

Article



# Automated Classification of Mental Arithmetic Tasks Using Recurrent Neural Network and Entropy Features Obtained from Multi-Channel EEG Signals

Abhishek Varshney<sup>1</sup>, Samit Kumar Ghosh<sup>1</sup>, Sibasankar Padhy<sup>2</sup>, Rajesh Kumar Tripathy<sup>1</sup> and U. Rajendra Acharya<sup>3,4,5,\*</sup>

- <sup>1</sup> Department of Electrical and Electronics Engineering, BITS-Pilani, Hyderabad Campus, Hyderabad 500078, India; f20170253@hyderabad.bits-pilani.ac.in (A.V.); samitnitrkl@gmail.com (S.K.G.); rajeshiitg13@gmail.com (R.K.T.)
- <sup>2</sup> Department of Electronics and Communication Engineering, IIIT Dharwad, Karnataka 580009, India; sibasankar@iiitdwd.ac.in
- <sup>3</sup> School of Engineering, Ngee Ann Polytechnic, Singapore 599489, Singapore
- <sup>4</sup> Department of Bioinformatics and Medical Engineering, Asia University, Taichung 41354, Taiwan
- <sup>5</sup> School of Management and Enterprise, University of Southern Queensland, Springfield 4300, Australia
- \* Correspondence: Rajendra\_Udyavara\_ACHARYA@np.edu.sg



Citation: Varshney, A.; Ghosh, S.K.; Padhy, S.; Tripathy, R.K.; Acharya, U.R. Automated Classification of Mental Arithmetic Tasks Using Recurrent Neural Network and Entropy Features Obtained from Multi-Channel EEG Signals. *Electronics* 2021, 10, 1079. https:// doi.org/10.3390/electronics10091079

Academic Editors: Teresa Orlowska-Kowalska and Cheng Siong Chin

Received: 14 March 2021 Accepted: 29 April 2021 Published: 2 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). Abstract: The automated classification of cognitive workload tasks based on the analysis of multichannel EEG signals is vital for human-computer interface (HCI) applications. In this paper, we propose a computerized approach for categorizing mental-arithmetic-based cognitive workload tasks using multi-channel electroencephalogram (EEG) signals. The approach evaluates various entropy features, such as the approximation entropy, sample entropy, permutation entropy, dispersion entropy, and slope entropy, from each channel of the EEG signal. These features were fed to various recurrent neural network (RNN) models, such as long-short term memory (LSTM), bidirectional LSTM (BLSTM), and gated recurrent unit (GRU), for the automated classification of mental-arithmeticbased cognitive workload tasks. Two cognitive workload classification strategies (bad mental arithmetic calculation (BMAC) vs. good mental arithmetic calculation (GMAC); and before mental arithmetic calculation (BFMAC) vs. during mental arithmetic calculation (DMAC)) are considered in this work. The approach was evaluated using the publicly available mental arithmetic taskbased EEG database. The results reveal that our proposed approach obtained classification accuracy values of 99.81%, 99.43%, and 99.81%, using the LSTM, BLSTM, and GRU-based RNN classifiers, respectively for the BMAC vs. GMAC cognitive workload classification strategy using all entropy features and a 10-fold cross-validation (CV) technique. The slope entropy features combined with each RNN-based model obtained higher classification accuracy compared with other entropy features for the classification of the BMAC vs. GMAC task. We obtained the average classification accuracy values of 99.39%, 99.44%, and 99.63% for the classification of the BFMAC vs. DMAC tasks, using the LSTM, BLSTM, and GRU classifiers with all entropy features and a hold-out CV scheme. Our developed automated mental arithmetic task system is ready to be tested with more databases for real-world applications.

**Keywords:** mental arithmetic task; multi-channel EEG; entropies; slope entropy; recurrent neural networks; accuracy

## 1. Introduction

The amount of mental effort performed by each person in response to certain cognitive tasks is called the cognitive workload [1]. The human brain produces different responses to various cognitive tasks, and these responses can be further investigated by analyzing the brain's electrical activity [2]. The information regarding the brain's electrical activity along both spatial and temporal directions can be visualized using multi-channel electroencephalogram (EEG) signals [3].

Multi-channel EEG signals have been used for various cognitive neuroscience applications, such as emotion recognition [4], drug effect monitoring [5], the categorization of sleep stages [6], sleep pathology detection [7], driver drowsiness detection [8], etc.

The mental arithmetic task classification is a type of cognitive workload categorization task that involves decision making, sequencing, the use of memory, and fact retrieval [9]. Mental arithmetic tasks can be classified into "bad" or "good" quality based on the type of mathematical calculations performed by the brain, and the responses can be observed in multi-channel EEG signals [10]. Mental arithmetic task classification provides an alternative method for the diagnosis of neurological ailments, such as autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD) [11,12].

The manual interpretation of this type of cognitive task based on the analysis of multi-channel EEG signals is a tedious process. Hence, automated algorithms based on the analysis of multi-channel EEG signals are required for the accurate classification of mental-arithmetic-based cognitive tasks [9]. The automated classification of mental-arithmetic-based cognitive workload tasks using various machine learning and deep learning-based methods is an active and interesting research topic in human–computer interface (HCI) and cognitive neuroscience.

**Related Works** In recent years, researchers have used various signal processing and classifier-based machine learning techniques for the automated classification of mentalarithmetic-based cognitive tasks. Fatimah et al. [13] extracted the  $L_2$  norm, mean, Shannon entropy, and energy features from the rhythms of each EEG channel. They employed a support vector machine (SVM) classifier for the classification of the cognitive workload classes, such as the before mental arithmetic calculation (BFMAC) or rest state and during mental arithmetic calculation (DMAC) or the active state.

From their study, an accuracy value of 95.80% was obtained using the decision treebased classification model. Similarly, in another study, the same authors considered the Fourier decomposition method to evaluate sub-band signals from the EEG data [14]. They extracted the variance, energy, and entropy features from the Fourier domain intrinsic band functions of EEG data and used an SVM classifier to classify the mental arithmetic tasks, such as BFMAC and DMAC, respectively, and reported an accuracy value of 98.60%. The stacked long short-term memory (LSTM)-based model was used to classify mental arithmetic tasks using spectral and instantaneous frequency features obtained from the multi-channel EEG signals [15].

A classification accuracy of 91.67% was reported using a stacked LSTM classifier for the BFMAC vs. DMAC classification task. Wang and Sourina [9] extracted features, such as the power of each rhythm, fractal, auto-regressive model coefficients, and statistical parameters, from different channels of EEG signals. They used principal component analysis (PCA) for the dimension reduction of the feature vector and SVM model to classify mental arithmetic tasks. The method based on multivariate autoregressive model-based analysis of EEG signal and SVM was also used for mental-arithmetic-based cognitive workload classification tasks [16].

The methods reported in [13–15] considered the classification schemes, such as BF-MAC vs. DMAC, using only the statistical and spectral features from EEG signals. Similarly, the method reported by [9] included a mental arithmetic task recognition framework using EEG signal features. In [16], the authors considered a three-class classification task, such as baseline vs. mental arithmetic vs. and mental letter, composed of EEG signal features.

The aforementioned works did not formulate the cognitive workload classification tasks, such as bad mental arithmetic calculations (BMAC) vs. good mental arithmetic calculations (GMAC), using the analysis of EEG signals. The above-reported methods only considered the time-domain and frequency-domain features of EEG signals for the mental arithmetic task classification. The state-space representation-based entropy features quantify non-linearity and randomness using EEG signal processing [17], and these fea-

tures have been used for various applications, such as the detection of generalized and partial epileptic seizures [17–19], emotion recognition [20], and brain-computer interface (BCI) [21] applications.

These non-linear entropy features, such as dispersion entropy [22], slope entropy [23], and other entropy measures [24,25], have not been explored for the mental-arithmetic-based cognitive task classification using EEG signals. Recently, deep learning approaches, such as the convolutional neural network (CNN) and recurrent neural network (RNN), have been extensively used for EEG signal processing applications [26]. A RNN is a deep neural network model used for the analysis of sequential data in natural language processing (NLP) and speech recognition applications [27].

This type of network explores the long-range dependencies for the modeling of timeseries data [28]. It considers the information from the previous state and input to evaluate the output at the present time-step. The LSTM-based RNN model has been used for mental arithmetic task classification using EEG signal features [15]. The RNN models can exploit the correlations between the features of EEG signals in different time-steps for the classification of mental arithmetic tasks.

Other RNN variants, such as bidirectional LSTM (BLSTM) and gated recurrent unit (GRU) [29], have not been used for the categorization of mental arithmetic tasks, such as BFMAC vs. DMAC and BMAC vs. GMAC, respectively, using EEG signal features. The novelty of this work is to explore various state-space domain non-linear entropy features and RNNs for mental-arithmetic-based cognitive workload task classification using EEG signals. The major contributions of this work are highlighted as follows:

- The slope entropy, dispersion entropy, permutation entropy, sample entropy, and approximation entropy measures were computed from each EEG channel.
- The LSTM, BLSTM, and GRU-based RNN models were used to classify mental arithmetic tasks.
- The classification strategies, such as BMAC vs. GMAC and BFMAC vs. DMAC, are considered in this work.

The remaining parts of this paper are organized as follows. In Section 2, a description of the multi-channel EEG database for mental-arithmetic-based cognitive workload classification is written. The proposed approach is explained in Section 3. In Section 4, the results are evaluated using the proposed approach, and a discussion on the results is presented. Finally, in Section 5, the conclusions of this paper are written.

#### 2. Multi-Channel EEG Database for Mental Arithmetic Tasks

In this work, we used a public database (MIT Physionet EEG mental arithmetic task dataset) to evaluate the proposed approach [30,31]. This database contains artifact-free multi-channel EEG recordings from 36 subjects. The multi-channel EEG signals were recorded using the neurocom monopolar 23 channel data acquisition system [30]. The electrode setup to record the multi-channel EEG signals is depicted in Figure 1.

The standard international 10–20 system for electrode placement and Ag/Ag-cl electrodes were used for the recording of EEG signals [30]. The electrodes were placed in the frontal region (F7, F8, Fz, F3, and F4), parietal region (P3, P4, and Pz), occipital region (O1 and O2), central region (C3, C4, and Cz), temporal region (T5, T6, T3, and T4), and anterior frontal (Fp1 and Fp2), respectively, during EEG recording [13,30]. Independent component analysis (ICA) was used for the filtering of ocular, cardiac overlapping, and muscle artifacts from the recorded multi-channel EEG signals.

The clinical protocols, such as no clinical manifestations of cognitive impairment, no verbal and non-verbal learning disabilities, and normal vision were followed by the subjects during the EEG recordings [30]. Those subjects who were drug or alcohol addicted, and with psychiatric disorders were excluded during the recording of the EEG signals. Each subject performed an arithmetic task such as the subtraction of two numbers. In this database, multi-channel EEG recordings from each subject were comprised of 180-s



resting-state EEG and 60-s mental arithmetic calculation-based cognitive state EEG data. The sampling frequency of the multi-channel EEG recordings was 500 Hz.

Figure 1. Electrode setup to record multi-channel EEG data for mental arithmetic tasks.

The multi-channel EEG recordings were divided into two groups—good (G) and bad (B)—based on the arithmetic calculations performed by each subject. The GMAC class is interpreted as the subjects who performed good quality arithmetic calculations with the number of calculations per four minutes as  $21 \pm 7.4$ . Similarly, the BMAC is termed as the subjects who performed bad quality mental arithmetic calculations with the number of calculations per four minutes as  $7 \pm 3.6$ .

The annotations or the count quality of each multi-channel EEG recording are given in the database. The symbols 'B' and 'G' denote bad and good mental-arithmetic-based EEG recordings based on the number of subtractions performed by the subject. The annotations of the multi-channel EEG recordings for all subjects before performing any mental arithmetic calculations (BFMAC or rest state) and during the mental arithmetic calculations (DMAC or active state) are also given in the database. In this work, the cognitive task classification strategies, such as BMAC vs. GMAC and BFMAC vs. DMAC, are studied.

## 3. Method

In this section, we describe the proposed method for the automated classification of BFMAC vs. DMAC and BMAC vs. GMAC using multi-channel EEG signals. The flowchart for the mental arithmetic calculation-based cognitive workload classification task is depicted in Figure 2. The step-by-step procedure for the automated categorization of mental arithmetic calculation-based cognitive workload classification tasks is as follows:

- Segmentation of multi-channel EEG recordings into multi-channel EEG frames.
- Evaluation of the state-space domain non-linear entropy features from each multichannel EEG frame.
- Classification of mental arithmetic tasks using RNN models.

We describe the detailed theory of each block of the flowchart in the following subsections.



Figure 2. Flowchart of the proposed approach for the automated classification of mental arithmetic tasks.

## 3.1. Segmentation of Multi-Channel EEG Recordings

In this study, we considered a non-overlapping window of 2 s duration ( $2 \times 500 = 1000$  samples) for segmenting each multi-channel EEG recording. A total of 30 frames were evaluated from each multi-channel EEG recordings. The total number of multi-channel EEG frames considered in this work was  $36 \times 30 = 1080$ . Figures 3 and 4 show the EEG frames of the Fp1 and O1 channels for BMAC (Figure 3a,b)) and GMAC (Figure 3c,d) tasks as well as DMAC (Figure 4a,b) and BFMAC (Figure 4c,d) tasks. We observed significant differences in the amplitude values and temporal characteristics of EEG signals between BMAC- and GMAC-, and between BFMAC- and DMAC-based cognitive tasks.

A study [32] reported that the positive level amplitude increased during mental arithmetic tasks. The temporo-centro-parietal activity in multi-channel EEG signal increased during the mental arithmetic calculation-based active state compared to the rest state [32,33]. These physiological changes affect both the temporal and spatial characteristics of multi-channel EEG signals. Therefore, the features evaluated from the EEG signals can be used for the automated classification of mental-arithmetic-based cognitive tasks. In the following subsection, the entropy features evaluated from the multi-channel EEG frames are described.



**Figure 3.** (a) Fp1 channel EEG signal corresponding to a BMAC task. (b) O1 channel EEG signal corresponding to a BMAC task. (c) Fp1 channel EEG signal corresponding to a GMAC task. (d) O1 channel EEG signal corresponding to a GMAC task.



**Figure 4.** (a) Fp1 channel EEG signal corresponding to a DMAC task. (b) O1 channel EEG signal corresponding to a DMAC task. (c) Fp1 channel EEG signal corresponding to a BFMAC task. (d) O1 channel EEG signal corresponding to a BFMAC task.

## 3.2. Non-Linear Entropy Features

In this study, we computed five entropy measures, viz. the slope entropy [23], dispersion entropy [22], permutation entropy [24], sample entropy [25], and approximation entropy [34], from each EEG channel for the classification of mental-arithmetic-based cognitive workload tasks. The slope entropy was evaluated using the difference between the consecutive amplitude values of each embedded vector extracted from the signal [23].

Here, we evaluated the slope entropy from each EEG channel. The step-by-step procedure to evaluate the slope entropy for *i*th channel EEG signal was as follows [23]:

**Step-1**: The *i*th channel EEG signal is denoted as  $\mathbf{x}^i = [x^i(n)]_{n=1}^N$ . The *j*th embedded vector extracted from the *i*th channel EEG signal is given by  $\mathbf{y}_{i}^{i,L} = [x^{i}(j), x^{i}(j+1), \dots, x^{i}(j+L-1)]$ where L is the embedded dimension. The factor,  $\mathbf{y}_{i}^{i,L}$  is termed as the *j*th embedded vector and j = 1, 2, ..., N - L + 1.

Step-2: The difference between the consecutive sample values of each embedded vector of *i*th channel EEG signal is evaluated, and the slope signal for *j*th embedded vector is evaluated as follows:

$$d_j^{i,L}(k) = \mathbf{y}_j^{i,L}(k+1) - \mathbf{y}_j^{i,L}(k)$$
(1)

where k = 1, 2... L - 1. In vector form, the slope signal *j*th embedded vector is denoted as follows:

$$\mathbf{d}_{j}^{i,L} = [d_{j}^{i,L}(1), d_{j}^{i,L}(2), \dots, d_{j}^{i,L}(L-1)]$$
<sup>(2)</sup>

**Step-3**: Each element of the slope signal for the *j*th embedded vector of *i*th channel is mapped to negative and positive integer values, and this mapping is given as follows:

- If  $d_i^{i,L}(k) < -\zeta$ , then assign  $d_i^{i,L}(k) = -2$ .
- If  $-\zeta \leq d_j^{i,L}(k) < -\eta$ , then assign  $d_j^{i,L}(k) = -1$ . If  $-\eta \leq d_j^{i,L}(k) < \eta$ , then assign  $d_j^{i,L}(k) = 0$ . If  $\eta \leq d_j^{i,L}(k) < \zeta$ , then assign  $d_j^{i,L}(k) = 1$ .
- •
- •
- If  $d_i^{i,L}(k) > \zeta$ , then assign  $d_j^{i,L}(k) = 2$ .

 $\eta$  and  $\zeta$  are the slope entropy parameters, and  $\zeta > \eta$ .

**Step-4**: For *j*th embedded vector of the *i*th channel, the mapped pattern containing the positive and negative integer values for the EEG signal is evaluated and is denoted as  $\pi_i$ . Step-5: The mapped patterns for all embedded vectors are evaluated. The relative frequency (RF) vector is computed using the pattern matching concept by considering all patterns and is denoted as  $\mathbf{r}^i = [r^i(1), r^i(2), \dots, r^i(T)]$ , where T is the total number of elements in the RF vector.

**Step-6**: The probability evaluated using the RF vector for *i*th channel is given as follows [23]:

$$P_{t}^{i} = \frac{r^{i}(t)}{\sum_{t=1}^{T} r^{i}(t)}$$
(3)

Thus, the slope entropy of *i*th channel EEG signal is evaluated as follows:

$$SlopeEN^{i} = -\sum_{t=1}^{T} P_{t}^{i} log_{2} P_{t}^{i}$$
(4)

Similarly, the approximation entropy (AppEN), permutation entropy (PermEN), dispersion entropy (DisEn), and sample entropy (SampEN) features were also calculated from each EEG channel, and for the *i*th EEG channel, these features are denoted as AppEN<sup>*i*</sup>,  $PermEN^{i}$ ,  $DesEN^{i}$ , and  $SampEN^{i}$ , respectively. For more details on AppEN, PermEN, DesEN, and SampEN, we encourage readers to refer to [22,24,25,34].

As five entropy features were computed from each EEG channel, a total of 95 dimensional feature vectors were obtained from each multi-channel EEG frame. This 95dimensional feature vector sequence, z(t) was used as the input to the RNN-based model for classification.

#### 4. Recurrent Neural Network (RNN)

In this study, we used three RNN variants-namely, LSTM, BLSTM, and GRU. The block diagram of the mental arithmetic task classification using the LSTM, BLSTM, and GRU classifiers is shown in Figure 5. The classification strategies, such as BMAC vs. GMAC and BFMAC vs. DMAC, were used in this work. The feature matrix is denoted as  $\mathbf{Z} \in R^{q \times s}$  where q denotes the number of instances or time-steps and s = 95 denotes the number of features. The training and the test instances for each type of RNN classifier were selected using both hold-out and 10-fold cross-validation (CV) techniques [6,35].

For hold-out CV, we considered 60%, 10%, and 30% of the instances for training, validation, and testing, respectively, for the LSTM, BLSTM, and GRU models [35]. The numbers of instances used for each class are shown in Table 1. From this table, for the BMAC vs. GMAC classification task, there is a class imbalance problem. We considered random over-sampling during the training of each type of RNN classifier to overcome the class imbalance problem [36]. The LSTM classifier is a type of RNN [37] and has been used in different biomedical applications [28,38,39].

The LSTM layer mainly comprises of the cell, candidate value, input gate, forget gate, and output gate. The cell is considered as a memory and is used to remember the information at different time-intervals [37]. Similarly, the information flow from and to the cell is performed using the gates. The architecture of the LSTM network is shown in Figure 6. In LSTM, the output at the *t*th time-step was evaluated using the input z(t) and the activation g(t - 1) from the (t - 1)th time-step [40]. Here, we denote different parameters, such as memory cells as m(t), candidate values as  $\tilde{m}(t)$ , forget gates as FG, update gates as UG, output gates as OG, and outputs at the *t*th level step as g(t) [37]. The mathematical expressions of LSTM for the forget gate, update gate, candidate value, memory cell, and output are given as follows: [37]:

$$FG(t) = f(W_{FG}.[g(t-1), z(t)] + b_{FG})$$
(5)

$$UG(t) = f(W_{UG}.[g(t-1), z(t)] + b_{UG})$$
(6)

$$\tilde{m}(t) = \tanh(\mathsf{W}_{\mathsf{m}}.[g(t-1), z(t)] + \mathsf{b}_{\mathsf{m}}) \tag{7}$$

$$m(t) = FG(t) \otimes M(t-1) + UG(t) \otimes \tilde{m}(t)$$
(8)

$$OU(t) = f(W_{OU}.[g(t-1), z(t)] + b_{OU})$$
(9)

$$g(t) = OU(t) \otimes \tanh(m(t)) \tag{10}$$

where  $W_{FG}$  and  $b_{FG}$  are the weight and bias values for forget gate. Similarly,  $W_{UG}$  and the  $b_{UG}$  are termed as the weight and the bias values for update gate. Moreover,  $W_{OU}$  and  $b_{OU}$  are the weight and bias values at the output gate. '*f*' is denoted as the sigmoid activation function. The operator  $\otimes$  is termed as the Hadamard product [41].

Table 1. Number of multi-channel EEG instances used in this work.



Figure 5. Classification of mental arithmetic task using EEG feature vector sequence with different RNN variants.





BLSTM considers both past and future time-step information to model the current time step for RNN model [42]. It consists of both forward and backward LSTM parts. The forward LSTM mathematical expressions are same as in Equations (5)–(10). For backward LSTM, however, the mathematical expressions of forget gate, update gate, candidate value, memory cell and output gate are given as follows [42]:

$$FG(t) = f(W_{FG}.[g(t+1), z(t)] + b_{FG})$$
(11)

$$UG(t) = f(W_{UG}.[g(t+1), z(t)] + b_{UG})$$
(12)

$$\tilde{m}(t) = \tanh(\mathsf{W}_{\mathsf{m}}.[g(t+1), z(t)] + \mathsf{b}_{\mathsf{m}}) \tag{13}$$

$$m(t) = FG(t) \otimes M(t+1) + UG(t) \otimes \tilde{m}(t)$$
(14)

$$OU(t) = f(W_{OU}.[g(t+1), z(t)] + b_{OU})$$
(15)

$$g(t) = OU(t) \otimes \tanh(m(t)) \tag{16}$$

The GRU is the simplified version of LSTM, which consists of two gates—namely, the update gate and the reset gate [43]. The GRU model architecture is shown in Figure 7. The mathematical expressions for the reset gate (RG), update gate (UG), candidate value,  $\tilde{m}(t)$ , and memory cell m(t) at the *t*th time-step are given below [43]:

$$RG(t) = f(W_{RG}.[m(t-1), z(t)] + b_{RG})$$
(17)

$$UG(t) = f(W_{UG}.[m(t-1), z(t)] + b_{UG})$$
(18)

$$\tilde{m}(t) = \tanh(W_{\rm m}.[{\rm RG}(t) \otimes M(t-1), z(t)] + b_{\rm m})$$
(19)

$$m(t) = (1 - \mathrm{UG}(t)) \otimes m(t - 1) + \mathrm{UG}(t) \otimes \tilde{m}(t)$$
<sup>(20)</sup>

In this study, the classification model shown in Figure 4 comprises the input layer, RNN variant layer, fully-connected (FC) layer, softmax layer, and classification layer. The input feature vector at the *t*th time-step is denoted as z(t). The LSTM/BLSTM/GRU layer was used to process the input feature sequence. The number of hidden neurons considered in this layer is denoted as  $n_h$ . The output obtained in the LSTM/BLSTM layer is denoted as  $\mathbf{g} = [g(t)]_{t=1}^T$  where *T* is the total number of time-steps considered for the LSTM, BLSTM, and GRU layers. For the GRU layer, the output is the same as the memory, m(t).

The FC layer was used to convert the LSTM/BLSTM/GRU layer activation to a feature vector containing two features. This feature vector for the *i*th instance is denoted as

 $FV^i = [f1^i, f2^i]$ . The softmax based activation function was used in the classification layer. The *k*th output neuron activation for the *t*th time-step was evaluated as follows:

$$\tilde{P}_k^i = \frac{e^{\mathrm{FV}_k^i}}{\sum_{k=1}^{K} e^{\mathrm{FV}_k^i}}$$
(21)

where K = 2 is the number of classes. The binary cross-entropy function was used as the cost function for the LSTM, BLSTM, and GRU classifiers [35]. The training parameters used for the LSTM, BLSTM, and GRU models are shown in Table 2. The Adam optimizer was used for the evaluation of the weight and bias parameters. The performance of three RNN variants for the classification of mental arithmetic calculation tasks was evaluated using the accuracy, sensitivity, F1-score, and specificity measures [35].



Figure 7. Block-diagram of a single GRU cell.

**Table 2.** Training parameters of LSTM-, BLSTM- and GRU-based classifiers for the automated classification of both BMAC vs. GMAC and BFMAC vs. DMAC tasks.

Classifiers	CV	Hidden Neurons	Epochs	Mini-Batch Size	L2-Regularization
LSTM	hold-out	256	200	256	0.001
BLSTM	hold-out	128	200	128	0.001
GRU	hold-out	256	200	128	0.001
LSTM	10-fold	512	200	256	0.001
BLSTM	10-fold	512	200	256	0.001
GRU	10-fold	512	200	256	0.001

#### 5. Results and Discussion

In this section, we discuss the statistical analysis of the selected entropy features for BMAC vs. GMAC and BFMAC vs. DMAC classification tasks and the results of classification using RNN-based models. Student's *t*-test was used to evaluate the statistical significance of the entropy features of the EEG signals for the BFMAC vs. DMAC and BMAC vs. GMAC classification tasks [44]. The significant level for the *t*-test of the entropy features of each channel for both classification tasks was considered as 0.05.

Box-plots showing the within-class variations for the Fp1-channel EEG signal dispersion entropy, F7-channel EEG signal slope entropy, C4-channel EEG signal approximation entropy, O1-channel EEG signal sample entropy, and O1-channel EEG signal permutation entropy features for the BFMAC vs. DMAC classification task are shown in Figure 8a–e, respectively. From these plots, the median values of each entropy feature were different for the BFMAC and DMAC classes. Similarly, we show the mean and standard deviation values of different entropy features for the Fp1, F7, C4, and O1-channel EEG signals in Table 3.

For channel Fp1, the p-values of the approximation entropy, dispersion entropy, sample entropy, and slope entropy features were less than 0.05. Significant differences in the mean values of the Fp1 channel EEG signal entropy features in between the BFMAC and DMAC classes were also observed. Moreover, for the F7-channel EEG signal, the p-values of all entropy features were found to be less than 0.05.

For the C4-channel EEG signal, except for the dispersion entropy, other entropy features showed p-values of less than 0.05, and these selected features were clinically significant for the classification of BFMAC vs. DMAC tasks. For the O1-channel EEG signal, higher differences in the mean values of the sample entropy and permutation entropy were observed between the BFMAC and DMAC classes. Similarly, statistical variations were also observed for the entropy features of other EEG channels.

Channel	Entropy	Class (BFMAC)	Class (DMAC)	<i>p</i> -Value
	Approximate Entropy	$0.075\pm0.036$	$0.069\pm0.032$	0.0017
	Dispersion Entropy	$3.134\pm0.613$	$2.916\pm0.906$	$2.96 imes10^{-6}$
Fp1	Permutation Entropy	$1.977\pm0.382$	$1.994\pm0.412$	0.4607
	Sample Entropy	$0.067\pm0.032$	$0.061\pm0.028$	0.0002
	Slope Entropy	$0.071\pm0.031$	$0.077\pm0.037$	0.0081
	Approximate Entropy	$0.071\pm0.029$	$0.075\pm0.030$	0.0159
	Dispersion Entropy	$3.096\pm0.545$	$2.979\pm0.905$	0.0092
F7	Permutation Entropy	$1.950\pm0.376$	$2.011\pm0.426$	0.0109
	Sample Entropy	$0.063\pm0.026$	$0.067\pm0.026$	0.0339
	Slope Entropy	$0.086\pm0.027$	$0.078\pm0.024$	$2.61  imes 10^{-7}$
	Approximate Entropy	$3.300\pm0.540$	$3.049 \pm 0.881$	$1.37  imes 10^{-8}$
	Dispersion Entropy	$1.893\pm0.388$	$1.896\pm0.414$	0.8871
C4	Permutation Entropy	$0.079\pm0.024$	$0.074\pm0.019$	$2.71 imes10^{-5}$
	Sample Entropy	$0.085\pm0.021$	$0.080\pm0.020$	$7.35 imes10^{-5}$
	Slope Entropy	$3.238\pm0.493$	$3.008\pm0.867$	$6.40 imes10^{-8}$
	Approximate Entropy	$1.864\pm0.361$	$1.873\pm0.408$	0.7143
	Dispersion Entropy	$0.069\pm0.019$	$0.067\pm0.018$	0.0337
O1	Permutation Entropy	$0.085\pm0.020$	$0.076\pm0.020$	$4.75 imes10^{-15}$
	Sample Entropy	$3.247\pm0.487$	$2.976\pm0.859$	$1.3610^{-10}$
	Slope Entropy	$1.826\pm0.365$	$1.847\pm0.429$	0.3873

**Table 3.** Statistical analysis of the entropy features obtained using the Fp1, F7, C4, and O1 channel EEG signals for the automated classification of the BFMAC and DMAC classes.

Similarly, the box-plots of the Fp1-channel approximation entropy, Fp1-channel sample entropy, F7-channel dispersion entropy, C4-channel slope entropy, and O1-channel permutation entropy features for the BMAC and GMAC classes are depicted in Figure 9a–e, respectively. The statistical analysis of all entropy features of the Fp1, F7, C4, and O1-channel EEG signals for the BMAC and GMAC classes are shown in Table 4. From the table, the Fp1-channel approximation entropy, dispersion entropy, sample entropy, and permutation entropy features had higher mean value differences between the BMAC and GMAC classes when compared to the slope entropy features.

Similarly, the F7-channel dispersion entropy, permutation entropy, and slope entropy features had p-values less than 0.05. For the C4-channel EEG signal, higher mean value differences in the slope entropy, approximation entropy, permutation entropy, and sample entropy features were observed. The p-values of these entropy features were found to be less than 0.05 when compared to the dispersion entropy features. Similarly, for the O1-channel EEG signal, the dispersion entropy, permutation entropy, and slope entropy features had p-values that were found to be less than 0.05 for the BMAC vs. GMAC classification.

The p-values of the approximation entropy and slope entropy features of the O1channel EEG signals were more than 0.05 for the BMAC vs. GMAC classification task. From the statistical analysis results, various entropy features computed from different channel EEG signals effectively captured the diagnostic information for the automatic classification of mental arithmetic cognitive tasks.



**Figure 8.** (a) Box-plots of the Fp1-channel EEG signal dispersion entropy features obtained for the BFMAC and DMAC classes. (b) Box-plots of the F7-channel EEG signal slope entropy features obtained for the BFMAC and DMAC classes. (c) Box-plots of the C4-channel EEG signal approximation entropy features obtained for the BFMAC and DMAC classes. (d) Box-plots of the O1-channel EEG signal sample entropy features obtained for the BFMAC and DMAC classes. (e) Box-plots of the O1-channel EEG signal permutation entropy features obtained for the BFMAC and DMAC classes.

The classification performance of the GRU, LSTM, and BLSTM models for the BMAC vs. GMAC classification tasks using hold-out CV is shown in Table 5. In this work, five trial-based random hold-out CVs were used, and the mean and standard deviation values of each performance measure were evaluated [45]. We observed from this table that all RNN models obtained accuracy, sensitivity, precision, specificity, and F-score values of more than 98% for the BMAC vs. GMAC classification task.

In Table 6, we show the performance of RNN classifiers using 10-fold CV-based multichannel EEG instance selection for the BMAC vs. GMAC classification task. For the GRU model, the accuracy, sensitivity, specificity, precision, and F-score values were more than 98% for each fold. The average values of all performance measures were more than 98% for each fold. The average values of all performance measures were more than 99% for the GRU model. Similarly, for the BLSTM model, the accuracy, sensitivity, and F-score values were higher than 98% in each fold. However, the specificity values for fold 3 and fold 4 were 96.77% and 97.00%, respectively.

For the GRU model, the specificity value at fold 1 was 96.43% and 100% for rest of the folds using all entropy features. The classification performance obtained using the LSTM, BLSTM, and GRU models with each entropy feature and all EEG channels is shown in

Table 7. The slope entropy features combined with each type of RNN model obtained an average accuracy value of more than 93% when compared to other entropy features. The dispersion entropy features coupled with each RNN model also demonstrated average accuracy values of more than 99% for the sensitivity, precision, and specificity, and the F-score values of each type of RNN model were high using slope entropy features when compared to other entropy features extracted from multi-channel EEG signals.

The classification performance of the LSTM, BLSTM, and GRU models was evaluated using the entropy features of each EEG channel, and this is shown in Table 8. The entropy features evaluated using the O2, Fz, Cz, and Pz channel EEG signals obtained average accuracy values of more than 84% using each type of RNN model when compared to the entropy features of other EEG channels. Therefore, the O2, Fz, Cz, and Pz EEG channels were found to be significant for the classification of mental arithmetic tasks.

Channel	Entropy	Class (BMAC)	Class (GMAC)	<i>p</i> -Value
	Approximate Entropy	$0.065\pm0.030$	$0.076\pm0.041$	$4.22  imes 10^{-5}$
	Dispersion Entropy	$2.707 \pm 1.052$	$3.155\pm0.660$	$4.30 imes10^{-10}$
Fp1	Permutation Entropy	$1.937\pm0.375$	$2.009\pm0.393$	0.0196
	Sample Entropy	$0.057\pm0.027$	$0.068\pm0.037$	$5.08 imes10^{-5}$
	Slope Entropy	$0.071\pm0.037$	$0.070\pm0.035$	0.7494
	Approximate Entropy	$0.070\pm0.030$	$0.074\pm0.031$	0.1079
	Dispersion Entropy	$2.758 \pm 1.060$	$3.144\pm0.564$	$2.36 imes10^{-8}$
F7	Permutation Entropy	$1.930\pm0.377$	$1.994\pm0.397$	0.0381
	Sample Entropy	$0.062\pm0.026$	$0.066\pm0.027$	0.0805
	Slope Entropy	$0.083\pm0.024$	$0.078\pm0.029$	0.0186
	Approximate Entropy	$2.956 \pm 1.101$	$3.237\pm0.497$	$4.51  imes 10^{-5}$
	Dispersion Entropy	$1.853\pm0.364$	$1.892\pm0.390$	0.1959
C4	Permutation Entropy	$0.078\pm0.021$	$0.074\pm0.020$	0.0040
	Sample Entropy	$0.083\pm0.021$	$0.078\pm0.022$	0.0098
	Slope Entropy	$2.888 \pm 1.072$	$3.166\pm0.481$	$3.52  imes 10^{-5}$
	Approximate Entropy	$1.848\pm0.357$	$1.867\pm0.364$	0.5031
	Dispersion Entropy	$0.069\pm0.020$	$0.065\pm0.018$	0.0020
O1	Permutation Entropy	$0.083\pm0.021$	$0.074\pm0.020$	$3.37 imes10^{-8}$
	Sample Entropy	$2.920 \pm 1.080$	$3.116\pm0.472$	0.0037
	Slope Entropy	$1.821\pm0.365$	$1.807\pm0.373$	0.6325

**Table 4.** Statistical analysis of the entropy features obtained for the Fp1, F7, C4, and O1-channel EEG signals for the automated classification of the BMAC and GMAC classes.



**Figure 9.** (a) Box-plots of the Fp1-channel approximation entropy features obtained for the BMAC and GMAC classes. (b) Box-plots of the Fp1-channel sample entropy features obtained for the BMAC and GMAC classes. (c) Box-plots of the F7-channel dispersion entropy features obtained for the BMAC and GMAC classes. (d) Box-plots of the C4-channel slope entropy features obtained for the BMAC and GMAC classes. (e) Box-plots of the O1-channel permutation entropy features obtained for the BMAC and GMAC classes.

**Table 5.** The classification performance obtained using RNN-based classifiers with the hold-out CV approach for the automated classification of the BMAC vs. GMAC task.

Model	Accuracy (%)	Specificity (%)	Precision (%)	Sensitivity (%)	F-Score (%)
GRU	$99.20\pm0.47$	$98.22\pm2.02$	$99.32\pm0.77$	$99.57 \pm 0.43$	$99.45\pm0.32$
LSTM	$99.32\pm0.51$	$98.00\pm2.14$	$99.24 \pm 0.80$	$99.83 \pm 0.23$	$99.53\pm0.35$
BLSTM	$99.69\pm0.31$	$99.56\pm0.61$	$99.83\pm0.23$	$99.74 \pm 0.38$	$99.79\pm0.21$

**Table 6.** The classification performance obtained using RNN classifiers with the 10-fold CV approach for the automated classification of the BMAC vs. GMAC task.

Model	Metrics	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Average ( $\mu \pm \sigma$ )
	Accuracy (%)	99.07	100.0	100.0	100.0	100.0	100.0	99.07	100.0	100.0	100.0	$99.81 \pm 0.39$
	Specificity (%)	96.43	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	$99.64 \pm 1.23$
GRU	Precision (%)	98.77	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	$99.88 \pm 0.39$
	Sensitivity (%)	100.0	100.0	100.0	100.0	100.0	100.0	98.70	100.0	100.0	100.0	$99.87 \pm 0.41$
	F-score (%)	99.38	100.0	100.0	100.0	100.0	100.0	98.35	100.0	100.0	100.0	$99.87\pm0.27$
	Accuracy (%)	100.0	99.07	100.0	100.0	100.0	100.0	100.0	99.07	100.0	100.0	$99.81\pm0.39$
	Specificity (%)	100.0	96.30	100.0	100.0	100.0	100.0	100.0	96.30	100.0	100.0	$99.26 \pm 1.56$
LSTM	Precision (%)	100.0	98.78	100.0	100.0	100.0	100.0	100.0	98.78	100.0	100.0	$99.76\pm0.51$
	Sensitivity (%)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0 + 0.00
	F-score (%)	100.0	99.39	100.0	100.0	100.0	100.0	100.0	99.39	100.0	100.0	$99.88 \pm 0.26$

Table 6. Cont.

Model	Metrics	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Average ( $\mu \pm \sigma$ )
	Accuracy (%)	100.0	99.07	99.07	97.22	100.0	99.07	100.0	100.0	100.0	100.0	$99.43 \pm 0.90$
	Specificity (%)	100.0	100.0	96.77	97.00	100.0	96.00	100.0	100.0	100.0	100.0	$98.98 \pm 1.67$
BLSTM	Precision (%)	100.0	100.0	98.72	96.30	100.0	98.81	100.0	100.0	100.0	100.0	$99.38 \pm 1.20$
	Sensitivity (%)	100.0	98.57	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	$99.86 \pm 0.45$
	F-score (%)	100.0	99.28	99.35	98.11	100.0	99.40	100.0	100.0	100.0	100.0	$99.61\pm0.61$

**Table 7.** The classification performance obtained for RNN classifiers for selected entropy features using the hold out CV approach for the BMAC vs. GMAC classification task.

Model	Feature Selection	Accuracy (%)	Specificity (%)	Precision (%)	Sensitivity (%)	F-Score (%)
GRU	Approximation entropy Sample entropy Permutation entropy Dispersion entropy Slope entropy	$\begin{array}{c} 83.40 \pm 1.24 \\ 90.74 \pm 1.57 \\ 88.27 \pm 1.85 \\ 91.05 \pm 0.90 \\ 93.83 \pm 1.23 \end{array}$	$\begin{array}{c} 76.44 \pm 4.67 \\ 83.11 \pm 4.40 \\ 82.22 \pm 6.76 \\ 84.00 \pm 2.17 \\ 87.33 \pm 4.55 \end{array}$	$\begin{array}{c} 90.51 \pm 1.64 \\ 93.54 \pm 1.58 \\ 93.04 \pm 2.46 \\ 93084 \pm 0.80 \\ 95.22 \pm 1.57 \end{array}$	$\begin{array}{c} 86.07 \pm 1.62 \\ 93.68 \pm 1.40 \\ 90.60 \pm 0.52 \\ 93.76 \pm 0.78 \\ 96.32 \pm 1.97 \end{array}$	$\begin{array}{c} 88.22 \pm 0.87 \\ 93.60 \pm 1.07 \\ 91.79 \pm 1.19 \\ 93.80 \pm 0.62 \\ 95.75 \pm 0.87 \end{array}$
LSTM	Approximation entropy Sample entropy Permutation entropy Dispersion entropy Slope entropy	$\begin{array}{c} 84.88 \pm 1.20 \\ 88.02 \pm 1.12 \\ 91.85 \pm 1.08 \\ 90.06 \pm 1.37 \\ 95.62 \pm 0.74 \end{array}$	$\begin{array}{c} 78.22 \pm 5.70 \\ 81.11 \pm 3.93 \\ 86.44 \pm 2.78 \\ 81.78 \pm 3.39 \\ 93.33 \pm 1.57 \end{array}$	$\begin{array}{c} 91.33 \pm 1.88 \\ 92.61 \pm 1.26 \\ 94.76 \pm 1.19 \\ 93.02 \pm 1.24 \\ 97.41 \pm 0.60 \end{array}$	$\begin{array}{c} 87.44 \pm 2.94 \\ 90.68 \pm 2.27 \\ 93.93 \pm 1.19 \\ 93.25 \pm 0.93 \\ 96.50 \pm 0.63 \end{array}$	$\begin{array}{c} 89.29 \pm 0.98 \\ 91.62 \pm 0.88 \\ 94.33 \pm 0.75 \\ 93.13 \pm 0.93 \\ 96.95 \pm 0.51 \end{array}$
BLSTM	Approximation entropy Sample entropy Permutation entropy Dispersion entropy Slope entropy	$\begin{array}{c} 84.69 \pm 1.55 \\ 88.15 \pm 1.75 \\ 91.79 \pm 1.21 \\ 90.25 \pm 1.85 \\ 94.20 \pm 0.96 \end{array}$	$78.44 \pm 4.56 \\82.67 \pm 2.30 \\84.22 \pm 4.67 \\83.11 \pm 2.98 \\91.78 \pm 1.86$	$\begin{array}{c} 91.35 \pm 1.57 \\ 93.13 \pm 0.86 \\ 94.01 \pm 1.67 \\ 93.47 \pm 1.11 \\ 96.79 \pm 0.68 \end{array}$	$\begin{array}{c} 87.09 \pm 2.68 \\ 90.26 \pm 2.37 \\ 94.70 \pm 1.64 \\ 92.99 \pm 2.02 \\ 95.13 \pm 1.67 \end{array}$	$\begin{array}{c} 89.14 \pm 1.22 \\ 91.66 \pm 1.33 \\ 94.33 \pm 0.83 \\ 93.22 \pm 1.33 \\ 95.94 \pm 0.70 \end{array}$

**Table 8.** Summary of the performance (%) obtained using each EEG channel and RNN-based classifiers for the automated classification of the BMAC vs. GMAC task.

EEG Channel	Fp1	Fp2	F3	F4	F7	F8	T3	T4	C3	C4
GRU	$62.22\pm2.37$	$62.96 \pm 10.72$	$55.31 \pm 3.35$	$66.33 \pm 3.28$	$64.51 \pm 5.82$	$59.51 \pm 7.53$	$59.51 \pm 7.53$	$66.42 \pm 5.87$	$67.10 \pm 3.81$	$63.95 \pm 2.96$
LSTM	$64.51 \pm 2.41$	$65.86 \pm 3.07$	$62.10\pm5.73$	$63.95 \pm 2.85$	$59.32\pm3.19$	$65.31 \pm 4.17$	$61.05 \pm 4.56$	$66.67 \pm 4.81$	$63.77 \pm 5.73$	$57.84 \pm 3.46$
BLSTM	$63.02\pm3.10$	$67.10\pm3.04$	$63.27\pm3.78$	$64.69 \pm 1.23$	$62.04 \pm 2.89$	$68.15 \pm 3.51$	$64.38 \pm 4.90$	$64.63\pm6.90$	$65.49 \pm 1.06$	$62.10 \pm 1.98$
EEG Channel	T5	T6	P3	P4	01	O2	Fz	Cz	Pz	
GRU	$68.40 \pm 4.04$	$62.96 \pm 5.78$	$60.74 \pm 9.47$	$60.49 \pm 7.03$	$61.54 \pm 3.09$	$85.74 \pm 1.06$	$93.89 \pm 1.14$	$89.51 \pm 1.15$	$93.58\pm0.86$	
LSTM	$64.81 \pm 5.59$	$69.81 \pm 6.52$	$61.98 \pm 5.79$	$61.23 \pm 9.17$	$63.02\pm5.31$	$86.48 \pm 1.79$	$95.31 \pm 0.51$	$89.94 \pm 0.64$	$92.59 \pm 1.18$	
BLSTM	$64.88 \pm 3.95$	$59.01 \pm 9.40$	$61.05\pm6.92$	$64.75 \pm 4.76$	$62.16\pm5.45$	$84.88 \pm 2.07$	$94.14 \pm 1.76$	$90.68 \pm 1.22$	$92.22 \pm 1.53$	

The plots of accuracy versus the number of iterations obtained using the LSTM, BLSTM, and GRU classifiers for the automated classification of the BFMAC vs. DMAC tasks are shown in Figures 10–12, respectively. Both the training and validation accuracy values were 100% for the LSTM classifier. Similarly, for the BLSTM classifier, the training and validation accuracy values obtained were 100% and 98.54% after 200 epochs. The training and validation accuracy values were more than 100% using the GRU classifier for the automated classification of the BFMAC vs. DMAC tasks.

Similar variations were also observed in plots of accuracy versus the number of iterations obtained using the LSTM, BLSTM, and GRU classifiers for the automated classification of the BMAC- vs. GMAC-based cognitive workload tasks. We also demonstrated the confusion matrices obtained using the LSTM, BLSTM, and GRU classifiers with one trial of hold-out CV for the automated classification of the BFMAC vs. DMAC and BMAC vs. GMAC tasks. In Figure 13a–c, we show the confusion matrices obtained using the LSTM, BLSTM, and GRU classifiers to the BFMAC vs. DMAC tasks. DMAC tasks.

The confusion matrix plots obtained using the LSTM, BLSTM, and GRU classifiers for the automated classification of both the BMAC vs. GMAC tasks are depicted in Figure 13d–f, respectively. The true positive and true negative values were high for all

three RNN classifiers for the BFMAC vs. DMAC classification tasks. The LSTM and BLSTM classifiers yielded higher false-positive values compared with the GRU classifier for the automated classification of the BMAC vs. GMAC tasks. We also evaluated the performance of the LSTM, BLSTM, and GRU classifiers using subject-independent CV cases for the automated classification of the BMAC vs. GMAC tasks.

The subject-wise accuracy values using the LSTM, BLSTM, and GRU classifiers from multichannel EEG instances are shown in Figure 14a–c. From these figures, the average accuracy values obtained using subject-independent CV for the LSTM, BLSTM, and GRU classifiers were 58.70%, 55.74%, and 60.55%, respectively. The GRU classifier obtained the highest classification accuracy with subject-independent CV over other types of RNN classifiers for the classification of the BMAC vs. GMAC tasks.



**Figure 10.** Plots of accuracy versus the number of iterations obtained using the LSTM classifier with hold-out CV for the automated classification of the BFMAC vs. DMAC tasks.



**Figure 11.** Plots of accuracy versus the number of iterations obtained using the BLSTM classifier with hold-out CV for the automated classification of the BFMAC vs. DMAC tasks.



**Figure 12.** Plots of accuracy versus number of iterations obtained using GRU classifier with hold-out CV for automated classification of BFMAC vs. DMAC tasks.



**Figure 13.** (a) Confusion matrix of the LSTM classifier for the BFMAC vs. DMAC classification task using one trial hold-out CV, (b) Confusion matrix of the BLSTM classifier for the BFMAC vs. DMAC classification task using one trial hold-out CV, (c) Confusion matrix of the GRU classifier for the BFMAC vs. DMAC classification task using one trial hold-out CV, (d) Confusion matrix of the LSTM classifier for the GMAC vs. BMAC classification task using one trial hold-out CV, (e) Confusion matrix of the BLSTM classifier for the GMAC vs. BMAC classification task using one trial hold-out CV, (f) Confusion matrix of the GRU classifier for the GMAC vs. BMAC classification task using one trial hold-out CV, (f) Confusion matrix of the GRU classifier for the GMAC vs. BMAC classification task using one trial hold-out CV, (f) Confusion matrix of the GRU classifier for the GMAC vs. BMAC classification task using one trial hold-out CV, (f) Confusion matrix of the GRU classifier for the GMAC vs. BMAC classification task using one trial hold-out CV, (f) Confusion matrix of the GRU classifier for the GMAC vs. BMAC classification task using one trial hold-out CV, (f) Confusion matrix of the GRU classifier for the GMAC vs. BMAC classification task using one trial hold-out CV, (f) Confusion matrix of the GRU classifier for the GMAC vs. BMAC classification task using one trial hold-out CV.



**Figure 14.** (a) Accuracy values obtained for the LSTM classifier using subject independent CV for the automated classification of teh BMAC vs. GMAC task. (b) Accuracy values obtained for the BLSTM classifier using subject independent CV for the automated classification of the BMAC vs. GMAC task. (c) Accuracy values obtained for the GRU classifier using subject independent CV for the automated classification of the BMAC vs. GMAC task. (c) Accuracy values obtained for the GRU classifier using subject independent CV for the automated classification of the BMAC vs. GMAC task.

The classification results evaluated using each type of RNN model with all entropy features of multi-channel EEG signals for the BFMAC vs. DMAC classification task with hold-out CV are shown in Table 9. The LSTM, BLSTM, and GRU models obtained accuracy values of more than 99% using all entropy features of the multi-channel EEG signals for the BFMAC vs. DMAC classification task. The results obtained using various types of RNN-based models with all entropy features of multi-channel EEG with 10-fold CV are shown in Table 10.

From these results, the LSTM and BLSTM models obtained accuracy, sensitivity, and specificity values of more than 97% for each fold. However, the specificity of the GRU -based classifier was less than 97% for fold 3 and fold 6. The average accuracy of each type of RNN classifier was more than 99% for the BFMAC- vs. DMAC-based cognitive workload classification task. Therefore, the entropy features computed using multi-channel EEG demonstrated higher classification performance for the classification of cognitive workload tasks.

**Table 9.** The classification performance obtained using the RNN-based classifiers with the hold out CV approach for the automated classification of the BFMAC vs. DMAC task.

Model	Accuracy (%)	Specificity (%)	Precision (%)	Sensitivity (%)	F-Score (%)
GRU	$99.63 \pm 0.26$	$99.62\pm0.34$	$99.63 \pm 0.34$	$99.62\pm0.33$	$99.63\pm0.26$
LSTM	$99.38 \pm 0.58$	$99.88 \pm 1.88$	$98.91 \pm 1.15$	$99.88 \pm 0.28$	$99.39 \pm 0.57$
BLSTM	$99.44\pm0.46$	$99.62\pm0.55$	$99.63\pm0.54$	$99.26\pm0.80$	$99.44\pm0.46$

Model	Metrics	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Average ( $\mu \pm \sigma$ )
	Accuracy (%)	99.07	100	100	100	100	99.07	100	100	99.07	100	$99.72\pm0.44$
	Specificity (%)	100	100	100	100	100	98.08	100	100	100	100	$99.80\pm0.60$
LSTM	Precision (%)	100	100	100	100	100	98.25	100	100	100	100	$99.82 \pm 0.55$
	Sensitivity (%)	97.96	100	100	100	100	100	100	100	98.11	100	$99.60\pm0.82$
	F-score (%)	98.97	100	100	100	100	99.12	100	100	99.05	100	$99.71\pm0.46$
	Accuracy (%)	100	99.07	100	100	100	100	99.07	100	100	99.07	$99.72\pm0.44$
	Specificity (%)	100	100	100	100	100	100	98.11	100	100	100	$99.81 \pm 0.59$
BLSTM	Precision (%)	100	100	100	100	100	100	98.21	100	100	100	$99.82\pm0.56$
	Sensitivity (%)	100	98.33	100	100	100	100	100	100	100	98.25	$99.65\pm0.72$
	F-score (%)	100	99.16	100.0	100.0	100.0	100.0	99.10	100	100	99.12	$99.73\pm0.42$
	Accuracy (%)	100	99.07	99.07	97.22	100	99.07	100	100	100	100	$99.43\pm0.90$
	Specificity (%)	100	100	96.77	97.00	100	96.00	100	100	100	100	$98.98 \pm 1.67$
GRU	Precision (%)	100	100	98.72	96.30	100	98.81	100	100	100	100	$99.38 \pm 1.20$
	Sensitivity (%)	100	98.57	100	100	100	100	100	100	100	100	$99.86 \pm 0.45$
	F-score (%)	100	99.28	99.35	98.11	100	99.40	100	100	100	100	$99.61 \pm 0.61$

**Table 10.** The classification performance obtained using RNN-based classifiers with the 10-fold CV approach for the automated classification of the BFMAC vs. DMAC task.

We compared the classification results obtained using our approach with existing methods to classify mental arithmetic calculation-based cognitive tasks using multi-channel EEG signals obtained from the same database. The comparison results are shown in Table 11. The existing methods used features, such as the mean amplitude, variance, Shannon entropy, energy, other statistical features, and various supervised learning-based classifiers, such as SVM, LSTM, and decision trees, for the BFMAC vs. DMAC based cognitive task classification scheme.

The L2-norm, mean amplitude, energy, and Shannon entropy features combined with the decision tree classifier obtained a classification accuracy of 95.80%. The statistical features of multi-channel EEG signals coupled with the decision tree model achieved an accuracy of 91.67%, which is less than the classification accuracy reported in [13]. Similarly, the SVM classifier combined with the variance, energy, and Shannon entropy features obtained an accuracy of 98.60% for the automated classification of the BFMAC vs. DMAC task. The performance of the SVM classifier depended on the proper selection of the kernel functions, kernel parameters, and number of iterations [46].

Similarly, the training parameters of the decision tree classifier were the depth of the tree, the number of times the split occurred in the decision tree, and the split criteria or information gain, respectively. The optimal training parameters of both the SVM and decision tree classifiers were selected using a grid-search in the nested cross-validation domain [4,47].

Each RNN-based classifier in the proposed approach yielded superior classification performance compared with the existing methods for the automated classification of BFMAC- vs. DMAC-based cognitive classification tasks using multi-channel EEG signals. The LSTM, BLSTM, and GRU models successfully quantified the dependencies in the entropy features of multi-channel EEG signals, which further helped to create a boundary for the classification of BFMAC- and DMAC-based cognitive tasks. The advantages of this mental-arithmetic-based cognitive classification task are given as follows:

- Various entropy features computed from various EEG channels were used for the classification of mental arithmetic tasks.
- The entropy features from the O2, Fz, Cz, and Pz channel EEG signals demonstrated higher classification accuracy using RNN-based classifiers.
- The slope entropy features combined with each type of RNN-based classifier obtained higher classification accuracy over the other entropy features.
- The proposed approach obtained the highest classification accuracy at a 99% classification accuracy for the BFMAC vs. DMAC and BMAC vs. GMAC cognitive workload classification tasks.

Features Used	<b>Classification Model</b>	<b>Classification Task</b>	Accuracy (%)
Variance, energy, and Shannon en- tropy features extracted from the IBFs of multi-channel EEG sig- nal [14]	SVM	BFMAC vs. DMAC	98.60
Statistical features evaluated from the multi-channel EEG signal [15]	LSTM	BFMAC vs. DMAC	91.67
L2-norm, mean amplitude value, Shannon entropy and energy features extracted from EEG signals [13]	Decision tree	BFMAC vs. DMAC	95.80
Slope entropy, permutation entropy, dispersion entropy, approximation entropy and sample entropy features extracted from multi-channel EEG signals (Proposed work)	RNN (LSTM, BLSTM, and GRU)	BFMAC vs. DMAC	99.38 (LSTM), 99.44 (BLSTM), 99.63 (GRU)

**Table 11.** Comparison with existing methods for the automated mental arithmetic cognitive task classification using multi-channel EEG signals from the same database.

In this work, we used the multi-channel EEG signals from 36 subjects to evaluate the proposed approach of automated classification of BMAC vs. GMAC and BFMAC vs. DMAC arithmetic tasks. Multi-lead EEG signals from more subjects are needed to develop accurate and robust automated classifications of mental-arithmetic-based cognitive workload tasks. Other entropy features, such as the distribution entropy [48], and bubble entropy [49] features obtained from multi-channel EEG signals, can be used for the automated classification of mental-arithmetic-based cognitive tasks.

In this work, the RNN-based models were used for the classification. In the future, we intend to use convolutional autoencoder [50], LSTM-autoencoder [51], convolutional neural network (CNN) [35,52] and CNN-RNN [53]-based deep learning models for feature extraction and classification tasks using a large database with multi-channel EEG signals.

## 6. Conclusions

An automated approach for the classification of mental arithmetic calculation-based cognitive tasks using various entropy features obtained from multi-channel EEG signals is proposed in this paper. The state-space domain entropy measures, such as the sample entropy, approximation entropy, dispersion entropy, permutation entropy, and slope entropy features, were computed from multi-channel EEG signals. We used recurrent neural network (RNN)-based models, such as long short-term memory (LSTM), bidirectional LSTM (BLSTM), and gated recurrent unit (GRU), as classifiers to perform cognitive task classification schemes, such as bad mental arithmetic calculation (BFMAC) vs. during mental arithmetic calculation (DMAC), respectively.

We obtained classification accuracies of 99.88%, 99.43%, and 99.81% using LSTM, BLSTM, and GRU-based RNN models for the automated classification of the BMAC vs. GMAC classification task. Our proposed approach demonstrated a classification accuracy of more than 99% using all RNN-based models for the automated classification of BFMAC vs. DMAC tasks. The slope entropy features coupled with each type of RNN model obtained the highest classification accuracy for both BMAC vs. GMAC and BFMAC vs. DMAC automated cognitive classification tasks. In the future, our proposed approach can be tested with multi-channel EEG signals to classify more types of mental-arithmetic-based cognitive tasks for brain-computer interface (BCI) applications.

**Author Contributions:** Conceptualization, A.V.; Data curation, R.K.T. and S.K.G.; Formal analysis, S.P. and U.R.A.; Methodology, A.V. and R.K.T.; Project administration, U.R.A.; Resources, R.K.T.; Supervision, R.K.T.; Writing—original draft, R.K.T.; Writing—review & editing, U.R.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The codes of this paper are available upon request to the authors.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Gevins, A.; Smith, M.E. Neurophysiological measures of cognitive workload during human-computer interaction. *Theor. Issues Ergon. Sci.* 2003, 4, 113–131. [CrossRef]
- 2. Vandewalle, G.; Archer, S.N.; Wuillaume, C.; Balteau, E.; Degueldre, C.; Luxen, A.; Dijk, D.J.; Maquet, P. Effects of light on cognitive brain responses depend on circadian phase and sleep homeostasis. *J. Biol. Rhythm.* **2011**, *26*, 249–259. [CrossRef]
- 3. Murugappan, M.; Rizon, M.; Nagarajan, R.; Yaacob, S. Inferring of human emotional states using multichannel EEG. *Eur. J. Sci. Res.* **2010**, *48*, 281–299.
- 4. Bhattacharyya, A.; Tripathy, R.K.; Garg, L.; Pachori, R.B. A Novel Multivariate-Multiscale Approach for Computing EEG Spectral and Temporal Complexity for Human Emotion Recognition. *IEEE Sens. J.* **2020**. [CrossRef]
- 5. Foster, B.L.; Liley, D.T. Nitrous oxide paradoxically modulates slow electroencephalogram oscillations: Implications for anesthesia monitoring. *Anesth. Analg.* 2011, *113*, 758–765. [CrossRef]
- Tripathy, R.K.; Ghosh, S.K.; Gajbhiye, P.; Acharya, U.R. Development of automated sleep stage classification system using multivariate projection-based fixed boundary empirical wavelet transform and entropy features extracted from multichannel EEG signals. *Entropy* 2020, 22, 1141. [CrossRef] [PubMed]
- 7. Sharma, M.; Dhiman, H.S.; Acharya, U.R. Automatic identification of insomnia using optimal antisymmetric biorthogonal wavelet filter bank with ECG signals. *Comput. Biol. Med.* **2021**, *131*, 104246. [CrossRef]
- Cao, Z.; Chuang, C.H.; King, J.K.; Lin, C.T. Multi-channel EEG recordings during a sustained-attention driving task. *Sci. Data* 2019, 6, 1–8. [CrossRef]
- 9. Wang, Q.; Sourina, O. Real-time mental arithmetic task recognition from EEG signals. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2013, 21, 225–232. [CrossRef] [PubMed]
- 10. Zarjam, P.; Epps, J.; Chen, F.; Lovell, N.H. Estimating cognitive workload using wavelet entropy-based features during an arithmetic task. *Comput. Biol. Med.* 2013, 43, 2186–2195. [CrossRef]
- Bühler, E.; Bachmann, C.; Goyert, H.; Heinzel-Gutenbrunner, M.; Kamp-Becker, I. Differential diagnosis of autism spectrum disorder and attention deficit hyperactivity disorder by means of inhibitory control and 'theory of mind'. *J. Autism Dev. Disord.* 2011, 41, 1718–1726. [CrossRef]
- 12. Jansen, A.G.; Dieleman, G.C.; Jansen, P.R.; Verhulst, F.C.; Posthuma, D.; Polderman, T.J. Psychiatric polygenic risk scores as predictor for attention deficit/hyperactivity disorder and autism spectrum disorder in a clinical child and adolescent sample. *Behav. Genet.* **2020**, *50*, 203–212. [CrossRef]
- Fatimah, B.; Pramanick, D.; Shivashankaran, P. Automatic detection of mental arithmetic task and its difficulty level using EEG signals. In Proceedings of the 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 1–3 July 2020; pp. 1–6.
- Fatimah, B.; Javali, A.; Ansar, H.; Harshitha, B.; Kumar, H. Mental Arithmetic Task Classification using Fourier Decomposition Method. In Proceedings of the 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 28–30 July 2020; pp. 46–50.
- Ganguly, B.; Chatterjee, A.; Mehdi, W.; Sharma, S.; Garai, S. EEG Based Mental Arithmetic Task Classification Using a Stacked Long Short Term Memory Network for Brain-Computer Interfacing. In Proceedings of the 2020 IEEE Vlsi Device Circuit and System (VLSI DCS), Kolkata, India, 18–19 July 2020; pp. 89–94.
- Dutta, S.; Singh, M.; Kumar, A. Automated classification of non-motor mental task in electroencephalogram based brain-computer interface using multivariate autoregressive model in the intrinsic mode function domain. *Biomed. Signal Process. Control* 2018, 43, 174–182. [CrossRef]
- 17. Acharya, U.R.; Molinari, F.; Sree, S.V.; Chattopadhyay, S.; Ng, K.H.; Suri, J.S. Automated diagnosis of epileptic EEG using entropies. *Biomed. Signal Process. Control* 2012, 7, 401–408. [CrossRef]
- Tripathy, R.K.; Deb, S.; Dandapat, S. Analysis of physiological signals using state space correlation entropy. *Healthc. Technol. Lett.* 2017, 4, 30–33. [CrossRef] [PubMed]
- 19. Arunkumar, N.; Ramkumar, K.; Venkatraman, V.; Abdulhay, E.; Fernandes, S.L.; Kadry, S.; Segal, S. Classification of focal and non focal EEG using entropies. *Pattern Recognit. Lett.* **2017**, *94*, 112–117.
- 20. Chen, T.; Ju, S.; Yuan, X.; Elhoseny, M.; Ren, F.; Fan, M.; Chen, Z. Emotion recognition using empirical mode decomposition and approximation entropy. *Comput. Electr. Eng.* **2018**, *72*, 383–392. [CrossRef]
- Martínez-Cagigal, V.; Santamaría-Vázquez, E.; Hornero, R. Asynchronous control of P300-based brain–computer interfaces using sample entropy. *Entropy* 2019, 21, 230. [CrossRef] [PubMed]
- 22. Rostaghi, M.; Azami, H. Dispersion entropy: A measure for time-series analysis. *IEEE Signal Process. Lett.* **2016**, *23*, 610–614. [CrossRef]
- 23. Cuesta-Frau, D. Slope Entropy: A New Time Series Complexity Estimator Based on Both Symbolic Patterns and Amplitude Information. *Entropy* **2019**, *21*, 1167. [CrossRef]

- 24. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [CrossRef] [PubMed]
- 25. Richman, J.S.; Lake, D.E.; Moorman, J.R. Sample entropy. Methods Enzymol. 2004, 384, 172–184. [PubMed]
- 26. Roy, Y.; Banville, H.; Albuquerque, I.; Gramfort, A.; Falk, T.H.; Faubert, J. Deep learning-based electroencephalography analysis: A systematic review. *J. Neural Eng.* **2019**, *16*, 051001. [CrossRef] [PubMed]
- 27. Goldberg, Y. Neural network methods for natural language processing. *Synth. Lect. Hum. Lang. Technol.* **2017**, *10*, 1–309. [CrossRef]
- Michielli, N.; Acharya, U.R.; Molinari, F. Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals. *Comput. Biol. Med.* 2019, 106, 71–81. [CrossRef] [PubMed]
- 29. Chhetri, M.; Kumar, S.; Pratim Roy, P.; Kim, B.G. Deep BLSTM-GRU Model for Monthly Rainfall Prediction: A Case Study of Simtokha, Bhutan. *Remote Sens.* 2020, *12*, 3174. [CrossRef]
- 30. Zyma, I.; Tukaev, S.; Seleznov, I.; Kiyono, K.; Popov, A.; Chernykh, M.; Shpenkov, O. Electroencephalograms during mental arithmetic task performance. *Data* **2019**, *4*, 14. [CrossRef]
- Goldberger, A.L.; Amaral, L.A.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* 2000, 101, e215–e220. [CrossRef]
- 32. Pauli, P.; Lutzenberger, W.; Rau, H.; Birbaumer, N.; Rickard, T.C.; Yaroush, R.A.; Bourne, L.E. Brain potentials during mental arithmetic: Effects of extensive practice and problem difficulty. *Cogn. Brain Res.* **1994**, *2*, 21–29. [CrossRef]
- 33. Inouye, T.; Shinosaki, K.; Iyama, A.; Matsumoto, Y. Localization of activated areas and directional EEG patterns during mental arithmetic. *Electroencephalogr. Clin. Neurophysiol.* **1993**, *86*, 224–230. [CrossRef]
- 34. Pincus, S.M.; Gladstone, I.M.; Ehrenkranz, R.A. A regularity statistic for medical data analysis. J. Clin. Monit. 1991, 7, 335–345. [CrossRef]
- 35. Panda, R.; Jain, S.; Tripathy, R.; Acharya, U.R. Detection of shockable ventricular cardiac arrhythmias from ECG signals using FFREWT filter-bank and deep convolutional neural network. *Comput. Biol. Med.* **2020**, *124*, 103939. [CrossRef]
- 36. Japkowicz, N.; Stephen, S. The class imbalance problem: A systematic study. Intell. Data Anal. 2002, 6, 429–449. [CrossRef]
- 37. Kong, W.; Dong, Z.Y.; Jia, Y.; Hill, D.J.; Xu, Y.; Zhang, Y. Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Trans. Smart Grid* 2017, *10*, 841–851. [CrossRef]
- 38. Yildirim, O.; Baloglu, U.B.; Tan, R.S.; Ciaccio, E.J.; Acharya, U.R. A new approach for arrhythmia classification using deep coded features and LSTM networks. *Comput. Methods Programs Biomed.* **2019**, *176*, 121–133. [CrossRef] [PubMed]
- Tripathy, R.K.; Bhattacharyya, A.; Pachori, R.B. Localization of myocardial infarction from multi-lead ECG signals using multiscale analysis and convolutional neural network. *IEEE Sens. J.* 2019, 19, 11437–11448. [CrossRef]
- 40. Greff, K.; Srivastava, R.K.; Koutnik, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* 2016, *28*, 2222–2232. [CrossRef]
- 41. Amin, J.; Sharif, M.; Raza, M.; Saba, T.; Sial, R.; Shad, S.A. Brain tumor detection: A long short-term memory (LSTM)-based learning model. *Neural Comput. Appl.* **2020**, *32*, 15965–15973. [CrossRef]
- 42. Sankaran, N.; Jawahar, C. Recognition of printed Devanagari text using BLSTM Neural Network. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, 11–15 November 2012; pp. 322–325.
- Chen, J.; Jing, H.; Chang, Y.; Liu, Q. Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process. *Reliab. Eng. Syst. Saf.* 2019, 185, 372–382. [CrossRef]
- 44. Neideen, T.; Brasel, K. Understanding statistical tests. J. Surg. Educ. 2007, 64, 93–96. [CrossRef] [PubMed]
- Tripathy, R.; Acharya, U.R. Use of features from RR-time series and EEG signals for automated classification of sleep stages in deep neural network framework. *Biocybern. Biomed. Eng.* 2018, 38, 890–902. [CrossRef]
- 46. Suykens, J.A.; Vandewalle, J. Least squares support vector machine classifiers. Neural Process. Lett. 1999, 9, 293-300. [CrossRef]
- 47. Siddharth, T.; Tripathy, R.K.; Pachori, R.B. Discrimination of focal and non-focal seizures from EEG signals using sliding mode singular spectrum analysis. *IEEE Sens. J.* 2019, *19*, 12286–12296. [CrossRef]
- 48. Li, P.; Liu, C.; Li, K.; Zheng, D.; Liu, C.; Hou, Y. Assessing the complexity of short-term heartbeat interval series by distribution entropy. *Med Biol. Eng. Comput.* **2015**, *53*, 77–87. [CrossRef]
- 49. Manis, G.; Aktaruzzaman, M.; Sassi, R. Bubble entropy: An entropy almost free of parameters. *IEEE Trans. Biomed. Eng.* 2017, 64, 2711–2718.
- 50. Wen, T.; Zhang, Z. Deep convolution neural network and autoencoders-based unsupervised feature learning of EEG signals. *IEEE Access* **2018**, *6*, 25399–25410. [CrossRef]
- Elessawy, R.H.; Eldawlatly, S.; Abbas, H.M. A long short-term memory autoencoder approach for EEG motor imagery classification. In Proceedings of the 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM), Dubai, United Arab Emirates, 9–10 January 2020; pp. 79–84.
- 52. Jain, P.; Gajbhiye, P.; Tripathy, R.; Acharya, U.R. A two-stage deep CNN architecture for the classification of low-risk and high-risk hypertension classes using multi-lead ECG signals. *Informatics Med. Unlocked* **2020**, *21*, 100479. [CrossRef]
- 53. Basiri, M.E.; Nemati, S.; Abdar, M.; Cambria, E.; Acharya, U.R. ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis. *Future Gener. Comput. Syst.* 2021, 115, 279–294. [CrossRef]