.7					99	98.7	99.3
	fEPSP	2.2	97	0.8				97.7	98.4	97
	PS1		0.7	98.6	0.6			98.1	97.5	98.6
	PS2			0.5	96.9	2.6		96.1	95.4	96.9
	DP				0.6	99.4		99.5	99.7	99.4
Mean								**98.08**	**97.94**	**98.24**

Fig. 5. A sample of HFP with: **a)** plateau phases; **b)** an extra peak following the population spike component; **c)** a constant value during the latent period.

Table 3
Overall metrics and per-segment f-scores of the best models in Experiment 2.

Model	Method	NU	Per-segment f-score (%)					Overall metrics (%)			
			Stm	fEPSP	PS1	PS2	DP	ACC	F1	PR	RE
Exp2	LSTM	400	97.84	94.08	92.95	90.61	98.93	98.96	94.88	94.19	95.67
	BiLSTM	**400**	**98.65**	**96.85**	**96.97**	**93.98**	**99.30**	**99.37**	**97.15**	**96.93**	**97.38**
	GRU	500	97.76	93.78	92.24	91.43	99.07	98.99	94.86	94.17	95.66

showing a "decay" phase. However, in some experiments, the HFPs recorded after the induction protocol decreased following a plateau phase (Fig. 5.a). In such cases, when labeling HFP signals, the first point where the slope of the field potential trace decreases and the signal flattens is labeled as the end of PS2. When labeling signals with plateaus, which are rarely observed compared with many recorded signals, we thought that the model cannot learn to label and looks for a decrease after the plateau and predicts the point where the decrease begins as the end of PS2. Even less commonly, an extra peak may occur in the signal because of desynchronization during firing (Fig. 5.b). Generally, neuroscientists ignore this extra peak when labeling and consider the first peak to be the end of the PS2 segment. Since the number of signals exhibiting an extra second peak is very low, we assumed that the model could not learn to ignore the extra peak and evaluate it as the end of the PS2 segment. Therefore, most of the PS2 segment prediction error occurred when PS2 was estimated as the DP segment.

In Table 2, the second lowest f-score belongs to the fEPSP segment. The recall value of the fEPSP segment was 97 %. In other words, 3 % of the segments that should actually be fEPSP were estimated incorrectly. 2.2 % of these incorrect predictions were due to the fEPSP segment being predicted as the STM segment. The precision of the fEPSP segment was 98.4 %. In other words, 1.6 % of the fEPSP segment predicted by the model were incorrect. As shown in the vertically normalized confusion matrix shown in Supplementary Table 2, 1.2 % of these incorrect predictions were actually STM segments.

The fEPSP component of the HFP provides a measure of the excitatory drive to dentate gyrus neurons. Generally, the onset of this component coincides with the potential reaching a minimum value, with the potential gradually and imperceptibly decreasing following the stimulus artifact in the latent period. Although the dataset largely consists of signals with a minimum potential value corresponding to the onset of the fEPSP component, the potential may rarely exhibit a constant value during the latent period. In such a case, an example of which is shown with the trace without a minimum point in the latent period in Fig. 5.c, the model cannot accurately determine the beginning of the fEPSP component. During labeling, the model presumably continues to search for the minimum for the starting point of the fEPSP, and if not, the model labels the beginning as the first point where the increase began in the latent period. This may cause the starting point of the fEPSP to remain incorrectly within the latent period.

3.2. Experiment 2

Table 3 shows the models that gave the best results among the tried models for each RNN-based algorithm in Experiment 2. Among the best models given for Experiment 2, Table 3 shows that the BiLSTM-based model was more successful than the LSTM- and GRU-based models. The BiLSTM-based model achieved 97.15 % f-score, 96.93 % precision, and 97.38 % recall. In other words, the BiLSTM-based model showed higher success in the HFP signal with added noise. BiLSTM is more robust in capturing patterns and understanding the overall context of data. This is helpful in scenarios where noisy data may introduce uncertainties or irregularities that need to be considered from different temporal perspectives [51,52].

Table 4 shows the row-normalized confusion matrix of the BiLSTM model that gave the best results in Experiment 2. When the confusion matrix was examined, the lowest f-score belongs to the PS2 segment. The recall value of the PS2 segment was 95.84 %. In other words, 4.17 % of
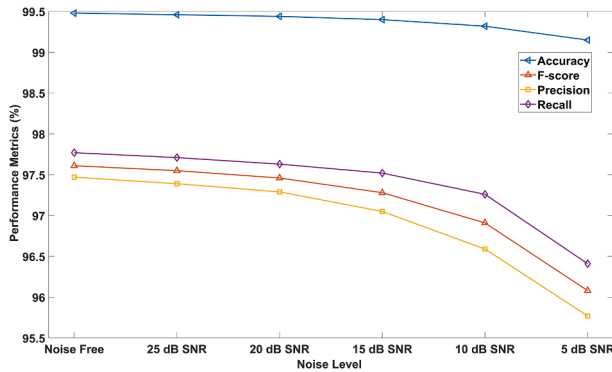
Table 4
Confusion matrix and per-segment evaluation metrics of the BiLSTM model for Experiment 2.

True Class (%)		Predicted Class (%)						Per-segment metrics (%)		
		Stm	fEPSP	PS1	PS2	DP		F1	PR	RE
	Stm	99.08	0.92	0	0	0		98.65	98.22	99.08
	fEPSP	2.99	96.24	0.77	0	0		96.85	97.46	96.24
	PS1	0.06	1.73	96.72	1.49	0		96.97	97.22	96.72
	PS2	0.04	0.12	0.59	95.84	3.42		93.98	92.19	95.84
	DP	0	0	0	0.95	99.04		99.30	99.55	99.04
Mean								**97.15**	**96.93**	**97.38**

Table 5
Overall metrics and per-segment f-scores of the BiLSTM model in Experiment 2 at different noise levels in the HFP signals.

Model	Noise Level	Per-segment f-score (%)					Overall metrics (%)			
		Stm	fEPSP	PS1	PS2	DP	ACC	F1	PR	RE
Exp2	Noise Free	98.77	97.18	97.45	95.23	99.44	99.48	97.61	97.47	97.77
	SNR 25 dB	98.76	97.16	97.41	95	99.41	99.46	97.55	97.39	97.71
	SNR 20 dB	98.75	97.12	97.32	94.71	99.37	99.44	97.46	97.29	97.63
	SNR 15 dB	98.73	97.05	97.16	94.15	99.31	99.40	97.28	97.05	97.52
	SNR 10 dB	98.66	96.80	96.72	93.17	99.20	99.32	96.91	96.59	97.26
	SNR 5 dB	98.21	95.79	95.73	91.58	99.06	99.15	96.08	95.77	96.41

Fig. 6. Evaluation metrics of the BiLSTM model at different noise levels in the HFP signals.

Table 6
Mean absolute percentage error between the actual and predicted EPSP slope and PS amplitude.

	MAPE	
	EPSP slope	PS amplitude
Experiment 1	0.50 %	0.71 %
Experiment 2	19.60 %	10.46 %

the segments that should actually be PS2 were predicted incorrectly. Most of these incorrect predictions (3.42 %) result from predicting the PS2 segment as the DP segment. The precision of the PS2 segment was 92.19 %. As can be seen in the vertically normalized confusion matrix in Supplementary Table 3, 7.81 % of the time points at which the system predicted as PS2 were estimated incorrectly. Most of these incorrect predictions (7.06 %) occurred as segments that were actually DP and were predicted as PS2 segments.

As expected, the performance of the model developed with the dataset with added noise was lower. When the f-scores of each segment are compared in the complexity matrices given in Table 2 and Table 4, the greatest decrease in f-scores were observed for the PS2 and PS1 segments. The increase in error in PS2 is due to the increase in the false positive value, that is, the increase in the rate of incorrect predictions as PS2, and the reason for the increase in error in PS1 is due to the increase in the false negative value, that is, the increase in the rate of incorrect prediction of time points that are actually PS1 as both fEPSP and PS2.

To determine the success of the best model obtained in Experiment 2 at different noise levels, the BiLSTM-based model was tested separately for raw and 5 different noise-added HFP signals. Table 5 lists the per-segment f-score values and overall accuracy, precision, and recall values for each noise level. The confusion matrices of the results given in Table 5 are presented in Supplementary Table 4. Fig. 6 shows how evaluation metrics change with noise level. As shown in Table 5 and Fig. 6, model success decreased as the noise level increased. This decrease is especially noticeable in signals with 10 and 5 dB added noise.

3.3. Comparing fEPSP parameters

The purpose of segmentation in this study was to determine the fEPSP slope and PS amplitude for synaptic plasticity studies. Therefore, in addition to the segmentation performance of the model, the fEPSP parameters calculated according to actual and predicted segmentation was also examined for comparison purpose. The difference between the real and predicted fEPSP parameters was determined by calculating the mean absolute percentage error (MAPE). In addition, the relationship between actual and predicted fEPSP parameters was evaluated using Pearson's correlation coefficient.

Table 6 lists the MAPE values obtained in both experiments. Fig. 7 shows scatter plots of the real and predicted fEPSP parameters. There is no standard for evaluating the errors in the fEPSP parameters calculated by different methods. However, considering the values given in Table 6, the MAPE is quite small and acceptable, especially for Experiment 1. Similarly, it can be seen that there is a high correlation between the fEPSP parameters obtained as a result of the automatic and manual segmentation of the noiseless HFP signals in Experiment 1 (Fig. 7.a and 7.b). Because of the noise added to the HFP signals, it is expected that MAPE would be higher than that in Experiment 1, and correlation between synaptic parameters would be lower than that in Experiment 1. However, although there was not much difference between Experiment 1 and Experiment 2 in terms of segmentation performance measures, the MAPE was higher than expected, and the correlation was lower than expected for noisy data. Because the fEPSP slope and PS amplitude calculations directly depend on the amplitude of the HFP signal, increasing the noise level increases the error in the calculation of these parameters. When the MAPE and correlations are calculated separately for different noise levels, as given in Supplementary Table 5 and Supplementary Fig. 1, it is understood that the amplitude change caused by the noise added to the HFP signal, especially at 5- and 10-dB SNR values, has a highly misleading effect in determining the fEPSP parameters. When the added noise level is high, the calculation of the fEPSP parameters is more prone to erroneous results in the noise data because even, small segmentation errors cause large changes in the amplitude.

In both experiments, the PS amplitude MAPE values were lower than the EPSP slope MAPE values listed in Table 5. Similarly, in both experiments, the correlation coefficients of the actual and predicted PS amplitudes were higher than the EPSP slope correlation coefficient. Although the PS2 segment's performance measurement was worse, notably the PS amplitude was calculated with less error. It was explained above that the error in finding the PS2 segment was mostly due to the estimation of the PS2 segment as the DP segment, and this error was thought to be due to the plateau shown in Fig. 5.a. Even if the PS2 segment is estimated incorrectly due to a plateau, as shown in Fig. 5.a, the PS amplitude parameter can be calculated correctly because the amplitude at the beginning and end of the plateau does not change. Although PS2 segment performance measurements were low, a more accurate calculation of the PS amplitude parameter confirmed the plateau explanation that we put forward as the cause of the PS2 segment error.
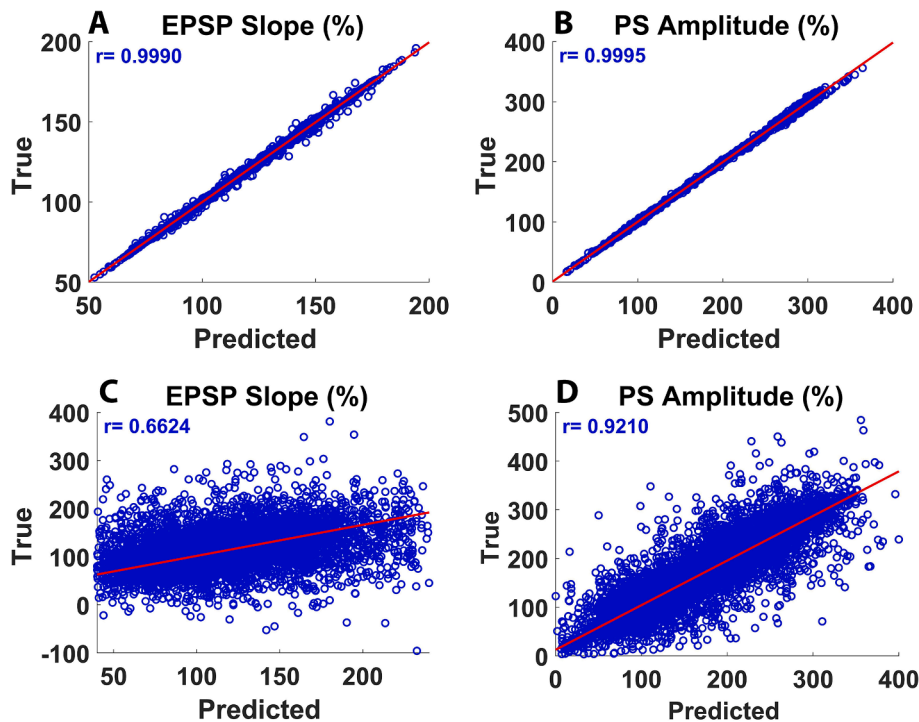
Fig. 7. Scatter plot of: **a.** Actual and predicted EPSP slopes for Experiment 1; **b.** Actual and predicted PS amplitudes for Experiment 1 (N=2408); **c.** Actual and predicted EPSP slopes for Experiment 2; **d.** Actual and predicted PS amplitudes for Experiment 2 (N=14448).

Table 7
Normalized mean square error between actual and predicted EPSP slope and PS amplitude.

	NMSE	
	EPSP slope	**PS amplitude**
Experiment 1	0.0095 %	0.017 %
Experiment 2	7.71 %	2.40 %

3.4. Comparison with other studies

Although hippocampal HFP signals are widely used in memory and learning research, there is a lack of dedicated tools to automate the segmentation of these signals during analysis. In fact, to the best of our knowledge, only two such studies have been conducted for this purpose. The first of these studies proposed an algorithm that can be used on specific hardware device for real-time inference of the amplitudes of the components (EPSP and PS) recorded from hippocampal slices. With this algorithm, the normalized mean square error (NMSE) was 2.48 % in PS amplitude calculations and 0.94 % in EPSP amplitude calculations [39]. In most plasticity studies, the slope value is preferred over the amplitude or other measures because at higher current intensity, especially if the electrodes are too close to the cell bodies. This is also because the slope is less subject to contamination from other sources of current flow [43]. Table 7 shows the NMSE values obtained using our proposed method. The NMSE values of all fEPSP parameters calculated with our proposed method were lower than the error obtained by Gholmieh et al. in Experiment 1. In Experiment 2, the PS amplitude fell within the acceptable range (lower than the 5 % error threshold set by Gholmieh et al.). However, the EPSP slope exceeded the 5 % error limit, indicating a potential issue. To identify the noise level at which the issue began, the authors calculated the NMSE at each noise level independently. The results are presented in Supplementary Table 5. The PS amplitude remained within the acceptable range (below the 5 % error threshold established by Gholmieh et al.) for all noisy data, except for the 5 dB SNR group. Conversely, the EPSP slope parameter exceeded the acceptable

limit (above the 5 % error threshold) in both the 5-dB and 10-dB SNR groups. Gholmieh's study, which provides the reference for the error limits, employed in vitro electrophysiological recordings on hippocampal slices that are inherently free from noise. This distinction should be considered when interpreting these error values in the context of noisy data. The main purpose of this comparison was to identify acceptable error limits because no standard exists for evaluating fEPSP parameters.

In the second study in the literature, the LSTM network and expert systems were used to extract the fEPSP parameters [20]. The study consists of two stages. In the first stage, an expert system was developed to eliminate contaminated HFP signals and generate labeled HFP signals. In the second stage, a LSTM network model was developed for segmenting the HFP signal using data tagged by an expert system. The segmentation accuracy of the LSTM model they developed is 96.73 %. No performance measurements other than accuracy were used in Zhao et al.'s study. The segmentation accuracy of our proposed BiLSTM network model was better than previous studies, with 99.6 % for Experiment 1 and 99.37 % for Experiment 2. Apart from the BiLSTM models that gave the best results, the accuracy of the other models obtained with LSTM and GRU networks was also higher than that obtained by Zhao et al.

While evaluating two studies based on a single accuracy parameter is generally challenging, several key differences stand out between our and Zhao et al. models: input data size, maximum epoch size, number of RNN units, and optimization algorithm. Our study utilized an input data size that was roughly five times larger than that employed by Zhao et al. While dataset size significantly affects deep learning model training, high data volume alone does not guarantee improved accuracy; model complexity must also be considered [53]. While Zhao et al.'s dataset was smaller, their maximum epoch size was six times larger, meaning that both models used roughly the same amount of data for training. In addition, the LSTM model performs as well for small data sizes as it does for high data sizes [54]. Our study found optimal performance with 400 or more RNN-based units, whereas Zhao et al. limited their experiments to 150 RNN-based units, achieving their best results at 100. This lower number of units may restrict the model's complexity and limit its ability

Table 8

Mean and standard deviation of performance measures obtained from repeated holdout evaluation.

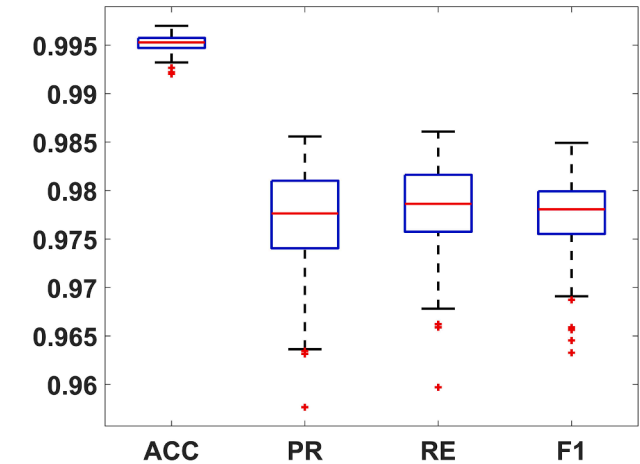|	ACC (%)	PR (%)	RE (%)	F1 (%)
Mean	99.5	97.6	97.8	97.7
Standard Deviation	0.09	0.54	0.48	0.41

Fig. 8. Box plot of performance measures obtained from repeated holdout evaluation (N=100).

to learn intricate relationships within the data. Notably, Zhao et al. used the SGDM optimization algorithm, which is more sensitive to hyperparameter tuning than our chosen ADAM algorithm [55]. These factors, combined with the smaller dataset and lower number of units, raise concerns about potential overfitting in their model. In summary, while it is challenging to definitively determine the reason for the success of our work, the ADAM optimizer and larger units appear to have a potentially higher capacity for learning complex patterns.

3.5. Generalizability of the model

The performance of the BiLSTM network model was assessed using the repeated holdout method, which is a robust technique for evaluating classification models [56,57]. To implement this method, 100 separate models were trained and tested using randomized 80/20 training-test data splits. Performance measures were calculated for each model and then averaged across all 100 runs. The mean and standard deviation of these measures are presented in Table 8, and Fig. 8 shows a box plot of their distribution. By analyzing these results, we can conclude that the model has a high generalizability capacity. This is due to the low standard deviation of the performance parameters, which indicates consistent performance across different data splits. In addition, the absence of extreme outliers in the box plot further supports this conclusion.

4. Conclusion

In this study, a new model based on deep learning was proposed to separate hippocampal HFP signals into stimulation, fEPSP, population amplitude, and decay phase segments to automatically calculate the EPSP slope and PS amplitudes used in fEPSP studies. Among the RNN-based models, the BiLSTM-based model has shown high success in segmenting both noise and clean HFP signals. Additionally, using the segmentation values of both the proposed model and the expert neuroscientist, the EPSP slope and PS amplitude values were calculated and compared. The comparison resulted in, an acceptable mean absolute percentage error was found for both parameters. The generalization properties of the model were examined using the repeated holdout

method, and high generalization capacity was observed. These investigations show that the proposed model can be easily used in studies on HFP. Thus, the time spent by neuroscientists to find HFP signal components will be reduced, and possible errors and subjectivity will be prevented.

Data statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author statement

The author confirms the sole responsibility for the conception of the study, presented results and manuscript preparation.

CRediT authorship contribution statement

Sabri Altunkaya: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The author declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbe.2024.09.005.

References

[1] Schomer DL, Lopes da Silva FH, editors. Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Wolters Kluwer: China; 2011.
[2] McColgan T, Liu J, Kuokkanen PT, Carr CE, Wagner H, Kempter R. Dipolar extracellular potentials generated by axonal projections. Elife 2017;6.
[3] Buzsaki G, Anastassiou CA, Koch C. The origin of extracellular fields and currents–EEG, ECoG LFP and spikes. Nat Rev Neurosci 2012;13(6):407–20.
[4] Rudy JW, editor. The Neurobiology of Learning and Memory. USA: Sinauer Associates; 2013.
[5] Lomo T. Patterns of activation in a monosynaptic cortical pathway: the perforant path input to the dentate area of the hippocampal. Exp Brain Res 1971;12:18–45.
[6] Lømo T. Potentiation of monosynaptic EPSPs in the perforant path-dentate granule cell synapse. Exp Brain Res 1971;12(1):46–63.
[7] Dash MB, Ajayi S, Folsom L, Gold PE, Korol DL. Spontaneous infraslow fluctuations modulate hippocampal EPSP-PS coupling. eNeuro 2018;5(1).
[8] Queiroz CM, Gorter JA, Lopes da Silva FH, Wadman WJ. Dynamics of evoked local field potentials in the hippocampus of epileptic rats with spontaneous seizures. J Neurophysiol 2009;101(3):1588–97.
[9] Tamura R, Eifuku S, Uwano T, Sugimori M, Uchiyama K, Ono T. A method for recording evoked local field potentials in the primate dentate gyrus in vivo. Hippocampus 2011;21(5):565–74.
[10] Liu XJ, Huang FS, Huang C, Yang ZM, Feng XZ. Analysis of high-frequency stimulation-evoked synaptic plasticity in mouse hippocampal CA1 region. Acta Physiologica Sinica 2008;60(2):284–91.
[11] Altunkaya M, Babur E, Barutcu Ö, Süer C, Dursun N. The expression of neurodegeneration-related genes in the hippocampus of hypothyroid rats following long-term potentiation. Eurasian J Med 2024;56(1):21–6.
[12] Sepahvand T, Nazari N, Qin T, Rajani V, Yuan Q. Olfactory threat extinction in the piriform cortex: an age-dependent employment of NMDA receptor-dependent long-term depression. PNAS Neurosci 2023;120(44):e2309986120.
[13] Abraham WC. Metaplasticity: tuning synapses and networks for plasticity. Nat Rev Neurosci 2008;9(5):387.

[14] Tan B, Tufan E, Barutçu Ö, Aslan-Gülpınar E, Dursun N, Süer C. Phosphorylation of tau protein based on the activity of kinases and phosphatases in various forms of synaptic plasticity. Physiol Int 2024;111(1):97–123.

[15] Babur E, Altunkaya M, Tufan E, Süer C, Dursun N. Hyperthyroidism-induced upregulation of neurodegeneration-related gene expression in metaplasticity-induced hippocampus. Neuroendocrinology 2024:400–10.

[16] Süer C, Dolu N, Artis AS, Sahin L, Yilmaz A, Cetin A. The effects of long-term sleep deprivation on the long-term potentiation in the dentate gyrus and brain oxidation status in rats. Neurosci Res 2011;70(1):71–7.

[17] Bliss TV, Cooke SF. Long-term potentiation and long-term depression: a clinical perspective. Clinics 2011;66(Suppl 1):3–17.

[18] O'Boyle MP, Do V, Derrick BE, Claiborne BJ. In vivo recordings of long-term potentiation and long-term depression in the dentate gyrus of the neonatal rat. J Neurophysiol 2004;91(2):613–22.

[19] Rubega M, Cecchetto C, Vassanelli S, Sparacino G. Algorithm and software to automatically identify latency and amplitude features of local field potentials recorded in electrophysiological investigation. Source Code Biol Med 2017;12:3.

[20] Zhao S, Shang Y, Yang Z, Xiao X, Zhang J, Zhang T. Application of expert system and LSTM in extracting index of synaptic plasticity. Cogn Neurodyn 2021;15(2):253–63.

[21] Nolde JM, Carnagarin R, Lugo-Gavidia LM, Azzam O, Kiuchi MG, Robinson S, et al. Autoencoded deep features for semi-automatic, weakly supervised physiological signal labelling. Comput Biol Med 2022;143.

[22] Zhao C, Li J, Guo Y. SleepContextNet: a temporal context network for automatic sleep staging based single-channel EEG. Comput Methods Programs Biomed 2022;220:106806.

[23] Gaugel S, Reichert M. PrecTime: a deep learning architecture for precise time series segmentation in industrial manufacturing operations. Eng Appl Artif Intel 2023;122:106078.

[24] Wang Z, Wang JL, Chen MY, Yang W, Fu R. Deep regression network with sequential constraint for wearable ECG characteristic point location. IEEE Access 2023;11:63487–95.

[25] Urtnasan E, Park JU, Joo EY, Lee KJ. Deep convolutional recurrent model for automatic scoring sleep stages based on single-lead ECG signal. Diagnostics 2022;12(5):1235.

[26] Londhe AN, Atulkar M. Semantic segmentation of ECG waves using hybrid channel-mix convolutional and bidirectional LSTM. Biomed Signal Process Control 2021;63:102162.

[27] He Y, Li WY, Zhang WQ, Zhang S, Pi XT, Liu HY. Research on segmentation and classification of heart sound signals based on deep learning. Applied Sciences-Basel 2021;11(2):651.

[28] Huang L, Yan JJ, Cai SY, Guo R, Yan HX, Wang YQ. Automated segmentation of the systolic and diastolic phases in wrist pulse signal using long short-term memory network. Biomed Res Int 2022;2022:1–9.

[29] Argha A, Celler BG, Alinejad-Rokny H, Lovell NH. Blood pressure estimation from korotkoff sound signals using an end-to-end deep-learning-based algorithm. IEEE Trans Instrum Meas 2022;71:1–10.

[30] Lu J, Yan C, Li J, Liu C. Sleep staging based on single-channel EEG and EOG with Tiny U-Net. Comput Biol Med 2023;163:107127.

[31] Xia K, Huang J, Wang H. LSTM-CNN architecture for human activity recognition. IEEE Access 2020;8:56855–66.

[32] Chen T-B, Soo V-W. A comparative study of recurrent neural network architectures on learning temporal sequences. In: Proceedings of International Conference on Neural Networks (ICNN'96). 1996.

[33] M.I. Jordan, editors. Advances in Psychology. Holland; 1997.

[34] Schuster M, Paliwal KK. Bidirectional recurrent neural networks. IEEE Trans Signal Process 1997;45(11):2673–81.

[35] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–80.

[36] Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw 2005;18(5):602–10.

[37] Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In: 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014.

[38] Anderson WW, Collingridge GL. The LTP Program: a data acquisition program for on-line analysis of long-term potentiation and other synaptic events. J Neurosci Methods 2001;108(1):71–83.

[39] Gholmieh G, Courellis S, Dimoka A, Wills JD, LaCoss J, Granacki JJ, et al. An algorithm for real-time extraction of population EPSP and population spike amplitudes from hippocampal field potential recordings. J Neurosci Methods 2004;136(2):111–21.

[40] Paxinos G, Watson C, editors. The rat brain in stereotaxic coordinates. USA: Academic Press; 2006.

[41] Barutçu Ö, Süer C, Dursun N, Tufan E, Gülpınar EA, Tan B. Insulin-induced long-term potentiation in the dentate gyrus of hippocampal formation. Psychoneuroendocrinology 2023;157:106343.

[42] Bliss TV, Lomo T. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. J Physiol 1973;232(2):331–56.

[43] Graziane N, Dong Y, Walz W, editors. Electrophysiological Analysis of Synaptic Transmission. New York: Springer; 2016.

[44] Madhavan PG. A new recurrent neural network learning algorithm for time series prediction. J Intell Syst 1997;7(1–2):103–16.

[45] Gers FA, Schmidhuber J, Cummins F. Learning to forget: continual prediction with LSTM. In: In: Ninth International Conference on Artificial Neural Networks ICANN; 1999. p. 99.

[46] Kusuma S, Jothi KR. ECG signals-based automated diagnosis of congestive heart failure using Deep CNN and LSTM architecture. Biocybernetics and Biomedical Engineering 2022;42(1):247–57.

[47] Kiperwasser E, Goldberg Y. Simple and accurate dependency parsing using bidirectional LSTM feature representations. Transactions of the Association for Computational Linguistics 2016;4:313–27.

[48] Wang H, Zhang Y, Liang J, Liu L. DAFA-BiLSTM: deep autoregression feature augmented bidirectional LSTM network for time series prediction. Neural Netw 2023;157:240–56.

[49] Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. Deep Learning and Representation Learning Workshop; 2014. in NIPS 2014.

[50] Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. In: International Conference on Learning Representations (ICLR). 2014.

[51] Cui X, Chen Z, Yin F. Multi-objective based multi-channel speech enhancement with BiLSTM network. Appl Acoust 2021;177:107927.

[52] S. Siami-Namini, N. Tavakoli and A.S. Namin. The Performance of LSTM and BiLSTM in Forecasting Time Series. In: IEEE International Conference on Big Data. 2019.

[53] Zhu X, Vondrick C, Fowlkes CC, Ramanan D. Do we need more training data? Int J Comput Vis 2016;119(1):76–92.

[54] Boulmaiz T, Guermoui M, Boutaghane H. Impact of training data size on the LSTM performances for rainfall–runoff modeling. Modeling Earth Systems and Environment 2020;6(4):2153–64.

[55] Soydaner D. A Comparison of optimization algorithms for deep learning. Int J Pattern Recognit Artif Intell 2020;34(13):2052013.

[56] Kim J-H. Estimating classification error rate: repeated cross-validation, repeated hold-out and bootstrap. Comput Stat Data Anal 2009;53(11):3735–45.

[57] Tomas B, Marcel Jr J, Pavel K, Marcel J, editors. Advances in Data Mining Knowledge Discovery and Applications. IntechOpen: Rijeka; 2012.