VLSI design of multiclass classification using sparse extreme learning machine for epilepsy and seizure detection

Yuanfa Wang1, 2, Qianneng Zhou1, Jiasai Luo1, Yi Lu1, Huiqian Wang1, Yu Pang1, and Zhiwei Huang1, 2, a)

Abstract An automatic detection system for distinguishing healthy, ictal, and interictal EEG signals is of importance in clinical practice. This paper presents a low-complexity three-class classification VLSI system for epilepsy and seizure detection. The designed system consists of a discrete wavelet transform (DWT)-based feature extraction module, a sparse extreme learning machine (SELM) training module, and a multiclass classifier module. A lifting structure of Daubechies order 4 wavelet is introduced in three-level DWT to save circuit area and speed up the computational time. The SELM which is a novel machine learning algorithm with low hardware complexity and high-performance is used for on-chip training. One-against-one multiclass non-linear SELM is designed for the first time due to its high classification accuracy. The designed system is implemented on an FPGA platform and evaluated using the publicly available epilepsy dataset. The experimental results demonstrate that the designed system achieves high accuracy with low-dimensional feature vectors.

Keywords: low-complexity, classification, DWT, multiclass, SELM

1. Introduction

Approximately 50 million people in the world suffer from epilepsy. Epileptic seizure impacts the quality of life for patients and their families, and even leads to unexpected death [1, 2]. So, detecting epilepsy and seizure with high efficiency is very necessary for patients. The electroencephalogram (EEG) provides the temporal and spatial information of brains’ electrical voltages to be used to diagnose epilepsy [3, 4]. Traditional seizure detection methods rely on highly trained professionals to visually inspect lengthy EEG recordings, which is time-consuming and inconvenient. Additionally, there are a number of conditions that look like epileptic seizures but are not. Thus, automatic detection of epilepsy and seizure is of great clinical significance. Detecting interictal (seizure free) EEG signals can be used to diagnose epilepsy in a clinical setting, and then with prescribed medicine for interictal patients we can significantly reduce their dangers and additionally, the detection of seizure is of importance for instant treatment [5, 6, 7]. Therefore, automatic classification of healthy, ictal (seizure), and interictal EEG signals is of great clinical significance.

With the rapid development of wearable device, it is necessary to develop a portable automatic seizure detection system with on-chip learning. Since the EEG signals of seizure patterns differ significantly between patients and between different ages of a patient, integrating a machine learning algorithm for the portable seizure detection system with on-chip learning is very valuable to be trained in time to adapt to the patient-to-patient and age-to-age variations using the patient-specific and up-to-date EEG data [1, 8]. Therefore, selecting a machine learning algorithm with low complexity and high accuracy is very important.

Currently, algorithms for seizure detection based on machine learning have been proposed in a large number of publications [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]. Artificial neural network (ANN) algorithms, such as the back-propagation (BP), convolutional neural network (CNN) and deep learning, have been widely used to classify epileptic EEG signals [18, 19, 20, 21, 22]. Although ANN has been found to exhibit good performance in seizure detection, it requires an extensive training process and a complicated design procedure since the connection weights and biases are time-consuming to be adjusted especially for lots of hidden layers [17, 22, 23, 24]. Research results [23] have shown that the average execution time of the BPNN is far more than that in other algorithms. Moreover, the conventional learning algorithms for ANN are prone to fall into a local minimum [23]. Therefore, the learning speed for conventional ANN is too slow to meet the requirements of seizure detection [23]. Thus, designing a system with quick in classification and response in the epilepsy detection is very important since the seizures occur suddenly in a few seconds. Support vector machine (SVM), which aims to find the optimal separating hyperplane on the basis of the statistical learning theory, is another popular machine learning method for epileptic EEG classification [10, 16]. However, the training speed of the SVM will be very slow when the sample size is large [25]. Extreme learning machine (ELM) is an emerging training algorithm proposed by Huang et al. [15, 26, 27]. The advantages of ELM include global searching ability, one-step learning, and fast learning speed since the hidden node parameters are randomly generated and the output weights are analytically computed. For hardware realization, however, the initial ELM would consume large storage space in calculating the inverse of matrix [15, 28]. Sparse ELM (SELM) is an extension method based on ELM, which is proposed by Bai et al. [29]. In the SELM, as fewer constraints need to be satisfied, and only one Lagrange variable needs to be up-
dated in each iteration, the training process would be easier, so the less storage space will be consumed if implemented in hardware. In order to classify three class EEG signals, multiclass SELM is necessary. However, the SELM was originally designed for binary classifier, while its multiclass problem can be implemented by combining several binary SELM classifiers. Similar to multiclass SVM, one-against-one method can be utilized to combine three binary SELM classifiers together for three-class SELM.

The feature extraction of EEG signals plays an important role in increasing the classification accuracy of three-class SELM [17, 29, 30, 31]. Several methods such as time domain, frequency domain, time-frequency domain, and nonlinear analysis have been developed for feature extraction in epileptic EEG classification [32, 33]. Empirical mode decomposition (EMD) and discrete wavelet transform (DWT) have been widely used for analyzing nonlinear and nonstationary EEG signals [34]. However, calculating cubic spline interpolation is a complex process in the hardware implementation of EMD. The DWT can be designed based on lifting architecture, which costs less hardware resource. The Daubechies-4 (db4) mother wavelet is used in the lifting-based DWT because its wave characteristic is similar to the spike wave of the epileptic EEG signals [35]. In this paper, three-level lifting-based discrete wavelet transform (LDWT) using Daubechies order 4 wavelet is introduced to decompose EEG signals into delta, theta, alpha, beta sub-bands. Subsequently, eight-dimensional feature vectors are created by computing maximum and standard deviation values of each sub-band. Finally, the eight-dimensional feature vectors are fed into SELM.

To the best of our knowledge, this is the first work to implement a three-class classification based on LDWT and multiclass non-linear SELM in hardware with on-chip training capability for detecting epilepsy and seizure. In order to reduce the computational complexity and hardware cost, several techniques are used in the circuit design.

2. SELM algorithm

The training of SELM is similar to SVM, but the training process would be easier than SVM, since fewer constraints need to be satisfied and only one Lagrange variable needs to be updated in each iteration. The training algorithm of the SELM is summarized as follows:

Step 1: Find the minimum of \( J_i \) and choose the index \( c \) of the updated Lagrange multiplier by \( c = \text{arg min}_i J_i \), where \( J_i = g_i d_i \) and \( d_i \) expressed as

\[
\begin{align*}
    d_i &= \begin{cases} 
    1, & \text{if } 0 < \alpha_i < C \\
    \text{sign}(g_i), & \text{if } \alpha_i = 0 \\
    -1, & \text{if } \alpha_i = C
    \end{cases}
\end{align*}
\]

Step 2: Update Lagrange multiplier \( \alpha_c \)

\[
\alpha_c^{new} = \alpha_c^{old} - g_c^{old}
\]

The unconstrained \( \alpha_c^{new} \) must be determined in the feasible range \([0, C]\)

The unconstrained \( \alpha_c^{new} \) is the minimum of \( g_i d_i \), where \( J_i = g_i d_i \) and \( d_i \) expressed as

\[
\begin{align*}
    d_i &= \begin{cases} 
    1, & \text{if } 0 < \alpha_i < C \\
    \text{sign}(g_i), & \text{if } \alpha_i = 0 \\
    -1, & \text{if } \alpha_i = C
    \end{cases}
\end{align*}
\]

Step 3: Update Lagrange multiplier \( \alpha_c \)

\[
\alpha_c^{new} = \alpha_c^{old} - g_c^{old}
\]

The unconstrained \( \alpha_c^{new} \) must be determined in the feasible range \([0, C]\)

The training algorithm using pseudo-code is shown in Table I.

| Table I | The SELM training algorithm using pseudo-code. |
|---------|---------------------------------------------|
| Initial step: \( \alpha_0, g_0, J_0, d_0 \) where \( i = 1, \ldots, N \). |
| Loop begin |
| Step 1: Obtain the minimum of \( J_i \) and the index \( c \) of the updated Lagrange multiplier by \( c = \text{arg min}_i J_i \). |
| Step 2: Update Lagrange multiplier \( \alpha_c \) by \( \alpha_c^{new} = \alpha_c^{old} - g_c^{old} \). |
| Step 3: Update \( g_i \) and \( J_i \) for \( i = 1, 2, \ldots, N \). |
| End loop (if \( \min J_i > -0.001 \)). |
| Final step: Save Lagrange multipliers into classifier model. |

3. Hardware implementation for three-class classification

3.1 System architecture

Fig. 1 shows the architecture of the proposed three-class classification, which consists of a MUX Bank, a LDWT-based feature extractor, a DEMUX, an SELM trainer and a Three-class SELM classifier. The operation of the signal processing path is described as follows. MUX Bank block selects the training data and labels or testing data. The training data is decomposed into four sub-bands through the three-level LDWT, and then an eight-dimensional feature vector is created by computing the maximum and standard deviation values of the wavelet coefficients. Three SELM classifier models are derived from SELM Trainer block. The classifier...
models are stored in the memory after they are trained using the training data and labels. After training, the testing data is also decomposed into four sub-bands through three-level LDWT, and eight-dimensional feature vectors are created. Hence, a hardware sharing technique is used. The eight-dimensional feature vectors and the pre-trained multiclass SELM classifier models are used as the Three-class SELM classifier inputs. A majority voting strategy is used in the Three-class SELM classifier to identify the testing data.

3.2 LDWT-based feature extraction

The EEG data are first processed by the LDWT-based feature extraction module. Fig. 2 shows the structure of the three-level LDWT. The three-level LDWT decomposes the EEG signal into four sub-bands, generating the approximation coefficients $A_3$ corresponding to the delta wave, detail coefficients $D_1$ corresponding to the beta wave, $D_2$ corresponding to the alpha wave, and $D_3$ corresponding to the theta wave. Fig. 3 shows the db4 LDWT circuit structure. The EEG signal sequences $X[n]$ are divided into two disjoint subsets of even and odd samples, respectively. Then the wavelet coefficients can be obtained using lifting steps. In hardware design of the LDWT, the multipliers can be implemented by shift-add operations, since the wavelet coefficients $-\sqrt{3}, \frac{\sqrt{3}}{4}, \frac{-\sqrt{3}}{4}$ and $\frac{1+\sqrt{3}}{2}$ are constant. After decomposing the EEG data using LDWT, eight-dimensional feature vectors are created by calculating maximum and standard deviation values of the four coefficients.

3.3 SELM trainer

In design of the three-class classification architecture, SELM trainer is the main unit. The architecture of SELM trainer is composed of a Lagrange multiplier updating block, a $g_i$ updating block, a control block, and a block of updating $J_i$, $d_i$ and finding $c$. Control block is a finite state machine (FSM) that controls the whole flow of the SELM algorithm as shown in Fig. 4.

The S0 is used to initialize the parameters $\alpha_1 = 0, g_1 = -1, J_i = -1, d_i = 1 (i = 1, \ldots, N)$. When the initialization is completed, the current state changes to S1, which is used to find the minimum of $J_i$, and the index $c$ of the updated Lagrange multiplier. State S2 is used to calculate $\alpha_c$. When the calculation is completed, the state then changes to S3 to update $g_i, d_i$ and $J_i$. Then the state changes to the next iteration. The training procedure is terminated when $\min_{i=1,\ldots,N} J_i > -0.001$, which could ensure great accuracy.

When the index $c$ of the updated Lagrange multiplier in each iteration is chosen, the unclipped Lagrange multiplier $\alpha_{c,\text{new}}$ is first updated by (2), and then the clipped Lagrange multiplier $\alpha_{c,\text{new,clipped}}$ is bounded in the range from 0 to $C$ by (3). As can be seen from Fig. 5, one subtractor, two comparators, and a pair of two-way multiplexers are used in this sub-module, which greatly saves the area of the system.

The circuit design of $g_i$ update is shown in Fig. 6. In order to reduce multiplication, two multiplexers are used. In Fig. 7, R is the interrupt signal. If the trigger condition is satisfied, R is set to 1, and the training procedure will be terminated.
4. Experimental results

The normal (set A), inter-ictal (set D), and ictal (set E) EEG dataset from Epilepsy Center at the University of Bonn, Germany, is used to verify the effectiveness of the presented epilepsy and seizure detection system [38]. Each dataset (set A, set D or set E) contains 100 single-channel EEG segments, with 4096 points in each segment. In data preprocessing, every segment is divided into 512-point sliding time epochs with 256-point overlap between adjacent epochs, the length of each epoch is 2.94 s, and there is an overlap of 1.47 s between adjacent epochs [15]. Moreover, 1000 and 600 epochs in each subset are randomly chosen as training and testing epochs, respectively. The performance of the three-class SELM system can be evaluated by sensitivity, specificity and accuracy, which are defined as follows [16]. Sensitivity=number of true positive decisions/number of actually positive cases. Specificity=number of true negative decisions/number of actually negative cases. Accuracy=number of correct decisions/total number of cases. Table II shows the sensitivity, specificity and accuracy results of the proposed three-class classification using software implementation. Table III shows the comparison of the proposed SELM and Gaussian SVM in classification accuracy, training time and testing time using the same feature vectors, training and testing EEG data. The designed SELM achieves higher accuracy with shorter training and testing time than SVM. The synthesizable Verilog-HDL description of the system is coded. The ISE 14.7 software and Xilinx Spartan 6 XC6SLX150-FGG484 FPGA are applied to implement the detection system. The resources utilized by the FPGA implementation in terms of the numbers of Slice Register, Slice Look-up Tables (Slice LUTs), and block RAMs (BRAMs) are given in Table IV.

Table V summarizes the comparison of this work with previous works using hardware implementation. In Table V, all of the comparisons are done using the same datasets (A, D and E). In Ref. [7], principal component analysis (PCA) is used to automatically reduce the feature dimension, and then ANN is used for epilepsy and seizure detection. In Ref. [20], DWT is used for feature extraction and genetic algorithm (GA) is used for feature selection, and then ANN and SVM are used for classification. However, PCA causes high complexity in calculating the matrix if implemented in hardware, and GA has relatively complex calculation process. As shown in Table VI, the proposed system achieves higher accuracy than [7] and uses fewer features than [20]. In addition, the system implemented in
hardware is faster than that implemented in software.

This work is the first time to implement the on-chip training multi-class SELM based on LDWT to classify seizure and epilepsy EEG signals. Our system achieves relatively high accuracy, and reduces training and testing time by adopting non-linear SELM and low-dimensional feature vectors.

5. Conclusion

A low-complexity three-class classification VLSI system based on LDWT and non-linear SELM for epilepsy and seizure detection is presented in this paper. Using the three-level LDWT, the EEG data are decomposed into four sub-bands. Then eight-dimensional feature vectors are extracted by calculating maximum and standard deviation values of each sub-band. Finally, the eight-dimensional feature vectors are input to the multi-class SELM. The three-class classification system is tested using the publicly available epilepsy dataset including normal, seizure activity, and seizure free EEG signals. The experimental results demonstrate that the designed three-class classification system achieves high classification accuracy, and reduces training and testing time by combining LDWT and the SELM with low computational complexity. The results also indicate that the designed three-class classification outperforms some previous methods.

Acknowledgments

This work is supported by the Science and Technology Research Program of Chongqing Municipal Education Commission under Grant Nos. KJQN202000643 and KJQN202000604, Natural Science Foundation Project of CQ CSTC under Grant Nos. cstc2021jcystx0590, cstc2019jcyj-msxmX0696, and Medicine & Engineering & Informatics Fusion and Transformation Key Laboratory of Luzhou City under Grant Nos. XGY202102, XGY202101, XGY202103 and XGY202104.

References

[1] J. Yoo, et al.: “An 8-channel scalable EEG acquisition SoC with patient-specific seizure classification and recording processor,” IEEE J. Solid-State Circuits 48 (2013) 214 (DOI: 10.1109/JSSC.2012.2221220).
[2] M.A.B. Alta, et al.: “A 1.83 J/classification, 8-channel, patient-specific epileptic seizure classification SoC using a non-linear support vector machine,” IEEE Trans. Biomed. Circuits Syst. 10 (2016) 49 (DOI: 10.1109/TBCAS.2014.2386891).
[3] S.Y. Lee, et al.: “A programmable EEG monitoring SoC with optical and electrical stimulation for epilepsy control,” IEEE Access 8 (2020) 92196 (DOI: 10.1109/ACCESS.2020.2994217).
[4] W.-M. Chen, et al.: “A fully integrated 8-channel closed-loop neural-prosthetic CMOS SoC for real-time epileptic seizure control,” IEEE J. Solid-State Circuits 49 (2014) 232 (DOI: 10.1109/JSSC.2013.2284346).
[5] Y. Li, et al.: “Epileptic seizure detection in EEG signals using a unified temporal-spectral-squeeze-and-excitation network,” IEEE Trans. Neural Syst. Rehabil. Eng. 28 (2020) 782 (DOI: 10.1109/TNSRE.2020.2973434).
[6] Y. Yuan, et al.: “A multi-view deep learning framework for EEG seizure detection,” IEEE J. Biomed. Health Inform. 23 (2019) 83 (DOI: 10.1109/JBHI.2018.2871678).
[7] S. Ghosh-Dastidar, et al.: “Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection,” IEEE Trans. Biomed. Eng. 55 (2008) 512 (DOI: 10.1109/TBME.2007.905490).
[8] L. Feng, et al.: “VLSI design of SVM-based seizure detection system with on-chip learning capability,” IEEE Trans. Biomed. Circuits Syst. 12 (2018) 171 (DOI: 10.1109/TBCAS.2017.2762721).
[9] K. Lee, et al.: “A low-power processor with configurable embedded machine-learning accelerators for high-order and adaptive analysis of medical-sensor signals,” IEEE J. Solid-State Circuits 48 (2013) 1625 (DOI: 10.1109/JSSC.2013.2253226).
[10] S.A. Huang, et al.: “A 1.9-mW SVM processor with on-chip active learning for epileptic seizure control,” IEEE J. Solid-State Circuits 55 (2020) 452 (DOI: 10.1109/JSSC.2019.2947775).
[11] S. Khammohammadi: “Adaptive seizure onset detection framework using a hybrid PCA-CSP approach,” IEEE J. Biomed. Health Inform. 22 (2018) 154 (DOI: 10.1109/JBHI.2017.2703873).
[12] V. Joshi: “Classification of ictal and seizure-free EEG signals using fractional linear prediction,” Biomedical Signal Processing and Control 9 (2014) 1 (DOI: 10.1016/j.bspc.2013.08.006).
[13] A.K. Tiwari: “Automated diagnosis of epilepsy using key-point based local binary pattern of EEG signals,” IEEE J. Biomed. Health Inform. 21 (2017) 888 (DOI: 10.1109/JBHI.2016.2589971).
[14] S. Surpaya: “Automatic epilepsy detection from EEG introducing a new edge weight method in the complex network,” Electron. Lett. 52 (2016) 1430 (DOI: 10.1049/el.2016.1992).
[15] J.L. Song, et al.: “Automated detection of epileptic EEGs using a novel fusion feature and extreme learning machine,” Neurocomputing 175 (2016) 383 (DOI: 10.1016/j.neucom.2015.10.070).
[16] Y. Tang: “A tunable support vector machine assembly classifier for epileptic seizure detection,” Expert Syst. Appl. 39 (2012) 3925 (DOI: 10.1016/j.eswa.2011.08.088).
[17] A.S. Muthanantha Murugavel and S. Ramakrishnan: “Hierarchical multi-class SVM with ELM kernel for epileptic EEG signal classification,” Med. Biol. Eng. Comput. 54 (2016) 149 (DOI: 10.1007/s11517-015-1351-2).
[18] R. Sari, et al.: “FPGA-based real-time epileptic seizure classification using artificial neural network,” Biomedical Signal Processing and Control 62 (2020) 102106 (DOI: 10.1016/j.bspc.2020.102106).
[19] W. Mardini, et al.: “Enhanced detection of epileptic seizure using EEG signals in combination with machine learning classifiers,” IEEE Access 8 (2020) (DOI: 10.1109/ACCESS.2020.2970012).
[20] M. Omidvar, et al.: “EEG signal processing for epilepsy seizure detection using 5-levelDb4 discrete wavelet transform, GA-based feature selection and ANN/SVM classifiers,” Journal of Ambient Intelligence and Humanized Computing 4 (2021) 1 (DOI: 10.1007/s11759-020-03564-0).
[21] R. Elakkiya, et al.: “Machine learning based intelligent automated neonatal epileptic seizure detection,” Journal of Intelligent & Fuzzy Systems 40 (2021) 8847 (DOI: 10.29918/jifs.200800).
[22] M. Rashid-Al-Mahfuz, et al.: “A deep convolutional neural network method to detect seizures and characteristic frequencies using epileptic electroencephalogram (EEG) data,” IEEE J. Transl. Eng. Health Med. 9 (2021) 1 (DOI: 10.1109/JTEHM.2021.3059025).
[23] L. Qin, et al.: “A novel framework based on clustering for automatic epileptic seizure detection,” International Journal of Machine Learning and Cybernetics 10 (2017) 1 (DOI: 10.1007/s13042-017-0716-2).
[24] R.H. Kumar, et al.: “FPGA implementation of wavelet neural network for epileptic seizure detection,” IEICE Electron. Express 10 (2013) 848 (DOI: 10.1587/elex.2013.00848).
[25] J. Dass, et al.: “Distributed training of support vector machine on a multiple-FPGA system,” IEEE Trans. Comput. 69 (2020) 1015 (DOI: 10.1109/TC.2020.2993552).
[26] G.-B. Huang: “Extreme learning machines: a survey,” Int. J. Mach. Learn. & Cyber. 2 (2011) 107 (DOI: 10.1007/s13042-011-0019-y).
[27] A. Safaei, et al.: “System-on-a-chip (SoC)-based hardware acceleration for an online sequential extreme learning machine (OS-ELM),” IEEE Trans. Comput.-Aided Design Integr. Circuits Syst. 38 (2019) 99 (DOI: 10.1109/TCAD.2018.2876162).
[28] S. Decherchi, et al.: “Efficient digital implementation of extreme learning machines for classification,” IEEE Trans. Circuits Syst. II, Exp. Briefs 59 (2012) 496 (DOI: 10.1109/TCSII.2012.2204112).
[29] Z. Bai, et al.: “Sparse extreme learning machine for classifi-
[30] Y. Xu, et al.: “An energy-efficient parallel VLSI architecture for SVM classification,” IEEE Electron. Express 15 (2018) 20180099 (DOI: 10.1587/elex.15.20180099).

[31] G.-B. Huang, et al.: “Extreme learning machine for regression and multiclass classification,” IEEE Trans. Syst. Man. Cybern. B 42 (2012) 513 (DOI: 10.1109/tsmcb.2011.2168604).

[32] S. Kayhan, et al.: “ECG denoising on bivariate shrinkage function exploiting interscale dependency of wavelet coefficients,” Turk J. Elec. Eng. & Comp. Sci. 19 (2011) 495 (DOI: 10.3906/elk-0908-176).

[33] H. Mahdavi, et al.: “Area-time-power efficient FFT architectures based on binary-signed-digit cordic,” IEEE Trans. Circuits Syst. I, Reg. Papers 66 (2019) 3874 (DOI: 10.1109/TCSI.2019.2922988).

[34] F. Riaz, et al.: “EMD-based temporal and spectral features for the classification of EEG signals using supervised learning,” IEEE Trans. Neural Syst. Rehabil. Eng. 24 (2016) 28 (DOI: 10.1109/TNSRE.2015.2441835).

[35] C.H. Hsia, et al.: “Memory-efficient hardware architecture of 2-D dual-mode lifting-based discrete wavelet transform,” IEEE Trans. Circuits Syst. Video Technol. 23 (2013) 671 (DOI: 10.1109/TCSVT.2012.2211953).

[36] Y.S. Wang, et al.: “Area-energy efficient CORDICs using new elementary-angle-set and base-2 exponent expansions scheme,” IEICE Electron. Express 18 (2021) 20200409 (DOI: 10.1587/exle.17.20200409).

[37] L. Chen, et al.: “Algorithm and design of a fully parallel approximate coordinate rotation digital computer (cordic),” IEEE Trans. Multi-Scale Comput. Syst. 3 (2017) 139 (DOI: 10.1109/TMSCS.2017.2696003).

[38] R.G. Andrzejak, et al.: “Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state,” Phys. Rev. E 64 (2001) 061907 (DOI: 10.1103/PhysRevE.64.061907).