Neural Comput & Applic
DOI 10.1007/s00521-017-3142-1

ORIGINAL ARTICLE

LU triangularization extreme learning machine in EEG cognitive task classification

Yakup Kutlu[1] · **Apdullah Yayık**[2] · **Esen Yildirim**[3] · **Serdar Yildirim**[4]

Received: 31 January 2016 / Accepted: 27 June 2017

Abstract Electroencephalography (EEG) has been used as a promising tool for investigation of brain activity during cognitive processes. The aim of this study is to reveal whether EEG signals can be used for classifying cognitive processes: arithmetic tasks and text reading. A recently introduced EEG database, which is constructed from 18 healthy subjects during a slide show including 60 slides of simple arithmetic tasks and easily readable texts, is used for this purpose. Multi-order difference plot-based time-domain attributes, number of values in specified regions after scattering the sequential difference values with several degrees, are extracted. For classification, improved extreme learning machine (ELM) scheme, namely luELM, by the use of lower–upper triangularization method instead of singular value decomposition which has disadvantages when used with huge data is proposed. As a result, higher accuracy results are achieved with reduced training time for proposed luELM classifier than traditional ELM classifier for both subject-dependent and subject-independent analysis.

✉ Yakup Kutlu
yakup.kutlu@iste.edu.tr

[1] Department of Computer Engineering, İskenderun Technical University, Hatay, Turkey

[2] Alparslan Defence Sciences, National Defense University, Ankara, Turkey

[3] Department of Electrical and Electronics Engineering, Adana Science and Technology University, Adana, Turkey

[4] Department of Computer Engineering, Adana Science and Technology University, Adana, Turkey

Keywords Cognitive processes · Lower–upper triangularization · Extreme learning machine · MoDP method · Optimized nodes

1 Introduction

Exploring the brain behaviour and dynamics during cognitive tasks has become a widely studied topic by researchers from various disciplines, such as mathematicians, physicist, neuroscientists, psychologists and bioengineers. Among those, cognitive tasks, in particular problem-solving and reading, have been investigated in many fields. In earlier studies, increase in the level of alpha band in Electroencephalography (EEG) spectrum during arithmetic tasks [21] and changes in rhythmic activities in frontal, parietal and occipital brain regions due to different levels of difficulty [12] were investigated. Recently, Galan and Beal investigated the use of amount of attention and cognitive workload during solving math problems for predicting whether the student solves the problem correctly, or not [9]. They reported that problem outcomes could be correctly predicted at rates better than chance. In another study, brain activities were examined while the subjects were solving math problems [19]. In the study, power decreases in alpha, beta and theta bands in various brain regions were reported as the task difficulty increased and differences in brain dynamics between fast and slow problem solvers were also examined. Besides, considerable amount of studies have focused on neuroimaging fields during text reading and comprehension. Based on the idea that processing the different parts of the text and making inferences is crucial on text comprehension, inference making while reading narrative and expository texts was studied [1]. Results showed that these types of texts are processed differently by the brain and exposition was more demanding than narration in terms of semantic processing [1]. Rosazza et. al. [26] investigated the brain behaviour in word recognition through ERPs. Ramsdem et al. [24] proved that the habit of reading, especially in early adolescence, has positive and major effects on advancing verbal intelligence. A recent study investigated the neurocognitive process of learning to read for early school age children [4]. These studies examined the brain behaviour for either problem-solving or reading. In brain–computer interface (BCI) studies, deciding the type of process would help on what to look for: difficulty level of the problem, cognitive workload, the text comprehension, etc. For that purpose, the authors have collected an EEG database and studied on classification of arithmetic and text processing using frequency domain features. Eraldemir et al. [6] used wavelet transform for feature extraction and Bayesian network (BayesNet) algorithm for a subject-dependent classification resulting in 89.1% true positive rate on the average. This classification scheme was repeated for reduced number of channels in Eraldemir et al. [8]. Channels were selected by feature selection using correlation-based feature selection (CFS) [13], and 90.0% true positive rate was obtained for only 10 channels. In another study, basic mathematical operations, addition/subtraction and multiplication/division, are classified using k-nn classifier with 79.3% true positive rate [7].

Feedforward neural networks have been extensively used with gradient descent algorithm in many disciplines. Since learning time of neural network is long, many researchers have worked on developing faster methods, such as early stopping criteria to speed up learning, gradient descent with adaptive learning rate [3]. Extreme learning machine (ELM) is introduced by Huang et al. [15] for single-hidden layer feedforward neural network (SLFN) to reach high generalization performance in shorter time. Contrary to traditional learning concepts, ELM arbitrary determines the input weights and hidden neuron biases and then calculates output weights using Moore–Penrose generalized inverse conditions [25]. Hidden layer biases are threshold values to prevent inner product to be zero. In order to improve generalized performance, different types of ELM, such as optimally pruned extreme learning machine (OP-ELM) [2], online sequential extreme learning machine (OS-ELM) [18] and evolutionary ELM [32], were introduced. Besides, QR decomposition technique with iteratively optimized neuron numbers using Gram–Schmidt and household reflection was introduced recently in Kutlu et al. [17]. This technique was used to detect P300 component of event-related potential (ERP) in a BCI smart home application, and 98.42% overall accuracy was obtained within milliseconds.

The traditional extreme learning machine is a slow method when working with huge data sets due to using singular value decomposition (SVD). In this study, a novel lower–upper triangularization-based [2] classification method, namely lower–upper triangulation ELM (luELM), a revised version of traditional extreme learning machine (ELM), which is a single-hidden layer feedforward neural network and is able to overcome pseudoinverse computation drawback in original ELM in terms of duration, is proposed. The proposed algorithm is tested by classification of a two-class EEG database. The second-order difference plot derived from the continuous signal is effective on visually demonstrating the degree of variability, which provides a powerful approach for summarizing long-duration time series, in the data set [22, 28]. Therefore, the features were constructed using time-domain multi-order difference plot (MoDP) features that are extracted by scattering the sequential difference values with several degrees and then determining the number of values in determined regions. Eventually, the proposed system ensures time-saving form thanks to both pace of luELM and time-domain MoDP algorithms. The remaining paper is organized as follows: in Sect. 2, database is described. Feature extraction and classification methods are explained in Sect. 3. Results and discussions are given in Sects. 4 and 5, respectively.

2 Materials and methods

2.1 Database description

In this study, the database named EEG mathematical and text processing (EEG-MaTeP) [5] is used. The experimental procedure and processes were approved by the Mustafa Kemal University Human Research Ethics Committee. EEG signals were recorded, using EEG-1200 Nihon Kohden system, from 18 healthy male subjects aged between 19 and 25 (average 20.39 $\pm$ 2.06) during basic mathematical operations and silent simple text reading. Nine of the subjects were undergraduate students at Department of Electrical and Electronics Engineering in Faculty of Engineering, Mustafa Kemal University, five of them were students at Department of Broadcast and Communications, and others were students at Department of Computer Technologies at Iskenderun Vocational School, Mustafa Kemal University. All subjects were notified in detail about ethical process before recordings and were informed about EEG in general, the aim of the study and the experiment protocol. They were warned about not to be medicated on the experiment day, not to move their eyes and/or limbs during the experiment, and only to focus on working mathematical process or reading text and not to worry about getting the exact answer or complete the text in order to minimize the EEG artifacts. EEG signals were recorded, in a noise-free, well-lighted room, from 22 electrodes constructing 26 channels, Fp1–A1, Fp2–A2, F3–A1, F4–A2, C3–A1, C4–A2, P3–A1, P4–A2, O1–A1, O2–A2, F7–A1, F8–A2, T3–A1, T4–A2, T5–A1, T6–A2, Fp2–O2, Fp1–O1, Fp1–Fp1, Fp1–Fp2, F7–F8, F3–F4, T3–T4, C3–C4, T5–T6 and P3–P4, with 1 kHz sampling frequency according to the international 10–20 protocol. Power line artifacts were automatically removed by EEG amplifier with 50 Hz stopband filter. Subjects were seated on a comfortable chair in front of 15-inch high-resolution screen during 60 slides shows each of which is 13.25 s long. First 30 of the slides include basic arithmetic processes, and the other 30 slides include a one paragraph easily readable Turkish text. Sixteen of the slides include arithmetic processes consisting of basic addition or subtraction (8 slides for each operator) of 4 and/or 5 digit numbers, and 14 of the slides include multiplication or division (7 slides for each operator) of 4 or 5 digit numbers by a one digit number. Minimum number of words in textual slides is 22, while the maximum number is 45, and the average is 29.83 ± 4.91. Sample slides are shown in Fig. 1.

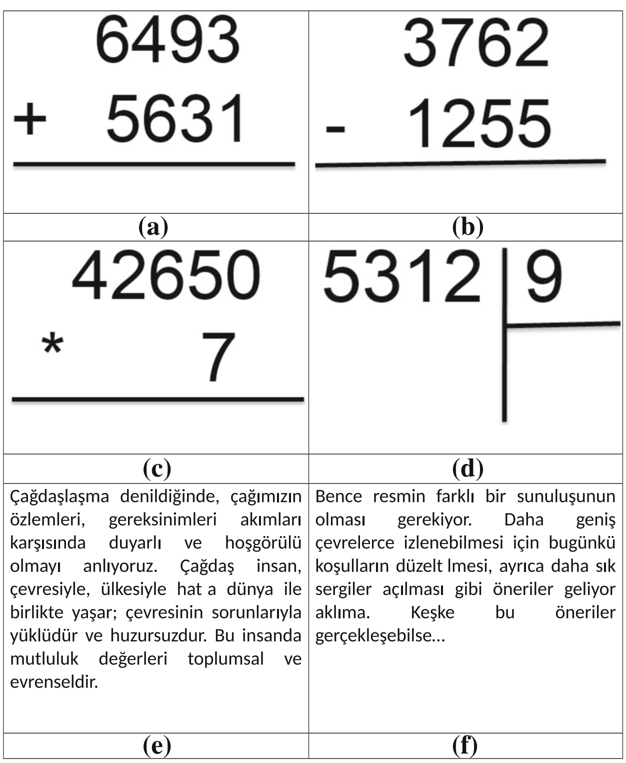

Fig. 1 Six presentation samples: **a** addition, **b** subtraction, **c** multiplication, **d** division, **e** long text and **f** short text

2.2 Preprocessing

EEG signals were filtered with 120 Hz low-pass filter to remove the high-frequency artifacts. Besides, the first and last one of mathematical and textual slides (1st, 30th, 31st and 60th slides) were removed to decrease the possible artifacts, and recorded signals were normalized between [−1,1].

3 Feature extraction and classification

3.1 Multi-order difference plot (MoDP)

In pattern recognition systems, extracted features directly affect the classification results. Time or frequency domain features or a combination of those can be used for this purpose. Frequency domain features, such as wavelet Fourier transform, and various classification schemes were used on the EEG-MaTeP database [5–8]. In this study, multi-order difference plot (MoDP), which is recently introduced by Kutlu et al. [17], is used for feature extraction. MoDP method is originated from Poincare [30] and second-order difference plot (SODP) [23, 27, 31] procedure. MoDP first scatters the sequential difference values with several degrees and then determines the number of values in determined regions.

3.1.1 Scattering of sequential difference values with several degrees

In this process, consecutive difference values with different degrees are scattered over the cartesian coordinate system. $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n] \varepsilon \Re^n$ is data matrix where $\mathbf{r}_i$ vectors are rows consisting of (x_i, y_i) coordinate pairs and n is the number of attributes. Difference values for degrees, $d = 1, \ldots, 4$, are calculated as follows:

$$
\begin{aligned}
&x_{1i} = x_i, y_{1i} = y_i,\\
&x_{2i} = x_{1(i+1)} - x_{1i}, y_{2i} = y_{1(i+1)} - y_{1i},\\
&x_{3i} = x_{2(i+1)} - x_{2i}, y_{3i} = y_{2(i+1)} - y_{2i},\\
&x_{4i} = x_{3(i+1)} - x_{3i}, y_{4i} = y_{3(i+1)} - y_{3i}, \, . \qquad (1)\\
&\ldots\\
&x_{di} = x_{(d-1)(i+1)} - x_{(d-1)i}\\
&y_{di} = y_{(d-1)(i+1)} - y_{(d-1)i}
\end{aligned}
$$

Note that this scheme is known as Poincare plot [30] when $d = 1$ and it is known as SODP [31] when $d = 2$. (x_{di}, y_{di}) data pairs for MoDP are computed as in Eq. (1). For this study, scattering of sequential difference values for $d \leq 4$ for $\forall d \in Z^+$ is used for feature extraction. Hence, four features 1st DP, 2nd DP, 3rd DP and 4th DP are obtained where DP means the difference plot.

3.1.2 Calculating values in specified regions

Data pairs, (x_{di}, y_{di}), determined in the first part are normalized between $[-1, 1]$ to fit into cartesian coordinate system, and values in specified regions are determined as described below.

In Cartesian coordinate system, 16 regions are specified using four circles (whose radius is 0.25, 0.5, 0.75 and 1, respectively), and numbers of scattered values in these areas are calculated analytically [27, 31]. These quadrants (shown in Fig. 2) are described as follows: 1st quadrant detects increase in both x_i and y_i, 2nd quadrant detects increase in x_i and decrease in y_i, 3rd quadrant detects decrease in both x_i and y_i, and 4th quadrant detects decrease in x_i and increase in y_i, with four different levels.

3.2 Extreme learning machine (ELM)

In this section, ELM [15] and the proposed method, namely luELM, are described.

Traditional ELM is a SLFN that has fixed number of nodes in input and output layers, and random number of nodes in hidden layer. Structure of ELM is shown in Fig. 3. Weights in the input layer are assigned randomly, and output layer weights are determined mathematically. ELM uses Moore–Penrose pseudoinverse conditions with singular value decomposition (SVD) for learning process. Let (x_i, t_i) be the sample set, with N distinct samples, where $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^{\mathrm{T}} \varepsilon \Re^n$ is the ith input sample and $t_i \varepsilon \Re^1$ is the desired output, single-hidden layer neural network with m hidden neurons, the output of the kth hidden layer is given by:

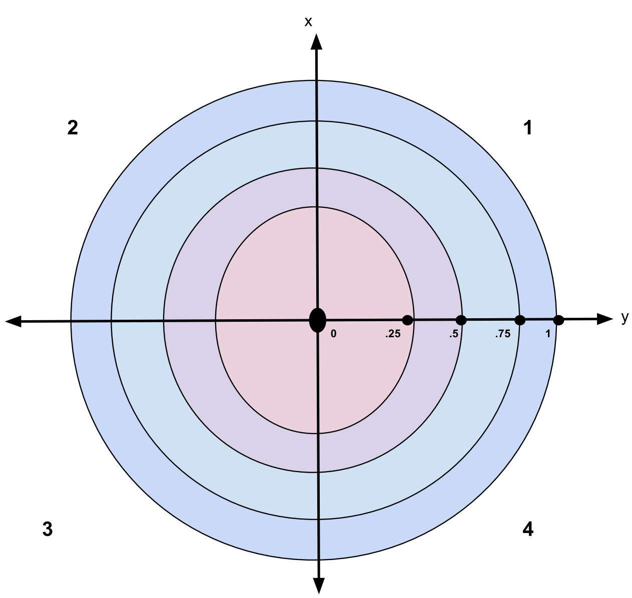

Fig. 2 Difference plot (DP) regions

$$\mathbf{h}_{ik} = g\left(\sum_{j=1}^{n} v_{ik} x_{ij} + b_k\right), \tag{2}$$

$$\mathbf{t} = \sum_{i=1}^{m} w_i h_i, \tag{3}$$

$$\mathbf{t} = \mathbf{H}\mathbf{w}, \tag{4}$$

where $g(\cdot)$ is activation function, $\mathbf{h}_i = [h_{i1} \ldots h_{im}]$ is output of hidden neurons, $\mathbf{v}_i = [v_{i1}, v_{i2}, \ldots, v_{in}]^T$ is input layer weight vector, $\mathbf{w} = [w_1, w_2, \ldots, w_m]^T$ is the output weight matrix, b_k is the bias value of the kth hidden neuron and t_i is the network output of the ith sample.

SLFN with m neurons in hidden layer approximates input samples with zero error such that $\sum_{j=1}^{N} ||o_j - t_j|| = 0$ where t_j is the network output and o_j is the desired output. Hidden layer neuron number m is selected randomly, such that $m \leq n$ to prevent overfitting, in traditional ELM. Hidden layer output weight vector, $\mathbf{w}$, is determined using Moore–Penrose pseudoinverse conditions. Therefore, algorithm to realize ELM is given as follows:

First, random input vectors $\mathbf{v}_i$ and bias are assigned. Then, the hidden layer output matrix $\mathbf{H}$ is calculated. Finally, the output weight vector $\mathbf{w}$ is calculated using:

$$\mathbf{w} = \mathbf{H}^{-1}\mathbf{t}. \tag{5}$$

Inverse of $\mathbf{H}$ cannot be determined directly if $\mathbf{H}$ is not a full-rank matrix. In this case, generalized inverse matrix, known as pseudoinverse, is used. In order to calculate the pseudoinverse, various approaches can be used. The most widely known type is the Moore–Penrose pseudoinverse [15]. Pseudoinverse of the matrix $\mathbf{H}$, namely $\mathbf{H}^+$, can be computed via least square solution:

$$\mathbf{H}^+ = (\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}. \tag{6}$$

This solution is accurate as long as square matrix $(\mathbf{H}^{\mathrm{T}}\mathbf{H})$ is invertible. However, it is singular in some applications. In SLFN, it will be singular in most of the cases since there is tendency to select $m << n$. In ELM, Huang et al.[15] have solved this problem using SVD [10] approach to compute the pseudoinverse. Although SVD can be used to calculate $\mathbf{H}^+$ in all cases, the algorithm is very slow when dealing with large data sets [14, 29]. In this study, lower–upper (LU) triangularization method is used to reach Moore–Penrose pseudoinverse of $\mathbf{H}$ matrix to overcome this problem. However, problem of $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ or $\mathbf{H}$ matrices being ill-conditioned or not-full rank still remains. Moreover, in this study this problem is solved using row-reduced

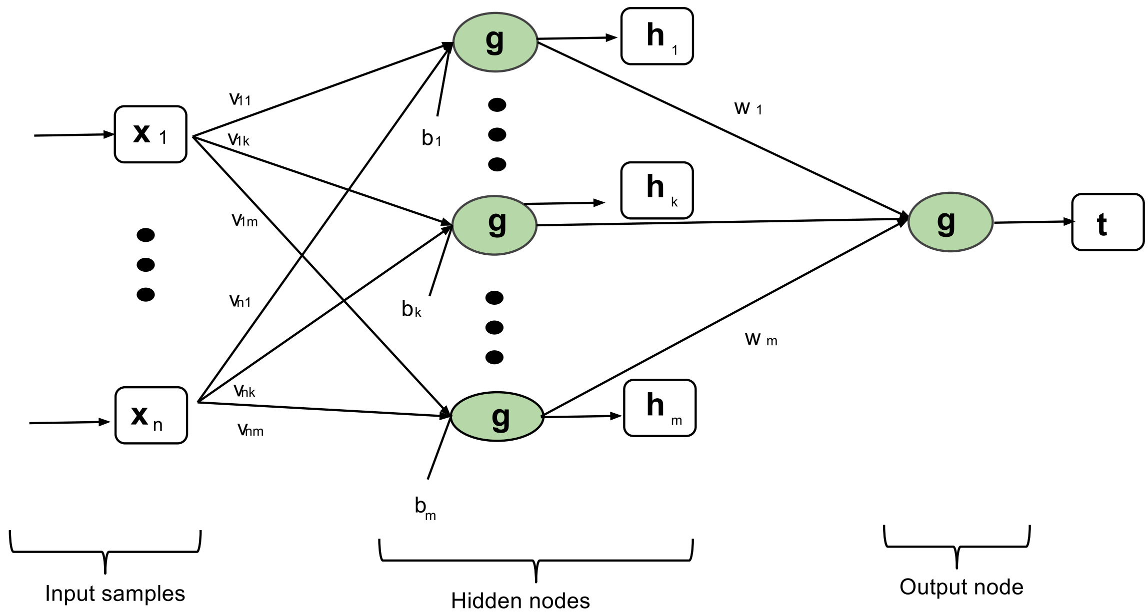

Fig. 3 Structure of ELM

echelon form of matrix which will be described in forward sections.

3.2.1 *Lower–upper triangularization and luELM*

Lower–upper (LU) triangularization method was introduced by Alan Turing in 1948 [2]. LU triangularization is used to determine determinant, inverse of matrix and solve linear equations. By LU triangularization, **H** matrix can be decomposed as $\mathbf{H} = \mathbf{LU}$, where **L** is a lower triangular matrix and **U** is an upper triangular matrix [11, 20]. Only if the matrix **H** is decomposed, $\mathbf{Hw} = \mathbf{t}$ can be solved directly. The overall procedure for solving $\mathbf{Hw} = \mathbf{t}$ is summarized as follows:

- Decompose **H**, such that $\mathbf{H} = \mathbf{LU}$. Hence, $\mathbf{LUw} = \mathbf{t}$.
- Let $\mathbf{Uw} = \mathbf{y}$ so that $\mathbf{Ly} = \mathbf{t}$. Solve this system for **y** using forward substitution.

$$\begin{aligned} y_1 &= t_1/L_{1,1} \\ y_2 &= (t_2 - (L_{2,1}y_1))/L_{2,2} \\ y_3 &= (t_3 - (L_{3,1}y_1) + (L_{3,2}y_2))/L_{3,3} \\ &\vdots \\ y_i &= \left(t_i - \sum_{j=1}^{i-1} L_{ij}y_j\right)/L_{ii} \end{aligned} \quad . \tag{7}$$

- Solve the triangular system $\mathbf{Uw} = \mathbf{y}$ for **w** using backward substitution.

$$\begin{aligned} w_e &= y_e/U_{e,e} \\ w_{e-1} &= (y_{e-1} - (U_{e-1,e}w_e))/U_{e-1,e-1} \\ w_{e-2} &= (y_{e-2} - (U_{e-2,e-1}w_{e-1}) + (U_{e-1,e}w_e))/U_{e-2,e-2} \\ &\vdots \\ w_i &= \left(y_i - \sum_{j=i+1}^{n} U_{ij}w_j\right)/U_{ii} \end{aligned} \quad , \tag{8}$$

where e is the end parameter and n is equal to length of y.

3.2.2 *Node number optimization*

Neuron numbers of a SLFN are iteratively changed from 1 to the number of attributes, and the neuron

Table 1 Subject-independent accuracies and time analysis

Classifier	Feature	Neuron	Accuracy (%)	Total time (s)
ELM	1	189	85.22	234.50
	2	164	84.26	214.00
	3	145	83.13	211.53
	4	217	81.05	211.56
luELM	1	165	87.80	12.64
	2	152	85.22	18.25
	3	190	85.71	13.84
	4	172	86.51	13.79

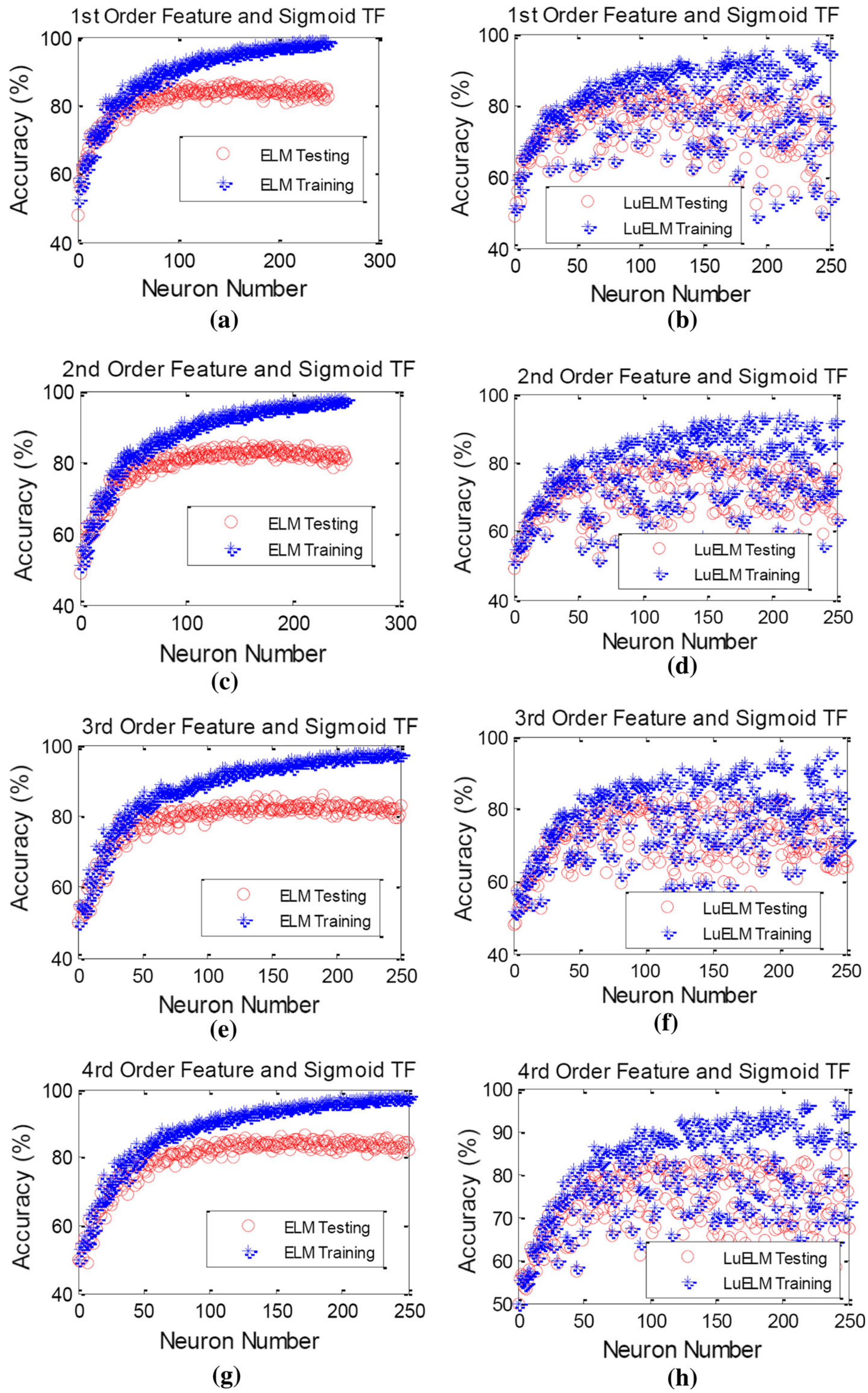

Fig. 4 ELM (1) classifiers with 1st DP **a**, 2nd DP **c**, 3rd DP **e** and 4th DP **g** features, luELM classifiers with 1st DP **b**, 2nd DP **d**, 3rd DP **f** and 4th DP **h** features train and test accuracies with iterative neuron numbers [transfer function (TF)]

number for which the classification result is the highest is selected. Neuron numbers giving the highest classification results for ELM and luELM are given in Table 1. Figure 4 shows the classification results for train and test data sets as the neuron number changes. In order to validate accuracy results, train and test data are specified according to *k*-fold cross-validation method [16]. *k* value is preferred as 10.

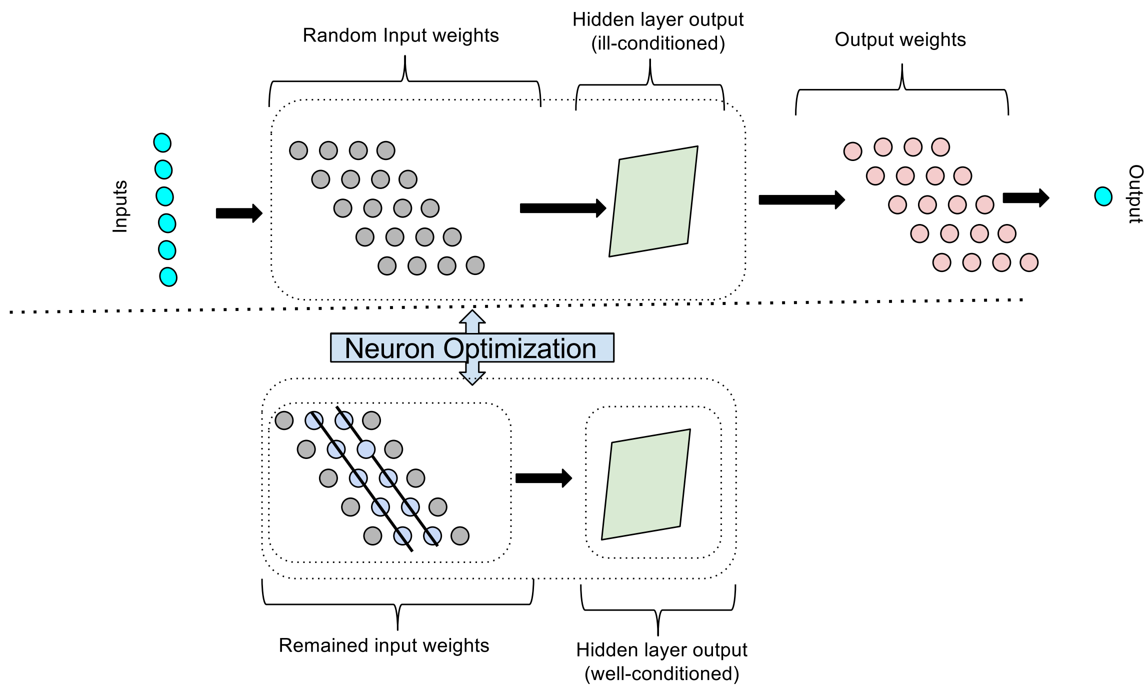

Fig. 5 Proposed neuron number optimization model and luELM

However, complexity of the system increases as the number of attributes increases. In this study, besides finding the optimum neuron number iteratively, row-reduced echelon (RRE) matrix form [11] is also used. Proposed structure is shown in Fig. 5. Performing row reducing on first layer output matrix, **H**, linearly dependent rows are eliminated and, hence, the neuron number is determined directly. Table 2 shows the results for 200 neurons, for optimum neuron number which gives the best classification results searched iteratively (from 1 to 200 neurons) and for the neuron number determined directly using RRE matrix form.

4 Results

In this study, a new approach for learning algorithm called luELM is proposed for SLFNs. In luELM, shortcomings of SVD in reaching pseudoinverse of hidden layer output in traditional ELM are overcome by using LU Triangularization. Besides, a new approach for identifying the optimum neuron number in hidden layer by using row-reduced echelon matrix form instead of choosing randomly, or iteratively, is proposed. This method is tested on classification of EEG signals collected during arithmetic and text reading processes. A recently introduced method, MoDP, which scatters the consecutive difference values with various degrees and then determines the number of values in specified regions, is used for feature extraction. In order to evaluate a system, either subject-dependent or subject-independent training is performed. For subject-dependent analysis, system should be trained and tested for each subject separately. In subject-independent analysis, on the other hand, all subjects are included in train and test sets. Most of the current studies with EEG data are designed by considering a single user. In case of insufficient data for training, such as having too little data during seizures and hours of seizure-free data for an epilepsy patient, or subjects that have difficulties for data collection, due to physical disabilities or unavailable laboratory conditions, etc., subject-independent analysis would help us build a better model and therefore would be more appropriate in constructing real-time BCI systems. Here, we performed both subject-dependent and subject-independent analyses, to be able to compare the results to the previous studies and to show the system behaviour when all subjects are included in the analysis at once. Subject-independent and subject-dependent classification results are given in Tables 1, 2, 3 and 4, respectively. In order to validate accuracy results, train and test data sets are specified using tenfold cross-validation method [16]. As expected, subject-independent results are lower than subject-dependent results, where data in training and testing phases belong to the same subject. Accuracies for each subject and the average accuracies are given in Tables 3 and 4 for ELM and luELM, respectively. As it is seen 100% accuracies in ELM and luELM, which are given in bold, prove that proposed machine learning system has ability to solve

Table 2 luELM classifier accuracy results when neuron number is optimized

Feature	Neuron	Remained	Accuracy (%)	Total time (s)
1	200	200	67.16	0.11
	200	196	86.01	4.47
	1–200	165	87.80	7.32
2	200	200	80.26	0.17
	200	192	80.06	4.32
	1–200	152	85.22	14.25
3	200	200	68.85	0.09
	200	194	68.25	4.14
	1–200	190	85.71	8.03
4	200	200	71.13	0.06
	200	192	79.76	4.70
	1–200	172	86.51	7.97

Initial and remained neuron numbers are shown

classification problem in cognitive tasks. Average subject-dependent true positive rates for ELM are 87.78, 92.05, 92.25 and 91.66% for first, second, third and fourth degree plots, respectively, while higher results, 95.55, 97.31, 96.96 and 97.83%, are obtained for luELM. Tables 1 and 2 show subject-independent accuracies. These results are average accuracies computed using tenfold cross-validation. Table 1 shows the accuracies for ELM and luELM for neuron number that is searched exhaustively for each feature set. Table 1 also shows the time period consumed for finding the neuron number that gives the best accuracy. Results for ELM/**luELM** are 85.22%/**87.80**%, 84.26%/**85.22**%, 83.13%/**85.71**% and 81.05%/**86.51**% are obtained for first, second, third and fourth difference plots, respectively. Analysing the total time periods consumed for finding the neuron number that results in best test accuracy, it can be seen that better performances are achieved for luELM in only about 1/15 of the time consumed for ELM. Total time for searching the optimum neuron number exhaustively, from 1 neuron to 250, exceeds 200 s (217.90 $\pm$ 11.13) for ELM for all cases, whereas it is between 12.64 and 18.25 s (14.63 $\pm$ 2.48) for luELM. In Table 2, luELM classifier accuracy results for 200 neurons, for neuron number that is reduced from 200 using RRE neuron optimization technique and optimum neuron number searched exhaustively between 1 and 200, are shown.

5 Discussion

Both subject-dependent and subject-independent analyses are performed to compare the results with previous studies and to show the system behaviour when all subjects are included in the analysis at once. In previous studies on MaTeP database, true positive rates of 89.1% [6] and 90% [8] were obtained for wavelet-based features and BayesNet classifier for all channels and most dominant 10 channels,

Table 3 Subject-dependent ELM accuracy results

Subject number	1st DP Acc/Nn	2nd DP Acc/Nn	3rd DP Acc/Nn	4th DP Acc/Nn
1	96.42/129	96.42/14	96.42/7	94.64/20
2	75.00/24	85.71/7	91.07/14	89.28/69
3	80.35/79	76.78/60	76.78/194	76.78/228
4	87.50/16	98.21/233	98.21/52	94.64/22
5	85.71/140	98.21/7	**100.00**/12	98.21/10
6	78.57/28	92.85/7	89.28/11	94.74/78
7	91.07/16	96.42/46	96.42/29	**100.00**/10
8	92.86/12	92.85/8	85.71/21	85.71/10
9	89.28/169	89.28/35	91.07/12	94.64/6
10	83.92/25	82.14/6	85.71/7	96.42/11
11	91.07/17	91.07/194	87.50/17	87.50/22
12	87.50/9	**100.00**/17	94.64/18	92.85/28
13	92.64/174	89.28/10	92.85/15	91.07/22
14	83.92/19	96.42/18	94.64/22	87.50/133
15	87.50/6	92.85/11	96.42/8	92.85/9
16	85.71/5	92.85/6	92.85/77	85.71/206
17	96.42/12	98.21/6	98.21/9	96.42/4
18	94.64/11	87.50/8	92.85/27	91.07/53
Average	87.78/49	92.05/38	92.25/30	91.66/52

Accuracy(overall)/neuron number (Acc/Nn)

Table 4 Subject-dependent luELM accuracy results

Subject number	1st DP Acc/Nn	2nd DP Acc/Nn	3rd DP Acc/Nn	4th DP Acc/Nn
1	**100.00**/24	98.00/13	96.66/177	98.00/24
2	89.33/213	94.33/35	94.33/7	96.00/22
3	92.66/154	93.00/23	79.00/9	98.33/130
4	96.33/19	**100.00**/31	**100.00**/17	**100.00**/35
5	96.67/167	**100.00**/14	**100.00**/5	**100.00**/232
6	91.33/27	98.33/213	98.33/20	**100.00**/157
7	**100.00**/31	**100.00**/25	**100.00**/193	**100.00**/154
8	96.66/16	96.66/30	96.66/17	94.66/19
9	**100.00**/250	98.33/26	**100.00**/30	98.33/14
10	93.00/27	89.33/4	92.66/194	98.33/11
11	91.33/28	94.33/29	96.33/233	96.33/17
12	90.33/27	**100.00**/14	**100.00**/18	**100.00**/45
13	**100.00**/29	98.33/8	98.33/22	96.66/27
14	96.66/199	**100.00**/25	**100.00**/30	95.00/23
15	98.33/141	98.00/15	**100.00**/224	98.00/25
16	87.33/128	98.33/240	94.66/5	96.33/233
17	**100.00**/29	**100.00**/16	**100.00**/9	**100.00**/5
18	**100.00**/18	94.66/20	98.33/32	95.00/25
Average	95.55/84	97.31/43	96.96/69	97.83/66

Accuracy(overall)/neuron number (Acc/Nn)

respectively. Besides, k-nn (k=1) and J48 decision tree classification results, when all channels are used, are given as 89.5 and 90.0%, respectively [5]. As it can be seen from tables both ELM and luELM classifiers with MoDP features give better performances than BayesNet, k-nn and J48 decision tree classifiers with wavelet features. Subject-dependent results show that MoDP features classified with either ELM or luELM classifiers give higher accuracies than wavelet transform-based features classified with BayesNet, k-nn and J48 decision tree classifiers. Similar inference can be made for subject-independent analysis since proposed luELM results in higher accuracies than ELM for all 1st, 2nd, 3rd and 4th degree plots features. Besides, analysing the total time periods in best test accuracies consumed for luELM is only about 1/15 of the time consumed for ELM. It is also worth mentioning that, besides giving better average accuracies, perfect classification result is obtained for more cases using luELM (26 cases in total) compared to ELM (for only 3 cases). Overall, luELM shows better performance than ELM for all cases. Comparing average accuracies and individual performances for each subject, better results are observed for luELM classifier than traditional ELM. In Table 1, it can be concluded that accuracies are higher than 80% for all cases, and all luELM results are better than ELM results. Maximum accuracy (87.80%) for subject-independent analysis is obtained using 1st degree difference plot features and luELM classifier. The results also show that luELM increases the accuracies for all scenarios. In Table 1, it is clear that when comparing consumed time periods for ELM and luELM, 93.29% time gain is observed. The results obtained using neuron number found directly using RRE neuron optimization are comparable only for 1st DP feature set while the time consumed is decreased about to half.

6 Conclusion

In this study, we have introduced a classification scheme, luELM, to improve the traditional ELM by employing lower–upper triangularization method to reach Moore–Penrose pseudoinverse of hidden layer output matrix instead of using SVD, which is a slow method when working with huge data sets. The proposed algorithm is applied with optimum neuron numbers searched iteratively and RRE form matrix. A recently introduced feature extraction method, namely MoDP, which uses scattering of consecutive difference values with different degrees of time-domain signal, is employed. ELM and luELM methods are compared in terms of accuracies of classifying time-domain features based on MoDPs (using different degrees) extracted from EEG data recorded during arithmetic and text processing. The results show that the proposed methods outperform the other approaches in classification accuracy and duration. Also, this paper demonstrated that using EEG signal arithmetic and text processing behaviours can be classified. Moreover, it should inspire researchers to investigate classifying different processing behaviours in human brain and to focus on determining novel algorithms to calculate pseudoinverse in traditional ELM. In fact, it is very likely that present methods in linear algebra can be tried or further improved, and we are waiting eagerly to seeing new works on this field.

Acknowledgements The authors are very grateful to Mr. Server Göksel Eraldemir for his efforts to construct the EEG-MaTeP database.

Compliance with ethical standards

Conflicts of interest The authors declare that there is no conflict of interest.

References

1. Baretta L, Tomitch LMB, MacNair N, Lim VK, Waldie KE (2009) Inference making while reading narrative and expository texts: An ERP study. Psychol Neurosci 2(2):137
2. Bunch JR, Hopcroft JE (1974) Triangular factorization and inversion by fast matrix multiplication. Math Comput 28(125):231–236
3. Duda RO, Hart PE, Stork DG (2001) Pattern classification. 2nd edition. A Wiley-Interscience publication
4. Eberhard-Moscicka AK, Jost LB, Raith M, Maurer U (2015) Neurocognitive mechanisms of learning to read: print tuning in beginning readers related to word-reading fluency and semantics but not phonology. Dev Sci 18(1):106–118
5. Eraldemir SG (2014) Analysis of EEG signals during text reading and mathematical processing
6. Eraldemir SG, Yildirim E (2014) Classification of simple text reading and mathematical tasks from EEG. In: Signal processing and communications applications conference (SIU). IEEE, pp 180–183
7. Eraldemir SG, Yıldırım E, Kutlu Y (2014) Classification of mathematical tasks from EEG signals using k-NN algorithm. In: Elektrik Elektronik Bilgisayar ve Biyomedikal Mhendislii Sempozyumu (Eleco 2014), Bursa, Turkey. pp 551–554
8. Eraldemir SG Yıldırım E, Yıldırım S, Kutlu Y (2014) Kognitif EEG isaretleri icin oznitelik secimi tabanli kanal secimi ve siniflandirma. In: Akilli Sistemlerde Yenilikler ve Uygulamalari Sempozyumu (ASYU). Izmir, Turkey
9. Galán FC, Beal CR (2012) EEG estimates of engagement and cognitive workload predict math problem solving outcomes. In: User modeling, adaptation, and personalization. Springer, pp 51–62
10. Golub GH, Reinsch C (1970) Singular value decomposition and least squares solutions. Numer Math 14(5):403–420
11. Golub GH, Van Loan CF (2012) Matrix computations. 4th edition. JHU Press, Baltimore
12. Gundel A, Wilson GF (1992) Topographical changes in the ongoing EEG related to the difficulty of mental tasks. Brain Topogr 5(1):17–25
13. Hall MA (1999) Correlation-based feature selection for machine learning. Ph.D. thesis, The University of Waikato
14. Horata P, Chiewchanwattana S, Sunat K (2013) Robust extreme learning machine. Neurocomputing 102:31–44
15. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70:489–501. doi:10.1016/j.neucom.2005.12.126
16. Krogh A, Vedelsby J et al (1995) Neural network ensembles, cross validation, and active learning. Adv Neural Inf Process Syst 7:231–238
17. Kutlu Y, Yayik A, Yildirim E, Yildirim S (2015) Orthogonal extreme learning machine based P300 visual event-related BCI. In: Neural information processing. Springer, pp 284–291
18. Liang NY, Huang GB, Saratchandran P, Sundararajan N (2006) A fast and accurate online sequential learning algorithm for feedforward networks. IEEE Trans Neural Netw 17(6):1411–1423
19. Lin CL, Jung M, Wu YC, Lin CT, She HC (2012) Brain dynamics of mathematical problem solving. In: Engineering in Medicine and Biology Society (EMBC), 2012 annual international conference of the IEEE. IEEE, pp 4768–4771
20. Meyer CD (2000) Matrix analysis and applied linear algebra. Siam publication
21. Osaka M (1984) Peak alpha frequency of EEG during a mental task: task difficulty and hemispheric differences. Psychophysiology 21(1):101–105
22. Pachori RB, Hewson D, Snoussi H, Duchêne J (2009) Postural time-series analysis using empirical mode decomposition and second-order difference plots. In: IEEE international conference on acoustics, speech and signal processing, 2009. ICASSP 2009. IEEE, pp 537–540
23. Pachori RB, Patidar S (2014) Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode functions. Comput Methods Programs Biomed 113(2):494–502. doi:10.1016/j.cmpb.2013.11.014
24. Ramsden S, Richardson FM, Josse G, Shakeshaft C, Seghier ML, Price CJ (2013) The influence of reading ability on subsequent changes in verbal IQ in the teenage years. Dev Cogn Neurosci 6:30–39
25. Regression AA (1972) The Moore–Penrose pseudoinverse. Academic, New-York
26. Rosazza C, Cai Q, Minati L, Paulignan Y, Nazir TA (2009) Early involvement of dorsal and ventral pathways in visual word recognition: an ERP study. Brain Res 1272:32–44
27. Thuraisingham R (2010) A classification system to detect congestive heart failure using second-order difference plot of RR intervals. Cardiol Res Pract. doi:10.4061/2009/807379
28. Thuraisingham RA (2009) A classification system to detect congestive heart failure using second-order difference plot of RR intervals. Cardiol Res Pract 2009:1–8
29. Tzeng J (2013) Split-and-combine singular value decomposition for large-scale matrix. J Appl Math. doi:10.1155/2013/683053
30. Yayık A, Kutlu Y (2011) Diagnosis of congestive heart failure using poincare map plot. In: Signal process and communication conference
31. Yayık A, Kutlu Y, Yıldırım E (2014) Epileptic state detection: pre-ictal, inter-ictal, post-ictal. In: International conference on advanced technology and sciences
32. Zhu QY, Qin AK, Suganthan PN, Huang GB (2005) Evolutionary extreme learning machine. Pattern Recogn 38(10):1759–1763