Received 12 October 2023, accepted 28 November 2023, date of publication 5 December 2023,
date of current version 12 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3339665

RESEARCH ARTICLE

Electroencephalogram Analysis Method to Detect Unspoken Answers to Questions Using Multistage Neural Networks

SHIN-ICHI ITO, (Member, IEEE), MOMOYO ITO, (Member, IEEE), AND MINORU FUKUMI
Graduate School of Technology, Industrial and Social Science, Tokushima University, Tokushima 770-8506, Japan

Corresponding author: Shin-Ichi Ito (s.ito@tokushima-u.ac.jp)

This work was supported in part by Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP17K12768 and Grant JP22K12221.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Research Ethics Committee, Division of Science and Technology, and Division of Bioscience and Bioindustry, Graduate School of Technology, Industrial and Social Sciences in Tokushima University.

ABSTRACT Brain–computer interfaces (BCI) facilitate communication between the human brain and computational systems, additionally offering mechanisms for environmental control to enhance human life. The current study focused on the application of BCI for communication support, especially in detecting unspoken answers to questions. Utilizing a multistage neural network (MSNN) replete with convolutional and pooling layers, the proposed method comprises a threefold approach: electroencephalogram (EEG) measurements, EEG feature extraction, and answer classification. The EEG signals of the participants are captured as they mentally respond with "yes" or "no" to the posed questions. Feature extraction was achieved through an MSNN composed of three distinct convolutional neural network models. The first model discriminates between the EEG signals with and without discernible noise artifacts, whereas the subsequent two models are designated for feature extraction from EEG signals with or without such noise artifacts. Furthermore, a support vector machine is employed to classify the answers to the questions. The proposed method was validated via experiments using authentic EEG data. The mean and standard deviation values for sensitivity and precision of the proposed method were 99.6% and 0.2%, respectively. These findings demonstrate the viability of attaining high accuracy in a BCI by preliminarily segregating the EEG signals based on the presence or absence of artifact noise and underscore the stability of such classification. Thus, the proposed method manifests prospective advantages of separating EEG signals characterized by noise artifacts for enhanced BCI performance.

INDEX TERMS Answer to question, convolutional neural networks, electroencephalogram, multistage neural networks, personal model, support vector machine.

I. INTRODUCTION

The faculty of communication is indispensable for human life. Individuals rendered speechless owing to afflictions such as stroke or post-neurosurgery encounter significant challenges in expressing themselves. Systems capable of facilitating nonverbal communication, specifically, the provision of answering questions, could substantially improve the therapeutic and rehabilitative processes for these individuals.

Historically, several paradigms for communication support have been advanced. These range from eye-tracking systems used for keyboard operations [1] to comprehensive brain activity analyses [2], [3], [4], [5], [6], [7], [8], [9], [10], [11]. Active systems like eye-tracking necessitate user training and may thus pose difficulties to certain individuals. In contrast, passive systems that interpret brain activity to decipher intent require no such learning curve. This study focuses on techniques that enable communication through a brain–computer interface (BCI).

A BCI functions by monitoring cerebral activity during the mental visualization of a motor action, such as

The associate editor coordinating the review of this manuscript and approving it for publication was Li Zhang.

lifting a hand, or speaking a specific word, e.g., "yes" or "no." To date, several methods have been developed to analyze brain activity, including functional near-infrared spectroscopy (fNIRS) and electroencephalograms (EEGs). In fNIRS-based approaches, motor imagery [2] and alterations in mental state [3], [4] are examined through metrics like oxygenated hemoglobin (HbO) levels to distinguish between "yes" and "no" responses. Naito et al. [3] classified answers to questions based on variations in the mental state (brain active = yes; relaxed state = no) by analyzing the alterations in the cerebral blood volume. Chiarelli et al. [2] classified the answers to questions based on motor imagery of right- and left-hand motions using a deep neural network (DNN). Naseer et al. [4] classified answers based on the changes in the mental state (mental arithmetic = yes; relaxed state = no) by analyzing changes in HbO using a support vector machine (SVM).

The equipment required for fNIRS is expensive and inconvenient to wear because the device may need to be worn for over 30 min. Therefore, although fNIRS-based assistive systems do not require any special training or prior practice to operate, their usage and application in daily life is limited. Moreover, direct recording of brain activity while imagining a word has not been reported in any previous fNIRS study. Conversely, EEG devices are generally easy to wear on a daily basis, and several studies have reported various approaches involving motor imagery EEG (MI-EEG) [5], [6], [7], [8], [9], [10] and word imagery [11], [12], [13], [14], [15]. Support systems employing EEG-based BCI [2], [5], [6], [16], [17] can be used for communication, rehabilitation, and environmental system and electric wheelchair controls.

EEG signals have been analyzed to create BCI systems using analysis techniques such as EEG feature extraction from an event-related potential (ERP) response [11], the power spectrum and spectral centroid [13], [17], [18], [19], Bayesian-based EEG feature extraction techniques [20], principal component analysis [21], [22], [23], [24], independent component analysis (ICA) [15], [23], [25], [26], [27], EEG pattern classifiers, artificial neural networks (ANN) [14], [23], [28], k-nearest neighbor algorithm [29], [30], linear discriminant analysis (LDA) [15], [31], [32], [33], SVM [12], [13], [21], [34], [35], [36], [37], self-organizing maps [38], and fuzzy entropy [39]. Furthermore, over the past few years, DNN, its improved models [7], [8], [40], [41], [42], [43], [44], and convolutional neural networks (CNN) [9], [10], [45], [46], [47], [48], [49], [50], [51] have been employed for EEG feature extraction and pattern classification. In particular, EEGNET was proposed as the dedicated EEG analysis model based on artificial intelligence techniques [46], [52], [53], [54], [55].

Prior research on MI-EEG has employed various computational models for classifying motor imagery. For instance, Lu et al. [7] and Sturm et al. [8] employed DNN models, while Yang et al. [6], Amin et al. [9], Tabar and Halici [10], Alazrai et al. [50], and Moussa et al. [51] favored CNN models. Furthermore, EEGNET was employed by Lawhern [46], Zhu et al. [55], and Strahnen and Kessler [54]. These models have demonstrated superior classification accuracy in comparison to other methods, including DNNs. However, the majority of MI-EEG methods require training sessions, which can be cumbersome for the user. Will an individual think through motor imagery when communicating with other individuals? Will an individual perform mental arithmetic operation while communicating with other people? We believe that these operations are unnatural even for controlling a BCI system. Theoretically, if EEG signals associated with responses to question can be directly extracted, the BCI systems will not require training and practice for their operation. Leydecker et al. [11] employed an ERP classification technique for word imagery, achieving an ERP classification accuracy of at least 60%. Adama and Bogdan [12] utilized spectral edge frequency extraction combined with random forests and SVM, attaining a classification accuracy range of 50.41–67.94% for the binary "yes" or "no" responses. Choi and Kim [13] applied event-related spectral perturbation extraction in tandem with SVM and reported a mean classification accuracy of 86.03%, with a standard deviation of 8.69. Sereshken et al. [14] used the ADJUST algorithm, discrete wavelet transform, and artificial neural networks to achieve a mean classification accuracy of 63.2% with a standard deviation of 6.4. Finally, Kim et al. [15] used ICA and LDA, achieving a classification accuracy of 70% or less. These varied outcomes indicate that current methods are still deficient in establishing a reliable BCI system for answer classification to support human communication.

EEG signals exhibit interparticipant and intraparticipant variabilities, which may compromise the stability of classification systems unless the user undergoes adequate training and practice. To mitigate the effects of intra-participant differences, personalized models can be developed for individual users. For addressing inter-participant variability, prior studies have employed the extraction of EEG features and have successfully employed CNN models to enhance feature extraction. Prior research has reported CNN models with excellent results; these may improve the EEG feature extraction even when the signals include inter-participant differences.

Sensors capturing EEG signals are strategically placed on the frontal and temporal cortices, regions of the brain implicated in cognitive functions such as reasoning, judgment, and decision-making. However, these signals often include substantial noise, specifically myogenic artifacts like eye blinks. Several methods, including ICA, frequency-based filters, and deep learning (DL) techniques, are employed to eliminate such artifacts. Although ICA enables separation of artifacts from EEG signals, identifying the artifact component remains challenging. Frequency-based filters are less effective when the EEG features and the artifacts share overlapping frequency bands. DL techniques facilitate feature extraction by automatically learning from EEG datasets, although it is challenging to separate noisy from non-noisy EEG signals for accurate analysis.

In this study, we propose a novel method for classifying question-related responses. Sensor placements were determined according to the international 10-10 system and targeted the frontal and temporal cortices. We developed personalized models for EEG signal analysis to diminish the impact of inter-participant differences. Consequently, our approach obviates the need for extensive training or practice by the BCI user. Multistage neural networks (MSNN) were utilized to segregate noisy and non-noisy EEG signals prior to feature extraction. This separation aims to enhance the accuracy of EEG analysis, and we hypothesize that common features can still be extracted via CNN models even if the BCI user has not undergone training or practice. Finally, SVM were employed for the classification of responses [12], [13], [34], [35], [36], [37], [49].

The primary contributions and salient points of the present research are summarized as follows:

1) We obtained high levels of accuracy on direct analysis of EEG signals associated with the imagined response of "yes" or "no" to a question.

2) Proposed models for developing a BCI system that does not require prior practice for effective operation.

3) Used multistage neural networks to separate EEG signals with obvious artifact noise related to blink from those without these noises and separately extract the features before classifying the responses.

4) Used CNNs to extract features and classify the question responses, with previous studies demonstrating the effectiveness of using CNN models for EEG feature extraction and noise removal [9], [10], [45], [46], [47], [48], [49], [50], [51], [56].

II. PROPOSED METHOD

In the present study, we introduce a MSNN designed to classify binary responses, namely, "yes" or "no", to questions. The methodological framework encompasses EEG measurement, EEG feature extraction, and subsequent classification of the responses. The MSNN, equipped with convolutional and pooling layers, serves as the cornerstone for EEG feature extraction. A CNN is deployed initially to segregate noisy from non-noisy EEG signals. Subsequently, two additional CNNs are utilized to extract features from these separated categories of signals. For the final classification of responses, a SVM is employed. The efficiency of the proposed method was experimentally substantiated using a simple electroencephalograph to record EEG signals as the participants non-verbally responded to questions through imagined "yes" and "no" answers.

A. EEG MEASUREMENT

For EEG data collection, 14-channel electrodes were positioned to capture activity from the frontal cortex [49], [57], [58], [59]. Two reference electrodes were affixed behind each ear on the temporal bone, while the exploring electrodes were located according to the international 10-10 system at positions AF3, F7, F3, T7, T8, F4, F8, and AF4, associated with frontal cortex activity. These positions are associated with frontal cortex activity. Table 1 and Figure 1 delineate the specifications of the electroencephalograph and the positioning of the electrodes, respectively.

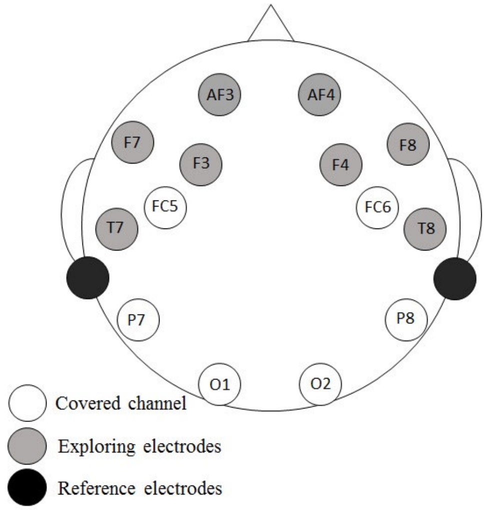


FIGURE 1. Sensing positions: white, gray, and black circles represent channels of electroencephalograph, exploring electrodes, and reference electrodes, respectively.

TABLE 1. Electroencephalograph specifications.

Number of channels	14 channels: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 Reference electrodes: attached over the bone just behind each lobe
Sampling rate	128 SPS
Resolution	14-bit
Bandwidth	0.16–43 Hz, notch filters: 50 and/or 60 Hz
Dynamic range	8,400 µV(pp)

B. EEG FEATURE EXTRACTION

The feature extraction stage involves multiple steps. Initially, EEG signals are normalized to serve as appropriate input data for the CNNs. This is followed by the separation of the EEG signals into two categories: those containing noise and those devoid of it. Thereafter, feature extraction is performed using CNNs. The network structures designated for this task is illustrated in Figure 2.

$$Nor(ch) = (EEGsig(ch) - Mini)/(Maxi - Mini), \tag{1}$$

where Nor and ch denote the normalized EEG signal and channel (AF3, F3, F7, T7, T8, F8, F4, and AF4), respectively. $EEGsig$, $Mini$, and $Maxi$ indicate the EEG signals recoded on each channel, the minimum value of the EEG signals of all channels, and the maximum value of the EEG signals from all channels, respectively. Based on the sensing positions in the international 10-10 system, the proposed method arranged the EEG signals in the order T7, F7, F3, AF3, AF4, F4, F8, and T8.

In the noise separation phase, the CNNs separated the EEG signals into those with and without obvious noise artifacts

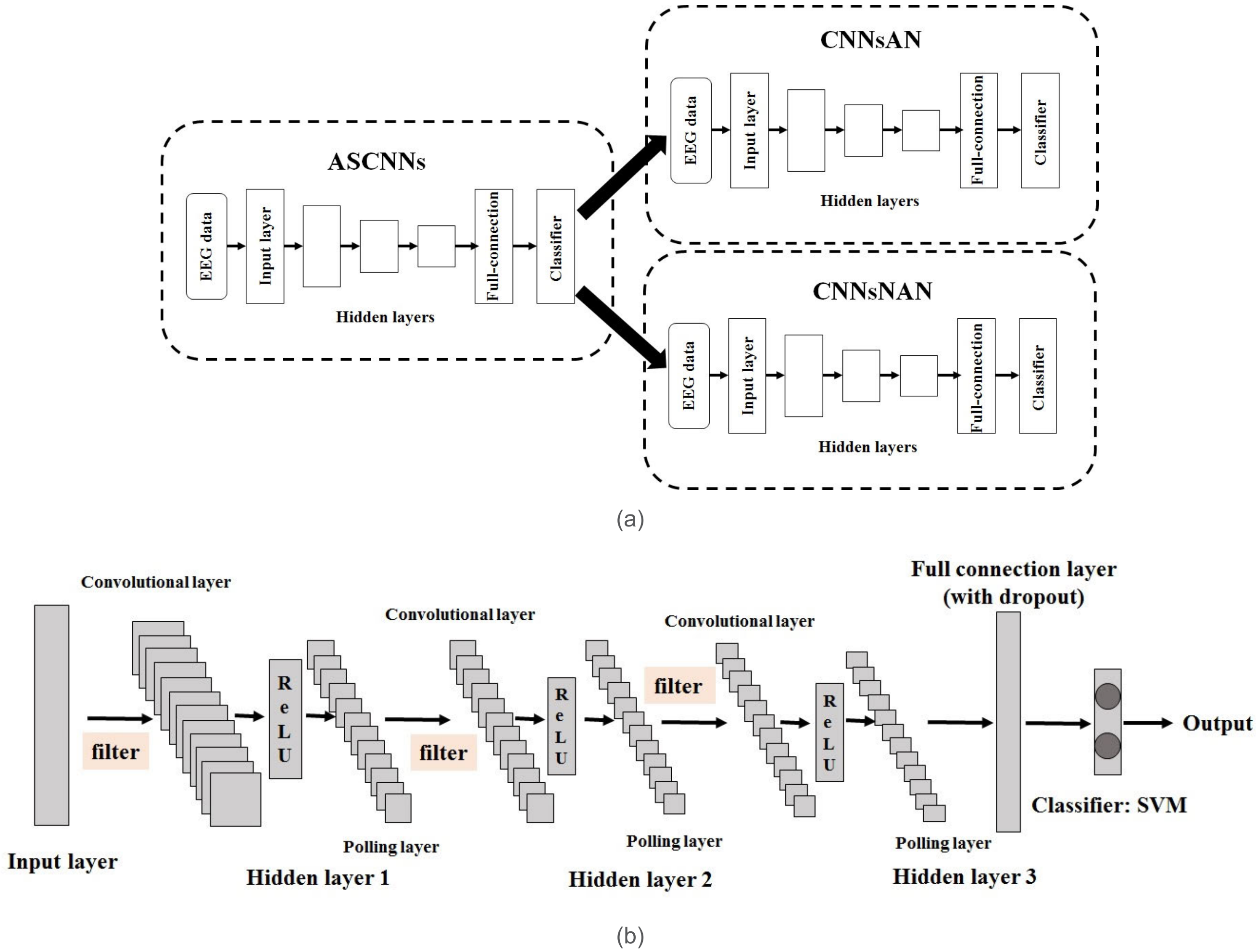


FIGURE 2. Structures and details of MSNN with convolutional and pooling layers: (a) MSNN (b) CNNs.

(ASCNNs). The CNNs included an input layer, three hidden layers, and a fully-connected layer. The hidden layer includes both convolutional and pooling layers. The EEG signal features relating to the question responses are extracted in the convolutional layer. Note that the convolutional layer uses an n × n filter, as depicted in Figure 2 (b). The convolutional layers serve to extract features that differentiate noisy from non-noisy EEG signals through the amalgamation of EEG signal data. Subsequently, the pooling layers compress this data. The fully-connected layer synthesizes the features, which are pivotal for segregating noisy and non-noisy signals, and includes a dropout function, with a rate of P% to protect against overfitting.

The feature extraction stage incorporates two specialized CNN models: CNNsAN and CNNsNAN. CNNsAN is designed to extract features from EEG signals that manifest discernible noise artifacts, while CNNsNAN targets EEG signals that are largely devoid of such artifacts. Each of these models consists of an input layer, three hidden layers, and a fully-connected layer. Within the hidden layers, convolutional and pooling layers are integrated. In terms of filters, the convolutional layers in CNNsAN and CNNsNAN employ m × m and l × l, respectively. The pooling layers contribute to noise reduction by compressing the EEG data.

In the fully-connected layer, the extracted EEG features are further processed and synthesized. To mitigate overfitting, dropout functions are integrated into these layers, with dropout rates of the hidden layer and includes a dropout function (dropout for CNNsAN and CNNsNAN is Q% and R%, respectively). In summary, CNNsAN and CNNsNAN are calibrated to learn and extract features from EEG signals with and without prominent noise, respectively. These features are subsequently instrumental in the classification of answers.

C. CLASSIFICATION OF ANSWERS TO QUESTIONS

Question responses are categorized utilizing EEG data: CNNsAN for data with noise and CNNsNAN for noise-free data. This study assesses the proposed methodology by computing both sensitivity and precision, employing k-fold cross-validation for testing. Sensitivity and precision are computed as follows:

$$sensitivity = TP/(TP + FN), \tag{2}$$

$$precision = TP/(TP + FP) \tag{3}$$

where *TP*, *FN* and *FP* denote true positive, false negative and false positive, respectively.

III. EXPERIMENTS

The participant cohort comprised nine healthy volunteers from Tokushima University, Japan, with a mean age of 22.7 years. Following a comprehensive briefing regarding the experiment, written informed consent was acquired from all participants in accordance with the Declaration of Helsinki. The study received approval from Research Ethics Committee, Division of Science and Technology, and Division of Bioscience and Bioindustry, Graduate School of Technology, Industrial and Social Sciences in Tokushima University. During the experimental sessions, participants were seated and wore an electroencephalograph, performing tasks with eyes open and closed. Eye opening led to a notable inclusion of blink-related noise in the EEG signals. Conversely, closing of eyes substantially eliminated such noise artifacts. The temporal structure of the experiments is illustrated in Figure 3. However, the EEG signals could not effectively include obvious noise, i.e., blinking noise, when the participants' eyes were closed. The time course of the experiments is displayed in Figure 3. EEG signals were recorded for 2 s, whereas the participant imagined a yes/no answer to a question (Imagine Ans. in Figure 3). Each participant answered a set of 25 questions randomly selected from a larger set of 30 questions (Table 2). The full set of 30 questions included nine questions that were easy to answer (Clear), 16 questions related to physical conditions (Conditions), and five questions related to demands (Demand). Note that participants did not practice for the experiments. In total, each participant was presented with eight sets of questions, with a maximum of two sets per day.

The analyzed EEG signals were segmented into intervals of 128 sample points with an incremental shift of ten sample points in 2 s (Imagine Ans. in Figure 3). This paper acquired ten data records per question.

The CNN parameters were determined *via* trial and error. The size of the input layer was 8 × 128 (number of channels × length of EEG signals), and the filter size *n* of hidden layers 1, 2, and 3 in the ASCNNs was 3, 3, and 2, respectively.

TABLE 2. 30 questions.

Clear	Do you understand who I am?
Clear	Can you hear me?
Clear	Do you know what your name is?
Clear	Did you eat breakfast?
Clear	Do you remember what you ate for dinner yesterday?
Clear	Do you understand who you are?
Clear	Do you understand where you are?
Clear	Did you drink a caffeine beverage today?
Clear	Did you drink alcoholic drinks?
Conditions	Did you catch a cold?
Conditions	Are you hungry?
Conditions	Do you feel pain in your butt?
Conditions	Are you sleepy now?
Conditions	Are you tired?
Conditions	Do you have sore place?
Conditions	Do you have a headache?
Conditions	Do you feel out of it?
Conditions	Do you have an appetite?
Conditions	Do you have a stuffed-up nose?
Conditions	Do you have a cough?
Conditions	Aren't you sleep deprived?
Conditions	Is your posture bad?
Conditions	Did you have a sound sleep?
Conditions	Have you taken medicine?
Conditions	Do you push yourself?
Demand	Would you like something to eat?
Demand	Would you like something to drink?
Demand	Would you like to go to the restroom?
Demand	Do you have a TV program you want to watch?
Demand	Would you like to smoke?

The filter sizes *m* and *l* for CNNsAN and CNNsNAN exhibited the same *n*. The number of convolutional layers in hidden layers 1, 2, and 3 in the ASCNNs, CNNsAN, and CNNsNAN were 50, 50, and 20, respectively. A max-pooling algorithm was employed in the pooling layer. P, Q, and R of the dropout rate in the fully-connected layers in the ASCNNs, CNNsAN, and CNNsNAN were all set to 50. The number of units in the fully-connected layer of the CNNs was set to 2,000. A linear SVM classifier was used, and the *k*-value for k-fold cross-validation was set to four, corresponding to the minimum number of days required for participation. Training and test datasets were mutually exclusive in terms of the data collection day for each participant. Although the architecture and parameters remained consistent across participants, the learning datasets varied.

Numerous studies have proposed enhanced CNN models for motor imagery classification, showcasing significant improvements over existing frameworks. Amin et al. [9]

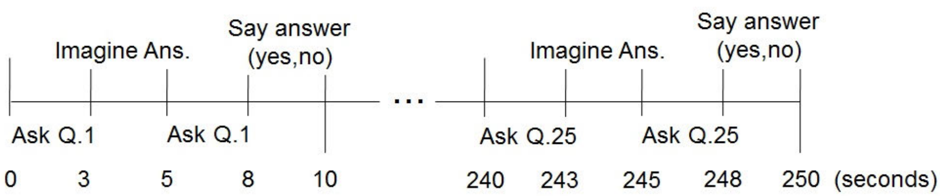


FIGURE 3. Time course of the experiments. "Ask Q." indicates that the system is presenting a question to the participant. "Imagine Ans." indicates the point at which the participant should imagine their "yes" or "no" answer to the question. "Say answer (yes or no)" indicates the point at which the participant should respond with a verbal "yes" or "no" to the question. Each participant answered a set of 25 questions.

TABLE 3. Total numbers of segmented EEG signals from YES/NO answers to questions for each participant.

		SubA	SubB	SubC	SubD	SubE	SubF	SubG	SubH	SubI	Total
Included	Yes	550	560	550	620	510	630	560	420	430	4,830
	No	440	440	450	380	490	360	440	580	570	4,150
Not included	Yes	530	560	530	620	530	620	550	420	440	4,800
	No	470	440	470	380	470	360	450	580	560	4,060

pitted their optimized CNN model against other CNN and DNN models, while Chen et al. [45] validated improved CNNs by comparing them with LDA, SVM, and other CNN models. Yang et al. [44] validated their CNN model by comparison with k-NN, SVM, LDA, and other CNN models. Moussa et al. [51] demonstrated the effectiveness of CNN models through comparison with LSTM, ANN and SVM. Similarly, Lim et al. [52] validated the efficiency of the simplified CNN models. Several authors demonstrated effectively improved DNN models (DBN): Results from Lu et al. [7], Ma et al. [43], Sturm et al. [8], and Yin et al. [44] suggested DNN to be a superior technique for BCIs. In our earlier work [49], we utilized CNN and SVM to discern yes-or-no responses using noise-free EEG data collected with the participants' eyes closed. Consequently, the present study incorporates a comparative analysis between CNN and DNN models. Moreover, Lawhern [46], Zhu et al. [55], and Strahnen et al. [54] employed EEGNET, whereas Zhu et al. [53] and Lim et al. [52] employed EEGNET for SSVEP. Despite our study not focusing on MI-EEG or SSVEP, we included EEGNET as a comparative model owing to its high performance in EEG analysis and specialization in artificial intelligence-driven EEG analysis. Upon comparing CNN models, the parameter settings remained consistent with those of our proposed method. The DNN architectures consisted of an input layer, seven fully-connected layers interspersed with a dropout layer, and an output layer. The fully-connected layers featured 2,000 units at both the initial and terminal layers, while the intermediary layers contained 1,000 units. The EEGNET model employed in our study is based on the architecture proposed by Lawhern [46]. Both the input and classification layers were identical to those in the comparative CNN models.

The input data for the MSNNs were generated using C/C++, and the CNN models for the MSNN, CNN, and DNN models for the comparative analyses were created using MATLAB 2018a, 2019a, 2022a and Neural Networks Toolbox.

IV. RESULTS AND DISCUSSIONS

Table 3 shows the number of EEG signals for each participant, where "Included" and "Not included" represent EEG signals with and without obvious noise, respectively. As displayed, the proportion of "yes" and "no" answers differed for each participant because each participant answered the questions differently. The total number of segmented EEG signals corresponding to answers was ~2,000 for each participant, equally divided between "yes" and "no" answers. In participants A and F, the number of EEG signals was less than 2000 because they did not answer certain questions.

Table 4 shows the mean and standard deviation for the sensitivity and precision of answer classifications derived from DNNs, DNNs+SVM, CNNs, CNNs+SVM, and EEGNET models. Note that the participants enumerated in Table 4 correspond with those identified in Table 3. The term "four-class" signifies that EEG signals were partitioned into categories—those with artifacts and those without—prior to their classification. Conversely, "mix" denotes a combination of EEG signals, with and without discernible artifacts, that could be classified. Utilizing the proposed method, we achieved mean and standard deviation values for sensitivity and precision of 99.6% and 0.2%, respectively. Under the "mix" condition, the mean values for sensitivity and precision using CNNs, DNNs, and EEGNET were equal to or exceeded 94%, 70%, and 75%, respectively. The standard deviation of the accuracy rate for these models was approximately 2% for CNNs, 15% for DNNs, and 7% for EEGNET. In the "four-class" scenario, the mean accuracy rates for CNNs, DNNs, and EEGNET were no less than 90%, 50%, and 70%, respectively, with maximum standard deviations being approximately 5.6%, 14.1%, and 5.1%, respectively. The proposed methodology yielded the highest mean (99.6%) and the lowest standard deviation (0.2%) in terms of sensitivity and precision. Initially, our approach utilized ASCNNs to segregate noisy from non-noisy EEG signals. Subsequently, CNN models tailored for artifact noise (CNNsAN) and those without (CNNsNAN) were employed for the classification of the respective signal types into "yes" and "no" categories. It is worth emphasizing that CNNsAN was not participated in training using non-noisy EEG signals, and conversely, CNNsNAN was not trained with noisy EEG signals. Consequently, we posit that the extraction of EEG features for the task of response classification is unequivocally straightforward.

Concerning the sensitivity and precision metrics for ASCNNs, CNNsAN, and CNNsNAN, Table 5 provides a comprehensive analysis of the recognition accuracy achieved via our proposed methodology. Specifically, "Noise Separation," "Included Artifact," and "Not Included Artifact" represent outcomes derived from the use of ASCNNs, CNNsAN, and CNNsNAN, respectively, focusing on the same participant cohort as enumerated in Table 3. The mean

TABLE 4. Mean and SD for sensitivity and precision for answer classification and mean for sensitivity and precision for each participant (%).

			mean	S. D.	SubA	SubB	SubC	SubD	SubE	SubF	SubG	SubH	SubI
mix	DNNs*2	Sen	72.3	13.1	86.3	55.8	63.6	87.7	50	75.7	72.5	88.4	70.8
		Pre	74.1	12.5	87.8	63.2	59.9	85.9	52	78.6	78	89.1	72.9
	DNNs*2+SVM	Sen	75.3	14.1	91.6	58.4	66.4	91.7	50.5	75.8	77.9	91.7	74
		Pre	77.8	11.2	91.5	76.1	61.9	90	59.7	78.2	77.2	91.8	74.2
	CNNs	Sen	94.5	1.8	94.9	93.4	95.8	97.2	91	95.5	92.3	95	95.6
		Pre	94.9	1.7	95.1	93.7	95.9	97.4	91.9	95.7	92.3	95.9	95.8
	CNNs +SVM	Sen	97	1	97.3	97.7	97.7	98	96.5	98	94.9	96.3	96.3
		Pre	97	1	97.1	97.9	97.6	97.9	96.8	97.9	94.8	96.7	96.2
	EEGNET	Sen	75.9	6.8	80.6	75.9	72.7	78.8	74.3	63.1	87.8	80.5	69.4
		Pre	80.3	4	81.2	75.6	74.2	79.3	79.5	83.6	88.8	80.1	80.7
four-class	DNNs*2	Sen	50.6	11	52.9	50.5	45.7	77.8	34.8	44.1	45.9	53.1	50.3
		Pre	58.3	10.5	59	55	46.5	80	43.4	59.9	53.5	69	58.3
	DNNs*2+SVM*1	Sen	54.2	12.3	58.9	50.8	47.7	84.1	38.9	44.3	49	59.3	54.9
		Pre	59	12.4	60	46.8	54.8	86.5	42.5	56.6	52.2	70.1	62
	CNNs	Sen	92.6	5.6	79	90.8	88.8	93.2	97.5	95.2	98.2	96.2	94.2
		Pre	93.6	3.9	85.4	92	89.5	93.3	97.6	95.6	98.2	96.1	94.8
	CNNs +SVM*1	Sen	93.6	4.5	82.7	91.3	91.6	95.4	97.6	95.7	97.2	97.2	94
		Pre	93.8	4.3	82.9	91.6	92.2	95.5	97.7	95.4	97.2	97.1	94.2
	EEGNET	Sen	71.4	5.1	77.6	73.8	70	71.4	75.3	65.6	75.2	73	60.6
		Pre	78.5	2.7	79.3	77.8	76.9	77.8	81.3	82.4	81.2	76.5	73.5
Proposed method		Sen	**99.6**	**0.2**	99.6	99.6	99.6	99.5	99.4	100	99.7	99.5	99.4
		Pre	**99.6**	**0.2**	99.6	99.6	99.7	99.6	99.5	100	99.7	99.6	99.5

*1 This study employed the multiclass SVM for four-class classification.
*2 The structure of the DNNs was one input layer, seven fully-connected layers, and one output layer with a Softmax layer.

TABLE 5. Mean and sd of sensitivity and precision for ascnns, cnnsan and cnnsnan in the proposed method (%).

		Mean	S. D.	SubA	SubB	SubC	SubD	SubE	SubF	SubG	SubH	SubI
Proposed method	Sen	99.6	0.2	99.6	99.6	99.6	99.5	99.4	100	99.7	99.5	99.4
	Pre	99.6	0.2	99.6	99.6	99.7	99.6	99.5	100	99.7	99.6	99.5
Noise separation (for ASCNNs)	Sen	100	0.1	99.9	100	100	100	100	100	100	100	99.8
	Pre	100	0.1	99.9	100	100	100	100	100	100	100	99.8
Included artifact (for CNNsAN)	Sen	99.7	0.2	99.7	99.6	99.7	99.5	99.4	100	99.7	99.5	99.9
	Pre	99.7	0.2	99.8	99.6	99.7	99.6	99.4	100	99.7	99.7	99.9
Not included artifact (for CNNsNAN)	Sen	99.6	0.2	99.7	99.6	99.7	99.5	99.5	100	99.7	99.4	99.4
	Pre	99.6	0.2	99.7	99.6	99.7	99.5	99.5	100	99.7	99.5	99.5

and standard deviation for both sensitivity and precision of the noise separation rate were recorded as 100% and 0.1%, respectively. For EEG signals containing overt noise artifacts, the mean and standard deviation for sensitivity and precision were 99.7% and 0.2%, respectively. Conversely, for EEG signals devoid of such artifacts, these metrics were 99.6% and 0.2%, respectively. These statistical outcomes underscore the imperative to discriminate between noisy and non-noisy EEG signals when both types coexist in the dataset.

In Figure 4, samples of recorded EEG signals, both with and without noise artifacts (obvious artifact noise related to blink), are illustrated. Our analysis confirmed that the signals identified as noisy exhibited conspicuous artifacts, while those marked as non-noisy were devoid of such anomalies. In a comparative analysis of noisy and non-noisy signals originating from the same response category, evident disparities between them were discernible, as depicted in Figure 4. If signals with identical semantic meaning but varying

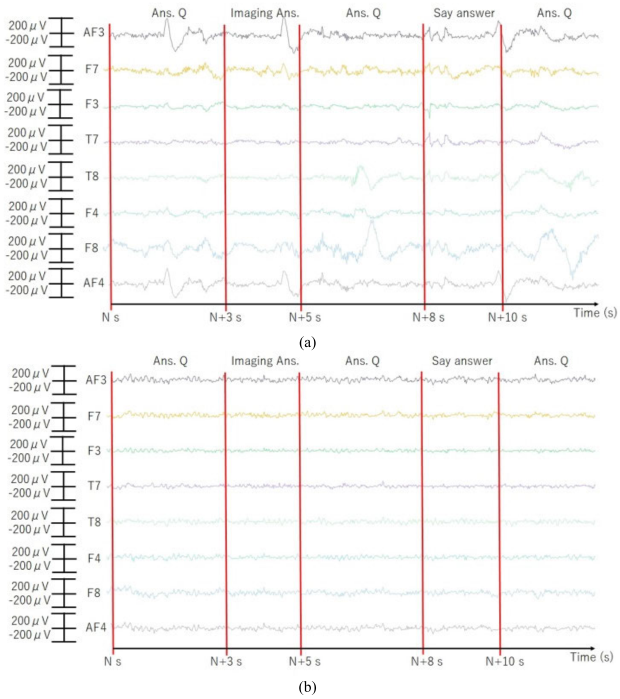


FIGURE 4. Samples of the recorded EEG signals. (a) and (b) represent signals with and without obvious noise artifacts, respectively. Ans. Q, Imagine Ans. and Say answer are in line with Figure 3. AF3, F7, F3, T7, T8, F4, F8, and AF4 are in accordance with Figure 1.

amplitude due to the presence or absence of noise artifacts are employed for CNN model training, the extraction of relevant EEG features could become convoluted. Nonetheless, pre-segregating the signals based on noise presence facilitates more efficient feature extraction during model training.

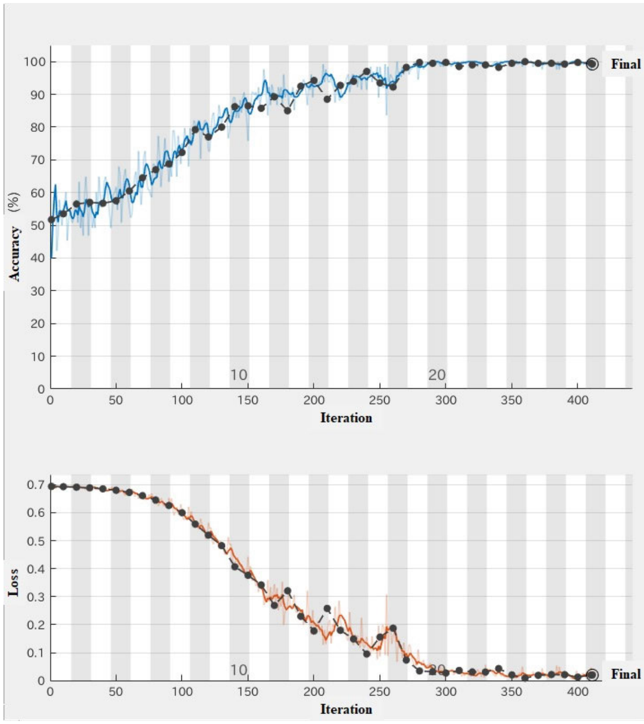
Our approach implemented distinct CNN models to handle EEG signals, categorizing them based on whether they exhibited artifact noise, notably those related to eye blinking. This differentiation enabled more efficacious extraction of EEG features pertinent to ''yes'' and ''no'' response classification, leading to improved standard deviations of 0.2% or less. In terms of architectural design, our employed CNN models—ASCNNs, CNNsAN, and CNNsNAN—featured compact structures, including one input layer, three hidden layers equipped with convolutional and pooling layers, and one fully-connected layer. The input dataset size for each participant approximated 2,000, with a bifurcated class pattern of ''yes'' and ''no.''

Figure 5 elucidates the loss rates experienced during the learning phase for EEG signals utilizing ASCNNs, CNNsAN, and CNNsNAN. Notably, these loss rates exhibited a clear convergence pattern. The compact nature of our CNN structures did not inhibit successful feature extraction. Inadequate CNN structures could result in overfitting or underfitting, manifesting as elevated loss rates or diminished recognition accuracy, respectively. Furthermore, the volume of input data allocated to each CNN model was sufficiently robust to facilitate learning, feature extraction, response classification, and validation of the method's effectiveness. The empirical findings affirm the capacity of the proposed methodology to achieve high recognition accuracy, thus substantiating its utility.

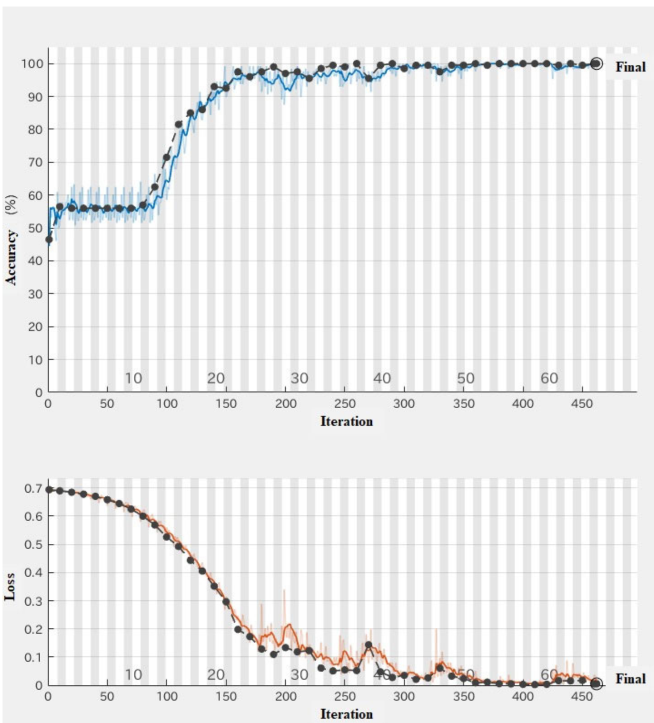
In the present study, CNNs were utilized for both noise separation and answer classification. Upon evaluation, both CNNs and EEGNET outperformed DNNs in terms of mean and standard deviation values. DNN models exhibited large standard deviations, leading to inconsistent results. For the ''mix'' condition, a mean recognition accuracy of 90% or higher indicates the efficacy of CNNs in feature extraction and noise removal. Likewise, in the ''four-class'' scenario, CNNs demonstrated proficiency in these tasks. Contrarily, DNNs and EEGNET encountered difficulties in feature extraction and noise removal when tasked with four-class classification. The classification results for the ''four-class'' case, verified using DNN, EEGNET, and CNN models, are shown in Figure 6, where the bottom row represents sensitivity and the rightmost column represents precision. Class numbers 1 and 2 indicate ''yes'' and ''no,'' respectively, with noise, whereas 3 and 4 indicate ''yes'' and ''no,'' respectively, without noise. In the DNN models, cases of ''yes'' with and without obvious noise related to blink were incorrectly classified as belonging to one of the other three classes. Cases of ''no'' with obvious noise were incorrectly classified as either ''yes'' or ''no'' without obvious noise. These results suggest that it is not easy to extract EEG features to classify four classes using DNN models.

In the EEGNET, the cases of ''yes'' with and without obvious noise related to blink were incorrectly classified as belonging to one of the remaining three classes. Cases of ''no'' with and without obvious noise related to blink were incorrectly classified as belonging to one of the remaining three classes. Although the EEGNET performance was not being influenced by noise or artifact signals [46], these results suggest that it cannot easily extract the EEG features to classify the four classes and separate the EEG signals with obvious noise related to blink from without noisy EEG signals using EEGNET in this case. Conversely, CNN models exhibited better discrimination between noisy and non-noisy EEG signals. Specifically, instances of ''no'' with obvious noise were erroneously classified as ''yes'' with obvious noise, while instances of ''yes'' lacking noise were incorrectly categorized as ''no'' without noise. These misclassifications notwithstanding, CNN models still demonstrate an inherent advantage. Owing to their convolutional and pooling layers, these models excel at reducing noise, extracting features, and compressing information, functionalities absent in DNN models, which primarily depend on the manipulation of network weight values for these tasks. Consequently, we ascertain that, in the context of this research, CNN models are unequivocally superior to DNN models. Moreover, the CNN models of Lim et al. [52] exhibited high performance for emotion recognition.

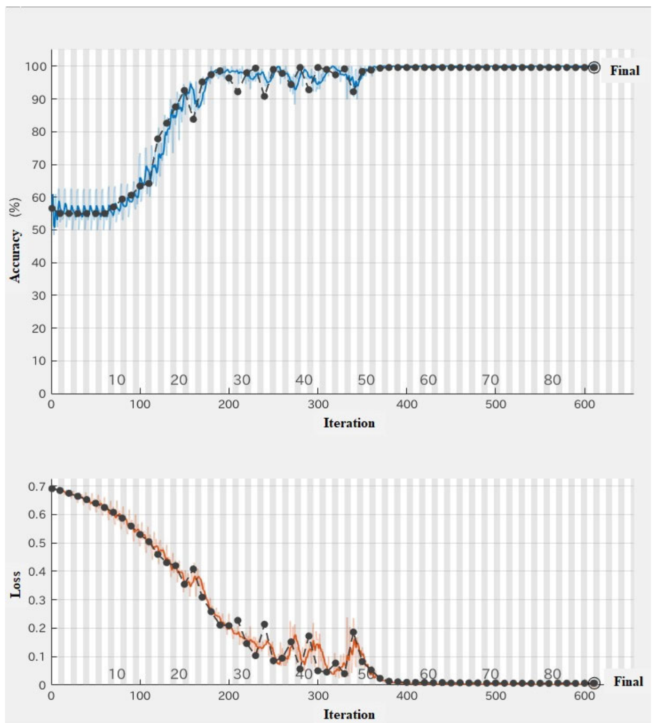
In the context of CNN models, both the mean sensitivity and precision exceeded 90% for the methods under consideration. It is crucial to highlight that the EEG signals



(a)



(b)



(c)

FIGURE 5. Samples of recognition accuracy and loss rate during learning. (a) to (c) are on ASCNNs, CNNsAN and CNNsNAN, respectively.

utilized in these experiments were linked to activities in the frontal cortex and that participants did not undergo specialized training in BCI operations. Such outcomes imply that frontal cortex-related EEG signals are pertinent for generating responses to questions. Moreover, these signals can be analyzed for categorizing such responses without necessitating specialized training. This assertion is supported by Naseer et al. [4], who demonstrated the feasibility of detecting question responses through analysis of frontal cortex activity using fNIRS. Similarly, this study confirms that "yes" and

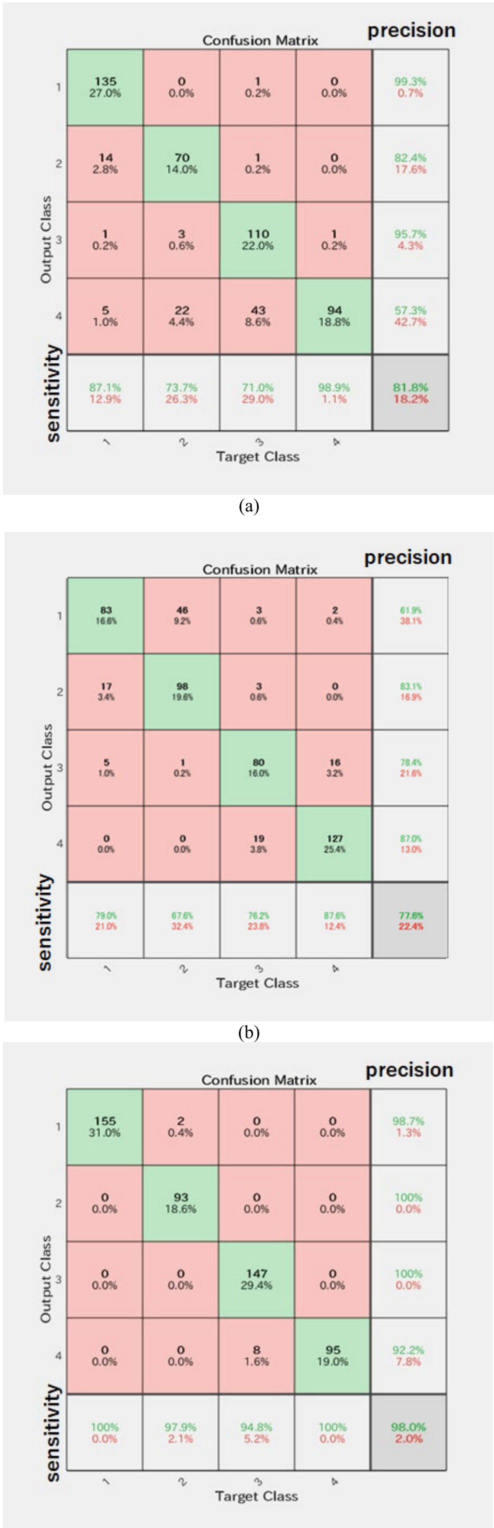


FIGURE 6. Samples of confusion matrix for four-class classification. (a) to (c) are the results on DNN+SVM, EEGNET and CNN+SVM, respectively. In the class number, 1–4 correspond to "yes" with obvious artifact, "no" with obvious artifact, "yes" not included obvious artifact and "no" not included obvious artifact, respectively.

"no" answers can be classified through the analysis of EEG signals from both the frontal and temporal cortices.

When compared with the CNN+SVM approach, which previously held the record for highest answer classification accuracy, our proposed method exhibited a 2% improvement in both mean sensitivity and precision. These findings underscore the imperative of segregating EEG signals obvious containing evident artifacts from those free of noise. Unlike CNN+SVM, our method successfully managed this separation. Additionally, the standard deviations for sensitivity and precision improved by 0.8%, suggesting that the proposed method facilitates stable recognition accuracy. Consequently, the intricacies of EEG feature extraction for answer classification can be reduced, and stable outcomes can be achieved when training algorithms on datasets comprising EEG signals that have been categorically separated based on the presence or absence of overt artifacts. This study thus validates the effectiveness of our proposed methodology in classifying "yes" and "no" responses to questions, negating the need for specialized training or practice on the part of the participants.

V. CONCLUSION

In the present study, we introduce a methodology for the direct examination of EEG signals generated during the mental conception of "yes" or "no" responses to queries. This method forms the basis of a novel brain–Computer Interface (BCI) system designed to facilitate human communication. The proposed approach encompasses EEG data collection, feature extraction, and answer classification, leveraging EEG signals generated as participants internally formulate "yes" or "no" answers without verbal articulation. We advocate the employment of MSNNs furnished with convolutional and pooling layers for EEG feature extraction. Specifically, we utilize ASCNNs, as well as CNN models designed for EEG signals with (CNNsAN) and without (CNNsNAN) manifest noise artifacts. For the task of answer categorization, a SVM was employed to distinguish between "yes" and "no" responses.

To validate the utility of our methodology, empirical tests were conducted using authentic EEG data. The acquired mean and standard deviation for both sensitivity and precision were 99.6% and 0.2%, respectively. These findings underscore the critical importance of differentiating between noisy and clean EEG signals, particularly when both types coexist in the collected data. Comparative analysis with conventional CNN, DNN, and EEGNET models revealed a mean recognition accuracy surpassing 90% for all CNN-based methods. Such outcomes indicate that EEG signals, particularly those emanating from the frontal and temporal cortices, are instrumental in formulating answers to questions. Furthermore, these answers can be reliably classified through the analysis of corresponding EEG signals, obviating the need for specialized BCI training. In relation to recognition accuracy, our proposed model exhibited a superior mean of 99.6% and a reduced standard deviation of 0.2% compared to alternative methods.

Although the immediate objective of this investigation was to categorize simple "yes" or "no" answers, it is pertinent to note that many queries necessitate more nuanced responses, such as expressions of physical discomfort or varying emotional states. These complex responses present significant classification challenges.

However, several questions demand more complex responses, such as "a bit of headache," "very tired," and "not too bad," which are extremely difficult to classify. In future, we intend to classify the degrees of answers—such as "very good," "good," borderline, "bad," and "severe"—by adopting and/or improving the proposed method.

REFERENCES

[1] H. Cecotti, "A multimodal gaze-controlled virtual keyboard," *IEEE Trans. Hum.-Mach. Syst.*, vol. 46, no. 4, pp. 601–606, Aug. 2016, doi: 10.1109/THMS.2016.2537749.

[2] A. M. Chiarelli, P. Croce, A. Merla, and F. Zappasodi, "Deep learning for hybrid EEG-fNIRS brain–computer interface: Application to motor imagery classification," *J. Neural Eng.*, vol. 15, no. 3, Apr. 2018, Art. no. 036028, doi: 10.1088/1741-2552/aaaf82.

[3] M. Naito, Y. Michioka, K. Ozawa, Y. Ito, M. Kiguchi, and T. Kanazawa, "A communication means for totally locked-in ALS patients based on changes in cerebral blood volume measured with near-infrared light," *IEICE Trans. Inf. Syst.*, vol. 90, no. 7, pp. 1028–1037, Jul. 2007, doi: 10.1093/ietisy/e90-d.7.1028.

[4] N. Naseer, M. J. Hong, and K.-S. Hong, "Online binary decision decoding using functional near-infrared spectroscopy for the development of brain–computer interface," *Exp. Brain Res.*, vol. 232, no. 2, pp. 555–564, Feb. 2014, doi: 10.1007/s00221-013-3764-1.

[5] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, Jun. 2018, Art. no. 031005, doi: 10.1088/1741-2552/aab2f2.

[6] J. Yang, S. Yao, and J. Wang, "Deep fusion feature learning network for MI-EEG classification," *IEEE Access*, vol. 6, pp. 79050–79059, 2018, doi: 10.1109/ACCESS.2018.2877452.

[7] N. Lu, T. Li, X. Ren, and H. Miao, "A deep learning scheme for motor imagery classification based on restricted Boltzmann machines," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 566–576, Jun. 2017, doi: 10.1109/TNSRE.2016.2601240.

[8] I. Sturm, S. Lapuschkin, W. Samek, and K.-R. Müller, "Interpretable deep neural networks for single-trial EEG classification," *J. Neurosci. Methods*, vol. 274, pp. 141–145, Dec. 2016, doi: 10.1016/j.jneumeth.2016.10.008.

[9] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Bencherif, and M. S. Hossain, "Multilevel weighted feature fusion using convolutional neural networks for EEG motor imagery classification," *IEEE Access*, vol. 7, pp. 18940–18950, 2019, doi: 10.1109/ACCESS.2019.2895688.

[10] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, no. 1, Feb. 2017, Art. no. 016003, doi: 10.1088/1741-2560/14/1/016003.

[11] A. Leydecker, F. Bießmann, and S. Fazli, "Single-trials ERPs predict correct answers to intelligence test questions," in *Proc. Int. Workshop Pattern Recognit. Neuroimaging*, Jun. 2014, pp. 1–4, doi: 10.1109/PRNI.2014.6858528.

[12] S. Adama and M. Bogdan, "Yes/no classification of EEG data from CLIS patients," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 5727–5732, doi: 10.1109/EMBC46164.2021.9629716.

[13] J. W. Choi and K. H. Kim, "Covert intention to answer 'yes' or 'no' can be decoded from single-trial electroencephalograms (EEGs)," *Comput. Intell. Neurosci.*, vol. 2019, pp. 1–11, Jul. 2019, doi: 10.1155/2019/4259369.

[14] A. Rezazadeh Sereshkeh, R. Trott, A. Bricout, and T. Chau, "EEG classification of covert speech using regularized neural networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 12, pp. 2292–2300, Dec. 2017, doi: 10.1109/TASLP.2017.2758164.

[15] H. Kim, N. Yoshimura, and Y. Koike, "Investigation of delayed response during real-time cursor control using electroencephalography," *J. Healthcare Eng.*, vol. 2020, pp. 1–9, Feb. 2020, doi: 10.1155/2020/1418437.

[16] Z. Emami and T. Chau, "Investigating the effects of visual distractors on the performance of a motor imagery brain–computer interface," *Clin. Neurophysiol.*, vol. 129, no. 6, pp. 1268–1275, Jun. 2018, doi: 10.1016/j.clinph.2018.03.015.

[17] A. Roman-Gonzalez, N. I. Vargas-Cuentas, M. Hoyos, J. Diaz, and M. Zimic, "Brain computer interface to answer yes-no questions," in *Proc. 2nd Int. Conf. Bio-Eng. Smart Technol. (BioSMART)*, Aug. 2017, pp. 1–4, doi: 10.1109/BIOSMART.2017.8095323.

[18] A. Bajoulvand, R. Z. Marandi, M. R. Daliri, and S. H. Sabzpoushan, "Analysis of folk music preference of people from different ethnic groups using kernel-based methods on EEG signals," *Appl. Math. Comput.*, vol. 307, pp. 62–70, Aug. 2017, doi: 10.1016/j.amc.2017.02.042.

[19] W. Park, D.-H. Kim, S.-P. Kim, J.-H. Lee, and L. Kim, "Gamma EEG correlates of haptic preferences for a dial interface," *IEEE Access*, vol. 6, pp. 22324–22331, 2018, doi: 10.1109/ACCESS.2018.2827023.

[20] H. Kang and S. Choi, "Bayesian common spatial patterns for multi-subject EEG classification," *Neural Netw.*, vol. 57, pp. 39–50, Sep. 2014, doi: 10.1016/j.neunet.2014.05.012.

[21] H. Lee and S. Choi, "PCA+HMM+SVM for EEG pattern classification," in *Proc. 7th Int. Symp. Signal Process. Appl.*, 2003, pp. 541–544.

[22] C. Wang, J. Zou, J. Zhang, M. Wang, and R. Wang, "Feature extraction and recognition of epileptiform activity in EEG by combining PCA with ApEn," *Cognit. Neurodyn.*, vol. 4, no. 3, pp. 233–240, Sep. 2010, doi: 10.1007/s11571-010-9120-2.

[23] T. Hoya, G. Hori, H. Bakardjian, S. Nishimura, T. Suzuki, Y. Miyawaki, A. Funase, and J. Cao, "Classification of single trial EEG signals by a combined principal + independent component analysis and probabilistic neural network approach," in *Proc. Integr. Comput.-Aided Eng.*, Jan. 2003, pp. 197–202.

[24] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation," *Sci. World J.*, vol. 2014, pp. 1–10, Sep. 2014, doi: 10.1155/2014/627892.

[25] N. Mammone, F. La Foresta, and F. C. Morabito, "Automatic artifact rejection from multichannel scalp EEG by wavelet ICA," *IEEE Sensors J.*, vol. 12, no. 3, pp. 533–542, Mar. 2012, doi: 10.1109/JSEN.2011.2115236.

[26] A. Kachenoura, L. Albera, L. Senhadji, and P. Comon, "ICA: A potential tool for BCI systems," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 57–68, Dec. 2008, doi: 10.1109/msp.2008.4408442.

[27] X. Bai, X. Wang, S. Zheng, and M. Yu, "The offline feature extraction of four-class motor imagery EEG based on ICA and wavelet-CSP," in *Proc. 33rd Chin. Control Conf.*, Jul. 2014, pp. 7189–7194, doi: 10.1109/ChiCC.2014.6896188.

[28] A. A. Azamimi, "Development of EEG based epileptic detection using artificial neural network," in *Proc. Int. Conf. Biomed. Eng.*, 2012, pp. 605–610.

[29] J. F. Borisoff, S. G. Mason, A. Bashashati, and G. E. Birch, "Brain–computer interface design for asynchronous control applications: Improvements to the LF-ASD asynchronous brain switch," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 985–992, Jun. 2004, doi: 10.1109/TBME.2004.827078.

[30] L.-W. Ko, R. K. Chikara, Y.-C. Lee, and W.-C. Lin, "Exploration of user's mental state changes during performing brain–computer interface," *Sensors*, vol. 20, no. 11, p. 3169, Jun. 2020, doi: 10.3390/s20113169.

[31] B. Blankertz, "Classifying single trial EEG: Towards brain computer interfacing," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 1, 2002, pp. 157–164.

[32] W. S. Rayens, "Discriminant analysis and statistical pattern recognition," *Technometrics*, vol. 35, no. 3, pp. 324–326, Aug. 1993.

[33] K.-R. Müller, C. W. Anderson, and G. E. Birch, "Linear and nonlinear methods for brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 165–169, Jun. 2003, doi: 10.1109/TNSRE.2003.814484.

[34] T. Felzer and B. Freisieben, "Analyzing EEG signals using the probability estimating guarded neural classifier," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 4, pp. 361–371, Dec. 2003, doi: 10.1109/TNSRE.2003.819785.

[35] H. Kang, "Composite common spatial pattern for participant-to-participant transfer," *IEEE Signal Process. Lett.*, vol. 16, no. 8, pp. 683–686, Aug. 2009.

[36] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, Feb. 2011, doi: 10.1109/TBME.2010.2082539.

[37] H. Morioka, A. Kanemura, J.-I. Hirayama, M. Shikauchi, T. Ogawa, S. Ikeda, M. Kawanabe, and S. Ishii, "Learning a common dictionary for subject-transfer decoding with resting calibration," *NeuroImage*, vol. 111, pp. 167–178, May 2015, doi: 10.1016/j.neuroimage.2015.02.015.

[38] R. Khosrowabadi, H. C. Quek, A. Wahab, and K. K. Ang, "EEG-based emotion recognition using self-organizing map for boundary detection," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 4242–4245.

[39] Z. Cao, C.-T. Lin, K.-L. Lai, L.-W. Ko, J.-T. King, K.-K. Liao, J.-L. Fuh, and S.-J. Wang, "Extraction of SSVEPs-based inherent fuzzy entropy using a wearable headband EEG in migraine patients," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 1, pp. 14–27, Jan. 2020, doi: 10.1109/TFUZZ.2019.2905823.

[40] S. Ding, N. Zhang, X. Xu, L. Guo, and J. Zhang, "Deep extreme learning machine and its application in EEG classification," *Math. Problems Eng.*, vol. 2015, pp. 1–11, Jan. 2015, doi: 10.1155/2015/129021.

[41] X. An, "A deep learning method for classification of EEG data based on motor imagery," in *Proc. ICIC*, Aug. 2014, pp. 203–210.

[42] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006, doi: 10.1162/neco.2006.18.7.1527.

[43] T. Ma, H. Li, H. Yang, X. Lv, P. Li, T. Liu, D. Yao, and P. Xu, "The extraction of motion-onset VEP BCI features based on deep learning and compressed sensing," *J. Neurosci. Methods*, vol. 275, pp. 80–92, Jan. 2017, doi: 10.1016/j.jneumeth.2016.11.002.

[44] Z. Yin, M. Zhao, Y. Wang, J. Yang, and J. Zhang, "Recognition of emotions using multimodal physiological signals and an ensemble deep learning model," *Comput. Methods Programs Biomed.*, vol. 140, pp. 93–110, Mar. 2017, doi: 10.1016/j.cmpb.2016.12.005.

[45] J. X. Chen, P. W. Zhang, Z. J. Mao, Y. F. Huang, D. M. Jiang, and Y. N. Zhang, "Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks," *IEEE Access*, vol. 7, pp. 44317–44328, 2019, doi: 10.1109/ACCESS.2019.2908285.

[46] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.

[47] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017, doi: 10.1002/hbm.23730.

[48] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain–computer interface using convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5619–5629, Nov. 2018, doi: 10.1109/TNNLS.2018.2789927.

[49] S.-I. Ito, M. Ito, S. Fujisawa, and M. Fukumi, "Electroencephalogram data for classifying answers to questions with neural networks and support vector machine," *Int. J. Signal Process. Syst.*, vol. 7, no. 4, pp. 118–122, Dec. 2019, doi: 10.18178/ijsps.7.4.118-122.

[50] R. Alazrai, M. Abuhijleh, H. Alwanni, and M. I. Daoud, "A deep learning framework for decoding motor imagery tasks of the same hand using EEG signals," *IEEE Access*, vol. 7, pp. 109612–109627, 2019, doi: 10.1109/ACCESS.2019.2934018.

[51] M. M. Moussa, U. Tariq, F. Al-Shargie, and H. Al-Nashash, "Discriminating fake and real smiles using electroencephalogram signals with convolutional neural networks," *IEEE Access*, vol. 10, pp. 81020–81030, 2022, doi: 10.1109/ACCESS.2022.3195028.

[52] H. Lim, S. Kim, and J. Ku, "Distraction classification during target tracking tasks involving target and cursor flickering using EEGNet," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1113–1119, 2022, doi: 10.1109/TNSRE.2022.3168829.

[53] Y. Zhu, Y. Li, J. Lu, and P. Li, "EEGNet with ensemble learning to improve the cross-session classification of SSVEP based BCI from ear-EEG," *IEEE Access*, vol. 9, pp. 15295–15303, 2021.

[54] M. Strahnen and P. Kessler, "Investigation of a deep-learning based brain–computer interface with respect to a continuous control application," *IEEE Access*, vol. 10, pp. 131090–131100, 2022, doi: 10.1109/ACCESS.2022.3228164.

[55] H. Zhu, D. Forenzo, and B. He, "On the deep learning models for EEG-based brain–computer interface using motor imagery," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2283–2291, 2022, doi: 10.1109/TNSRE.2022.3198041.

[56] L. Farokhah, R. Sarno, and C. Fatichah, "Simplified 2D CNN architecture with channel selection for emotion recognition using EEG spectrogram," *IEEE Access*, vol. 11, pp. 46330–46343, 2023, doi: 10.1109/ACCESS.2023.3275565.

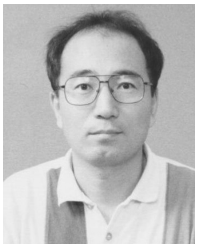
[57] G. B. S. Seco, G. J. L. Gerhardt, A. A. Biazotti, A. L. Molan, S. V. Schönwald, and J. L. Rybarczyk-Filho, "EEG alpha rhythm detection on a portable device," *Biomed. Signal Process. Control*, vol. 52, pp. 97–102, Jul. 2019, doi: 10.1016/j.bspc.2019.03.014.

[58] R. Majid Mehmood, R. Du, and H. J. Lee, "Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors," *IEEE Access*, vol. 5, pp. 14797–14806, 2017, doi: 10.1109/ACCESS.2017.2724555.

[59] H. Ullah, M. Uzair, A. Mahmood, M. Ullah, S. D. Khan, and F. A. Cheikh, "Internal emotion classification using EEG signal with sparse discriminative ensemble," *IEEE Access*, vol. 7, pp. 40144–40153, 2019, doi: 10.1109/ACCESS.2019.2904400.

SHIN-ICHI ITO (Member, IEEE) received the B.E. and M.E. degrees from Tokushima University, in 2002 and 2004, respectively, and the D.E. degree from the Tokyo University of Agriculture and Technology, in 2007. He was with Japan Gain the Summit Company Ltd. He was with the Tokyo University of Agriculture and Technology, as a System Engineer and a Specially Appointed Assistant Professor, in 2004 and 2007, respectively. He was an Assistant Professor with Tokushima University, from 2009 to 2019, where he has been an Associate Professor, since 2020. His current research interests include EEG analysis, bio-signal processing, and information visualization. He is a member of IEICE, JSMBE, and IEEJ.

MOMOYO ITO (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in computer science from Akita University, Japan, in 2005, 2007, and 2010, respectively. She was an Assistant Professor with Tokushima University, from 2010 to 2016, where she has been an Associate Professor, since 2016. Her current research interests include human behavior analysis and intelligent transportation systems for active safety. She is a member of JSAE, IEICE, and IPSJ.



MINORU FUKUMI received the B.E. and M.E. degrees from the University of Tokushima, in 1984 and 1987, respectively, and the Ph.D. degree from Kyoto University, in 1996. Since 1987, he has been with the Department of Information Science and Intelligent Systems, Tokushima University. In 2005, he became a Professor with the Department of Information Science and Intelligent Systems, Tokushima University. His research interests include human sensing, neural networks, and digital signal processing. He is a member of IEEJ, RISP, JSAI, and IEICE. He received the best paper awards from SICE, in 1995, and the Research Institute of Signal Processing, in 2011, Japan, and best paper awards from some international conferences.