This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBCAS.2020.2974154, IEEE Transactions on Biomedical Circuits and Systems

A Review of Algorithm & Hardware Design for AI-Based Biomedical Applications

Ying Wei, Jun Zhou*, Yin Wang, Yinggang Liu, Qingsong Liu, Jiansheng Luo, Chao Wang, Fengbo Ren, Li Huang

***Abstract*— This paper reviews the state of the arts and trends of the AI-based biomedical processing algorithms and hardware. The algorithms and hardware for different biomedical applications such as ECG, EEG and hearing aid have been reviewed and discussed. For algorithm design, various widely used biomedical signal classification algorithms have been discussed including support vector machine (SVM), back propagation neural network (BPNN), convolutional neural networks (CNN), probabilistic neural networks (PNN), recurrent neural networks (RNN), Short-term Memory Network (LSTM), fuzzy neural network and etc. The pros and cons of the classification algorithms have been analyzed and compared in the context of application scenarios. The research trends of AI-based biomedical processing algorithms and applications are also discussed. For hardware design, various AI-based biomedical processors have been reviewed and discussed, including ECG classification processor, EEG classification processor, EMG classification processor and hearing aid processor. Various techniques on architecture and circuit level have been analyzed and compared. The research trends of the AI-based biomedical processor have also been discussed.**

***Index Terms*—AI, Biomedical Application, Algorithm, Processor**

I. Introduction

THe large-scale multi-group and multi-level biomedical data has provided us with the possibility to study the mechanism of biological operations. Compared with traditional data analysis, artificial intelligence is progressing rapidly and is exerting an increasing impact in the field of biomedical processing algorithms and hardware due to its powerful ability of model building and parallel learning.

There is a wide spectrum of AI-based biomedical applications such as intelligent ECG monitoring, EMG monitoring, blood pressure monitoring and hearing aids, where the AI-based classification algorithms are used for automated diagnosis and classification. Compared with traditional signal processing algorithms, the AI-based algorithms help the doctors enhance the intelligence and accuracy of the diagnosis or classification since it can automatically extract and analyze the signal features while being less affected by subjective factors. Many AI based classification algorithms have been proposed in the past, including support vector machine (SVM), back propagation neural network (BPNN), convolutional neural networks (CNN), probabilistic neural networks (PNN), recurrent neural networks (RNN), Long Short-term Memory Network (LSTM), fuzzy neural network and etc. Different algorithms are suitable for different applications. For example, SVM and CNN are widely used in ECG signal processing while fuzzy classifiers are popular for blood pressure monitoring.

For hardware design, dedicated biomedical processors have been proposed to accelerate the AI-based biomedical processing algorithms and achieve low power consumption. Various hardware platforms have been adopted for the development of the biomedical processors, including microcontroller, FPGA and ASIC. The machine learning algorithms with relatively low complexity such as SVM and KNN can be implemented on microcontrollers for lower design cost and effort [1]. Compared to the microcontroller, the FPGA provides better performance and is suitable for accelerating algorithms with higher complexity such as neural network computation [2][3]. ASIC has been used to achieve even higher computation speed and ultra-lower power consumptions that FPGA cannot achieve [4][5].

Most of the AI-based biomedical processors focus on ECG signal classification, epilepsy seizure detection and EMG signal classification. Some other applications include AI-based acoustic processor for hearing aid and stress detection processor for health monitoring. While the applications vary, these processors have similar system architectures, consisting of pre-processing module for filtering/denoising & segmentation, feature extraction module for feature extraction & dimension reduction, and classification module for classification & diagnosis.

For the design of the AI-based biomedical processors, a major focus is to optimize the power and area consumption while achieving high classification accuracy, as many of these processors are used in wearable/portable devices. Firstly,

This work was supported in part by National Key Research and Development Program of China (2019YFB2204500, 2019YFB1311302), National Science Foundation (61671277), Shandong Key Research and Development Program (2018JHZ007, 2018GGX101051). Jun Zhou is the corresponding author.

Ying Wei and Yinggang Liu are with Shandong University, Jinan, China. (e-mail: eleweiy@sdu.edu.cn).

Jun Zhou, Yin Wang, Qingsong Liu and Jiansheng Luo are with University of Electronic Science and Technology of China (e-mail: zhouj@uestc.edu.cn).

Chao Wang is with Huazhong University of Science and Technology (e-mail: chao_wang_me@hust.edu.cn).

Fengbo Ren is with Arizona State University (e-mail: renfengbo@asu.edu).

Li Huang is with imec Netherlands (e-mail: li.huang@imec-nl.nl)

algorithm-level exploration and optimization are used to reduce the computational complexity while maintaining classification accuracy, such as sparse neural network and spiking neural network. Second, architecture-level design techniques are used to reduce the power consumption, such as weak-strong hybrid classification architecture, sparse matrix computing architecture and serial processing architecture. Thirdly, circuit-level design techniques are adopted to further reduce the power consumption, such as near-threshold circuit design, switchable clocking circuit design, and analog computing circuit design. In addition, area reduction techniques have been proposed to lower the cost such as structured compression of neural network, low-precision representation of weights and reconfigurable hardware design.

II. AI-BASED BIOMEDICAL PROCESSING ALGORITHM DESIGN

A. *AI-based ECG Processing Algorithm Design*

In the last decade, AI-based Electrocardiogram (ECG) applications have achieved fast development. The two most popular approaches are SVM and artificial neural network (ANN).

SVM is a powerful classifier that categories data using supervised learning. In [6] SVM was used to automatically identify obstructive sleep apnea syndrome (OSAS) types from night-time ECG recordings. The heart rate variability (HRV) and ECG-derived respiration (EDR) are adopted as features to train the SVM and achieving the OSAS+ recognition accuracy rate of 92%. HRV is also adopted as a feature in [7] for driver drowsiness. Receiver operation curve (ROC) analysis and SVM classifiers were used for feature selection and classification. More complicated features are used in [8]. Fourteen metrics in the ECG were used in the genetic algorithm to select the best combinations of variables, and then these variables were entered into the SVMs for ventricular fibrillation (VF) and rapid ventricular tachycardia (VT) detection. More different SVM models were developed in [9]. Both the subject-independent model and the subject-dependent model were used to monitor real-time sleep apnea based on ECG recordings.

To improve the performance, other techniques are combined with SVM. In [10], Least Square-Support Vector Machine (LS-SVM) and feed forward neural network (NN) were used for the classification of different types of ECG beats. They compared three different feature extraction methods and found that the highest accuracy was obtained when the principal component of segmented ECG beats was used. In [11], the classification was done based on particle swarm optimization (PSO) and SVM. Experiment results verified that PSO-SVM has improved classification accuracy compared with the previous work. PSO-SVM structure were also used in [12]. It proposed a system with PSO-based wavelet filter bank followed by SVM classifier, which obtained satisfactory accuracy and stability. In general, the combined algorithms have better performance than using SVM directly.

The artificial neural networks (ANN) used in ECG applications include many different types, such as convolutional neural networks (CNN), probabilistic neural networks (PNN), recurrent neural networks (RNN) and back propagation neural network (BPNN). An individual CNN was trained for each patient's ECG in [13]. It is verified by experiments that this system has superior performance in detecting ventricular ectopic beats and supraventricular ectopic beats. In [14-17], CNNs are used to detect atrial fibrillation (A(fib)), atrial flutter (A(fl)), ventricular fibrillation (V-fib), coronary artery disease (CAD), myocardial infarction (MI) and five main classes of non-life-threatening arrhythmias. The same 11-layer deep CNN structure was used. In [18], a 9-layer deep CNN structure was used. The noise of the ECG signal is removed and then fed into the CNN network. The work of [14-18] validated the reliability of using different training sets to detect multiple diseases in the same CNN framework.

PNN classifiers are also commonly used in ECG applications. In [19], a PNN model was proposed to distinguish the normal sinus rhythm (NSR) and seven types of arrhythmias. Fuzzy C-mean (FCM) clustering was used for feature reduction. Experiments results show that the proposed model obtained good classification accuracy. PNN can also be used together with feature reduction algorithms [20]. The performances of three different feature reduction methods combined with PNN classifiers were compared. It is shown that Principal component analysis (PCA) + linear discriminant analysis (LDA) + PNN is the best solution.

RNNs was applied in [21] and [22] to classify normal beat, congestive heart failure beat, ventricular tachyarrhythmia beat and atrial fibrillation beat. In [21], multiple signal classification (MUSIC) and Minimum-Norm were used to estimate the power spectral density (PSD) which is fed into the RNN network. In [22], maximum, minimum, mean and standard deviation of the Lyapunov exponents of each ECG beat are the features fed into the RNN network. The results of applications showed that PSD estimates and Lyapunov exponent are good representations of ECG signals. RNNs are good at processing time series. However, the conventional RNN has the problem of long-term dependency. The subsequent RNN variant, Long Short-term Memory Network (LSTM), can solve this problem well, so it has become the attractive choice for ECG signal processing. In [23], a bi-directional LSTM model was proposed to classify ECG signals. A wavelet-based layer was used to produce the ECG sequence which were then divided into different subbands before they were sent to the LSTM network. Experimental results showed that the recognition accuracy is as high as 99.39%. In [24], a new two-stage LSTM structure was proposed for anomaly detection of the ECG signals. The-second stage predictor was applied to further process the error outputs of the first stage network to find the true anomalies, which can enhance the detection accuracy.

BPNN is also used in ECG applications. In [25], BPNN classifier was applied to automatic ECG arrhythmia classification. Dual tree complex wavelet transform (DTCWT)-based feature extraction technology was used to extract feature sets. In [26], the genetic algorithm was used to decrease the dimension of the feature set and optimized the weight and biases of the BPNN. The statistical characteristics of the wavelet packet coefficients were used as feature sets.

The main methods and the corresponding performance evaluation parameters are summarized in Table I for comparison. In addition, we investigated the proportional relations of several commonly used ANNs in the past ten years (from 2010 to 2019) using Web of Science. The scale map of various network used is shown in Fig. 1. It can be seen that among these networks, CNN is the most popular one. The other three networks, PNN, RNN and BPNN are used in about the same proportion. The reason that CNN, which performs well in the image field, can be successfully applied to ECG signals could be that ECG signals also have the characteristics of local correlation and translation invariance just as some image signals.

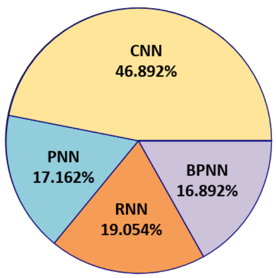


Fig. 1 The proportional relations of several commonly used AI models for ECG processing in the past ten years (2010 to 2019).

B. AI-based EEG Processing Algorithm Design

AI is widely used in EEG signal processing such as the discovery of epilepsy and various psychiatric diseases. Compared with these areas which have been concerned for a long time, emotional recognition based on EEG has received widespread attention. [27]. Compared with pulse, blood pressure and other physiological signals, as a direct central nervous response, EEG has real-time differences and thus associated with emotions far beyond other signals [28]. In the research of emotion recognition based on EEG signals, many experts and scholars have obtained remarkable academic achievements.

Experts and scholars have combined psychology, physics and neurology to find EEG features that can be used to identify emotions. Earlier, the feature extraction methods using Fourier transform, wavelet transformation and other methods to calculate the time or frequency domain features, such as standard deviation, power spectrum, entropy etc. [29-32]. Recently, deep learning network (DLN) is used to discover unknown feature correlations between input signals [33]. The DLN is implemented by a stacked automatic encoder (SAE) using a hierarchical feature learning method and achieved a good classification accuracy.

Choosing a suitable classification algorithm is also significant in the study. Both supervised learning and unsupervised learning are used for the emotional classification. Unsupervised learning methods commonly used by researchers include fuzzy clustering, K-proximity algorithms, etc. Supervised learning methods generally include linear decision analysis, SVM, neural networks, and Gaussian Bayesian network. Fuzzy C-means and fuzzy K-means algorithms were proposed to cluster three emotions in [34]. The clustering results were observed experimentally to help select the feature set that is most conducive to emotion recognition. In [35], Bayesian linear discriminant was used to classify six kinds of emotions (happy, angry, disgusted, sad, surprised, and scared). The accuracy rate is over 80%. More classifiers, including quadratic discriminant analysis and support vector machine, were used and compared in [36]. The classification accuracy of using quadratic discriminant analysis and support vector machine reached 62.3% and 83.33%, respectively. In addition, in [37] the effectiveness of second-order discrimination, K-nearest neighbors and support vector machine were compared. The emotions were divided into surprises and anger. The highest average recognition rate is 83.33% when using support vector machine. In [38] the isometric feature mapping algorithm (Isomap) in manifold learning was used to predict the change of emotion. The emotional curve predicted by the isometric feature map and the collected emotional curve have very good fit.

Using one machine learning method is to find a classifier which is closest to the actual classification function. However, the performance and stability of a single classifier are usually affected by the data when building the model. EEG signal is a time-varying non-stationary random signal with strong background noise. Thus only using one classifier to perform EEG signal processing may lead to unstable classification results. One possible solution is to use ensemble learning to get a better and more comprehensive strong supervision model. For example, in [39], a very effective classifier in ensemble learning, random forest (RF), was used for emotion recognition based on EEG. The EEG features are hybrid, which is from time, frequency and wavelet domains respectively. The RF method obtained an accuracy of 75.6%, which outperforms the accuracy of 69.8% using SVM method and the accuracy 60.4 % using LDA classifier. The summary of recent emotion recognition algorithms based on EEG signals is shown in Table II.

The diagnosis of epilepsy by analyzing EEG signals has also become a hot topic. The intelligent diagnosis technology plays a more and more important role for improving the detection efficiency and reducing misjudgment rate [40]. The support vector machine algorithm (SVM) were most widely used in epilepsy detection [41-43]. In [41] and [42], wavelet analysis was applied to decompose EEG signals. The characteristics such as energy, entropy, volatility index, etc. were then calculated and input into the SVM classifiers. The classification accuracy of detecting epilepsy from normal EEG signals was 91.2% in [41]. In [42], the performance was enhanced by post-processing such as smoothing and multi-channel decision fusion. The sensitivity and specificity were 94.46% and 95.26%, respectively. In [43], combined wavelet transform with high order moments was used to extract the time-frequency domain joint distribution information that is the input of the SVM classifier. The classification accuracy was 94.5%. In [44] and

[45], least squares support vector machine (LS-SVM) was used to minimize errors and maximize classification performance. In [44], a time-dependent signal decomposition method was proposed. Wavelet analysis and fractal method were used to extract signal features, which are then sent to LS-SVM classifier. The classification accuracy rate was 96.5%. In [45], the statistical characteristics such as standard deviation, skewness coefficient and kurtosis coefficient were passed to the LS-SVM classifier for nonlinear kernel function for classification. The final classification accuracy rate reached 97.19%. Beside SVM technique, artificial neural network was also used for epilepsy detection. In [46], a two-layer feed-forward neural network was used to perform classification. EMD was applied to obtain high-order eigenmode functions to extracted linear and nonlinear features such as energy and related dimensions of EEG signals. The classification accuracy rate reached 96%. The summary of recent epilepsy detection algorithms based on EEG signals is also shown in Table II.

C. AI-based Blood Pressure Processing Algorithm Design

Many AI algorithms have been used for blood pressure monitoring. Among them, fuzzy classifiers are widely used. In [47], a Mamdani-type fuzzy model with 21 rules is established. The blood pressure level classification is performed based on the expert knowledge represented in the fuzzy rules. The main contribution of this work is the design of optimal interval type-2 fuzzy systems with trigonometric membership functions. The Mamdani fuzzy inference system was also used in [48]. The fuzzy system was built in accordance with the European Hyertensin Guidelines besides the expert's knowledge. Both the systolic and diastolic blood pressures were used as inputs. Experimental results illustrated the effectiveness of this system. In [49], a Neural Fuzzy Hybrid Model (NFHM) was applied to classify blood pressure based on rules provided by experts. GA is used to obtain the optimal number of rules for classifiers with the lowest classification error. The fuzzy expert system is also used for hypertension diagnosis [50][51]. Different with [49], the expert systems in [50] and [51] do not give a classification for blood pressure but give the evaluation of the risk.

Other kinds of AI models are also applied in blood pressure control. A multilayer feed forward model was used in [52] to predict blood pressure. The results show that the model can simulate multiple heterogeneous results at the same time, and is more suitable for predicting the complex interactions and nonlinear effects between variables and ideal results. In [53], two neural network models, BP and RBF network were used to predict blood pressure based on body mass index (BMI), age, motion level, etc. In [54], a Deep Belief Network (DBN)-Deep Neural Network (DNN) was designed to understand the complex nonlinear relationship between artificial eigenvectors obtained with DBN-DNN-based oscilloscopes and reference blood pressure. Compared with the traditional methods, this model provides lower error standard deviation, average error and average absolute error for supine blood pressure (SBP) and diastolic blood pressure (DBP). In [55], a CNN model is built to recognize Korotkoff sounds one by one. Then, a mapping algorithm is developed to associate the identified Korotkoff beats with the corresponding cuff pressure for SBP and DBP measurements.

The methods and performance of recent blood pressure algorithms was listed in table III for a better overview. In addition, we investigated the proportional relations of several commonly used AI models in the past ten years (from 2010 to 2019) using Web of Science. The scale map of various network used is shown in Fig.2. It can be seen that among these networks, fuzzy classifiers are most widely used. The other networks commonly used are CNN, RBF and BPNN. The reason fuzzy classifiers were attractive is that they can take a better use of the experts' knowledge compared with other AI models, which is extremely important in blood pressure control.

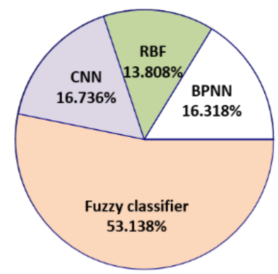


Fig. 2The proportional relations of several commonly used AI models for blood pressure control (2010 to 2019).

D. AI-based Hearing Aid Processing Algorithm Design

In hearing aid application, AI is widely used in speech enhancement, multi-talker separation and dereverberation. According to the number of microphones, there are two classes of approach, monaural-microphone-based approach and array-based approach [56].

Much study has been invested into monaural separation approach. A DNN-SVM with ideal binary mask (IBM) as target was proposed in [57]. In first stage, the raw features were fed into feedforward neural network to learn new features which contain the relationship between raw features. Then, raw features and learned features were fed into SVMs to estimate IBM. In [58], back-propagation DNN was used. The LPS log power spectrum (LPS) features of noisy speech and clean speech were used as the input and the training target respectively to train the network. In [59], both a positive DNN and a negative DNN were trained with mixture signals that have positive and negative SNRs, respectively. In the separation stage, a general DNN was firstly used to estimate the SNR of the mixture signal. Then, the positive or negative DNN is selected for separation according to SNR. The experimental results show that the proposed framework outperforms the conventional DNN. In [60], a framework with two-stage DNNs was introduced. The denoising stage is followed by the dereverberation stage. Each stage has their own DNN. The two DNNs firstly trained independently to generate initial weights of DNNs. Then, the two DNNs trained jointly. The output of dereverberation DNN was used for signal reconstruction. In [61], two multi-DNN systems for speaker separation were proposed. Each DNN was trained to estimate an Ideal Ratio Mask (IRM) or signal approximation (SA) from a mixture speech. The output of the first network is an average of the outputs of the DNNs. The second network has multi-layers. Each layer includes one or serval DNNs. The outputs of DNNs in lower layer are used as the inputs of DNNs in higher layer. The output of the network is the output of DNN in the last layer.

To obtain a good generalization in different noisy environments, in [62], a large-scale training was carried out on DNN. The training set consists of 10000 noise signals. The results suggest that the obtained DNN has better generalization performance under different signal-to-noise ratios. Although the generalization to noises can be improved through apply large-scale training to a DNN, it is proved in [63], that DNN cannot gain a good generation to speakers even through a largescale training of speakers. In [63], long short-term memory (LSTM) network was adopted. Because LSTM block has the ability to memory the previous information, it captures long-term speech contexts to trace a target speaker. The proposed module shows effective generalization to speaker and noise.

Compared to monaural microphone, binaural microphones can provide spatial information which is benefit to sound sources separation. DNN model is also the most popular model used for binaural processing. Compared with monaural case, the binaural features are added to the inputs of the network. In [64], interaural time difference (ITD) and interaural level difference (ILD) were used. In [65], Log-power spectra (LPS) (monaural feature) and ILD (binaural feature) were extracted. The LPS feature of the left-ear is selected as the training label. The speech can be reconstructed through the estimated LPS. In [66], the signals from left-ear and right-ear are preprocessed by two modules. In the first module, one enhanced signal is obtained from the two signals through beamforming before monaural feature extraction. In the second module, ILD and ITD are obtained. Because the DNN is trained to estimate ratio mask, IRM is selected as the training target.

Microphone arrays are widely used to speech enhancement through beamforming. In [67-69], conventional beamforming techniques are combined with neural networks. In [67], the Minimum Variance Distortionless Response (MVDR) was improved. A LSTM network was used to enhance each microphone signal and get the corresponding initial noise and speech components. Then, both of them were used to calculate spatial covariance matrices by a predicted mask. The spatial covariance matrices and the initial signals are finally used for MVDR beamformer. In [68], a bi-directional Long Short-Term Memory (BLSTM) network was used to estimate IBMs for noise and for target speech from each channel. All the masks were condensed to get a single speech and a single noise mask through median operation, based on which the Cross-Power Spectral Density (PSD) matrices of speech and noise can be obtained. The PSD matrices are used for beamforming. The experiment results show that the performance of Generalized Eigenvalue (GEV) beamformer is better than that of the MVDR beamformer. In [69], a DNN module was used to estimate the IRM mask from was selected as the initial mask to calculate target speech covariance matrix. Using the matrix, the steering vector is obtained to be used for the MVDR beamforming.

The typical methods and corresponding performance of recent Hearing aid algorithms was summarized in Table IV. In addition, we investigated the work in this area from 2010 to now at web of knowledge, as shown in Fig. 3. Compared with multi-talker separation and dereverberation, the study of speech enhancement based on AI-model increases fast during these years. When considering the number of microphones, we found that there is not much difference in proportions between using a single microphone and using multiple-microphones, as shown in Fig. 4.

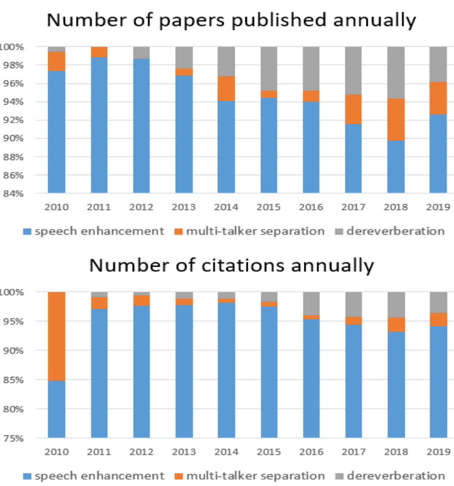


Fig.3 Papers published and cited annually (from 2010 to 2019).

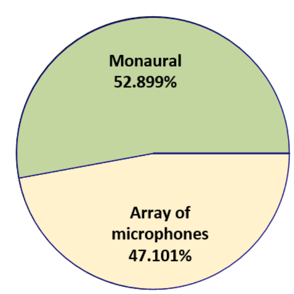


Fig. 4 The ratio of using a single microphone to use a multiple microphone for hearing aid (2010 to 2019).

E. AI-based Electromyography Processing Algorithm Design

Pattern recognition and classification of EMG signals can be used to diagnose neuromuscular diseases. In [70], a PSO-SVM method based on the combination of particle swarm optimization (PSO) and SVM was proposed. PSO was used to optimize the parameters of SVM. This method has good generalization ability and high sparsity. The classification accuracy was 97.41% on 1200 EMG signals. In [71], an EMG signal classification framework based on multi-scale principal component analysis for de-noising, discrete wavelet transform (DWT) for feature extraction, and decision tree (DT) for classification was proposed. The combination of DWT and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBCAS.2020.2974154, IEEE Transactions on Biomedical Circuits and Systems

random forest (RF) classifier obtained an accuracy of 96.67%. In [72], an EMG classification scheme based on multiple one-versus-one classifiers was proposed, which used the dimensionality reduction method of uncorrelated linear discriminant analysis (LDA). The results showed that this method is better than the other nine popular classifiers. In [73], several representative classifiers, including linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), k-nearest neighbor (KNN), decision tree (DT), Naive Bayes (NB) were used to verify the recognition accuracy. The signal was processed by differential transformation, which led to 2-8% increasing in accuracy.

In addition to various classification algorithms in disease diagnosis, there are many specific applications of EMG pattern analysis, such as gesture analysis, prosthesis control, etc. In [74], a twin SVM model was proposed to implement gesture classification. Twin SVM generates two separate hyperplanes and does not require each class to have a similar distribution. Thus, different kernels can be used for different classes. The method is effective and can be extended to other similar biomedical applications. In [75], Kernel Regularized Least Squares (KRLS) algorithm was applied to the control of the prosthesis. The results showed that the kernel was better than the radial basis function kernel in various situations. In addition, combining the complementary electromyogram and accelerometry in the multimodal classifier can significantly improve the accuracy. In [76], an efficient four-channel EMG pattern recognition technique based on decision tree (DT) was proposed for the control of prosthetic hands. Its processing time and storage space are small. The recognition accuracy can reach 91+1.9%. In [77], different machine learning techniques including the least squares kernel (LSK) and linear discriminant analysis (LDA) was used to determine the muscle changes of arthritis patients. It was shown that LSK algorithm had the higher classification ability than LDA in this specific application. For a clear overview, the performance of different algorithms was shown in Table V.

F. *Other AI-based Biomedical Processing Algorithm Design*

Beside the above applications, the use of AI technology in biomedical and healthcare field is more and more extensive, especially in the area of various medical auxiliary equipment and medical image processing. By combining deep learning, artificial neural network (ANN), Reinforcement learning (RL), logistic regression and other machine learning methods, AI-based applications have been implemented and achieved good results.

AI-based medical diagnostic or detection systems can effectively promote the development of medical service, and improve the quality of medical treatment. Recent work includes detecting diabetic retinopathy [78], Laparoscopy [79], grading OCT images [80], monitoring blood glucose [81], classifying benign and malignant cervical lymph nodes [82], arrhythmia screening [83] etc.

Medical image segmentation and fusion is another area in which AI models are widely used. The advantages of AI-based technology are reflected in accuracy and convergence time compared with the non-AI-based technology. The popular AI models used in this area include CNN, RNN, deconvolution networks (DeCN), Generative Adversarial Networks (GAN), simple fusion network (SFNet) [84], fuzzy C-means (FCM) [85], deformable U-Net (DUNet) [86] etc.

Table I Summary of ECG processing algorithms

content	Ref	Method	Accuracy	Sensitivity	Specificity:
Apnea Episodes	[6]	SVM	92.85%	–	–
	[9]	SVM	90% (F-measure)	96%	–
Driver Drowsiness	[7]	ROC+ SVM	95%	95%	95%
Different types of ventricular arrhythmia：Normal beat (N)，Premature Ventricular Contraction (PVC), Paced beat (PACE), Right Bundle Branch Block beat (RBBB), Left Bundle Branch Block beat (LBBB), Atrial Premature Contraction (APC), Ventricular flutter wave (VLWAV), and Ventricular Escape beat (VESC).	[8]	SVM	96.3%±3.4%	96.2%±2.7%	96.2%±4.6%
	[10]	LS-SVM	98.11%	99.90%	99.10%
	[11]	PSO-SVM	89.72%	–	–
	[12]	PSO-SVM	89.18%	–	–
	[14]	CNN	94.90%	99.13%	81.44%
	[16]	CNN	93.18%	95.32%	91.04%
	[18]	CNN	89.07%	–	–
	[19]	PNN	99.58%	–	–
	[20]	PCA-LDA-PNN	99.71%	–	–
	[21]	RNN	98.06%	–	–
	[22]	RNN	94.72%	–	–
	[23]	LSTM	99.39%.	–	–
	[25]	BPNN	94.64%	–	–
	[26]	GA-BPNN	99.33%	–	–
Myocardial Infarction	[15]	CNN	94.95%	93.72%	95.18%
	[17]	CNN	93.53%	–	–

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBCAS.2020.2974154, IEEE Transactions on Biomedical Circuits and Systems

Table II Summary of EEG processing algorithms

Ref.	Method	Content	Recognition accuracy
	Emotional recognition		
[29]	Support vector machines (SVM)	Four (joy, relax, sad, and fear)	65.1%
[30]	Linear discriminant analysis (LDA)	Three (anger, fear, and surprise)	66.3%
[31]	Support vector machines (SVM)	Two (positive and negative)	84.22%
	k-nearest neighbors (KNN)		76.56%
[32]	Deep Neural Networks (DNN)	Three (positive, neutral and negative)	86.08%
[33]	Deep learning network (DLN)	Three different levels of valence and arousal	Valence:49.52%+5.55% Arousal:46.03%+6.53%
[34]	Fuzzy C-Means (FCM) clustering	Four (happy, disgust, surprise and fear)	97-99%
[35]	Bayes linear discrimination (BLD)	Six (Happy, angry, sad, surprised, afraid and disgusted)	80%
[36]	Quadratic discriminant analysis (QDA)	Six (happiness, anger, fear, disgust, surprise, and sadness)	62.3%
	support vector machines (SVM)		83.33%
[37]	Support vector machine (SVM), multi-layer perceptron (MLP) and k-nearest neighbors (KNN)	Two (class-1: performing a complex cognitive task; class-2: eyes open condition)	95-99%
[38]	Support vector machines (SVM)	Six (joy, anger, fear, disgust, relaxation, and sad)	71.38%-78.41
[39]	Random forest (RF)	Four (happy, sad, angry and relaxed)	75.6%
	support vector machines (SVM)		69.8%
	linear discriminant analysis (LDA)		60.4%
	Epilepsy detection		
[40]	Feed forward artificial neural networks (ANN)	healthy, interictal, seizure, epileptogenic and hippocampus	99.28%
[41]	Support vector machines (SVM)	normal and epileptic	91.2%
[42]	Support vector machines (SVM)	normal and epileptic	94.46% (sensitivity) 95.26% (specificity)
[43]	Support vector machines (SVM)	normal and epileptic	94.5%
[44]	Least square support vector machine (LS-SVM)	ictal and interictal	96.5%
[45]	Least square support vector machine (LS-SVM)	epileptogenic, hippocampus, episodes	97.17%
[46]	Feed forward artificial neural networks (ANN)	healthy, interictal, seizure, epileptogenic and hippocampus	94%

Table III Summary of blood pressure processing algorithms

Ref.	Method	Content	Classification accuracy	Risk	Standard deviation
[47]	Fuzzy classifier	blood pressure level classification	99.11%	–	–
[48]	Fuzzy classifier	blood pressure and hypertension risk diagnosis	Modular 1 98% Modular 2 97.62% Modular 3 97.83%	1.03 0.89	–
[49]	Neural Fuzzy Hybrid Model	blood pressure diagnosis	66.7%	_	_
[50]	Fuzzy expert system.	Hypertension diagnosis	–	0.461	–
[51]	Support Vector Machine Recursive Feature Elimination (SVM-RFE)	gene expressions for hypertension diagnosis	_	_	6.302%
[52]	Multilayer feed-forward neural network (ANN)	predict Kt/V, fluid volume removal, heart rate, and BP to management of Blood Pressure	_	_	13.3%
[54]	Deep belief network (DBN)	Oscillometric Blood Pressure Estimation	_	_	7.07%
[55]	CNN	Supine blood pressure (SBP) and diastolic blood pressure (DBP) determinations.	_	_	SBP:5mmHg-DBP:2mmHg -9mmHg

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBCAS.2020.2974154, IEEE Transactions on Biomedical Circuits and Systems

Table IV Summary of hearing aid processing algorithms

The number of microphone	Object	Ref.	The adopted techniques	HIT-FA with different input SNR			PESQ with different input SNR			STOI with different input SNR			SNR with different input SNR /dB			LSD with different input SNR/dB			WER
				SNR 5	SNR0	SNR-5	SNR 5	SNR 0	SNR -5	SNR5	SNR 0	SNR -5	SNR 5	SNR 0	SNR -5	SNR 5	SNR 0	SNR -5	
Monaural microphone	Speech enhancement	[57]	SVMs, DNNs, IBM	–	68.9± 2.0%	64.9± 0.9%	–			–			–	10.5	–	–			–
	Speech enhancement	[58]	DNNs	–			2.78	2.41	1.97	–			2.7	1.7	0.0	3.6	4.5	5.7	–
	Speech enhancement	[59]	DNNs	–			2.46	2.10	1.74	–			–			5.87	8.23	11.0	–
	Speech enhancement, dereverberation	[60]	DNNs, IRM	–			2.76	2.57	2.25	90.0 %	87.6 %	83.3 %	–			–			–
	Speaker separation	[61]	DNNs, IRM, Signal Approximati on (SA)	–			–			–	91.1 %	–	–			–			–
	Speech enhancement	[62]	DNN, IRM, large-scale training	–			–			89.9 ± 0.6%	82.9 ± 0.5%	70.8 ± 0.1%	–			–			–
	Speaker separation	[63]	LSTM	–			–			92.0 ± 1.0%	86.5+ 1.5%	77.5 ± 2.5%	–						
Bionaurual microphones	Speech enhancement, dereverberation	[64]	DNN, IBM	86.27 %	86.54 %	84.82 %	–			–			16.6 8	15.0 8	11.8 6	–			–
	Speech enhancement	[65]	Regression DNN	–			–	2.46	–	–	83.3 %	–	–			–			–
	Speaker separation, dereverberation	[66]	DNN, IRM, Beamformin g	–			–			–	–	74.7 %	–	–	4.5	–			–
Microphone arrays	Speech enhancement	[67]	Beamformin g, LSTM	–			1.70±0.16(SNR not given)			–			–			–			29.04± 7.84%
	Speech enhancement	[68]	Bi-directional LSTM, IBM, Beamformin g	–			2.65(SNR not given)			–			–			–			15.42%
	Speech enhancement	[69]	DNN, IRM, Beamformin g	–			–			–			–			–			5.05%

Table V Summary of the EMG processing algorithms

Ref.	Method	content	Accuracy	Sensitivity	Specificity
[70]	PSO-SVM	1200 EMG signals selected from 27 subject records	97.41%	-	-
[71]	Random forest	EMG signals from25 patients (10 controls, 7 myopathy and 8 amyotrophic lateralizing sclerosis)	96.67%	-	-
[73]	linear discriminant analysis (LDA)	Eight motions (forearm pronation (FP), forearm supination (FS), wrist extension (WE), wrist flexion (WF), wrist radial deviation (WRD), wrist ulnar deviation (WUD), hand open (HO), and hand close (HC))	83.33% (6.4%)	-	-
	quadratic discriminant analysis (QDA)		90.12% (5.9%)	-	-
	k-nearest neighbor (KNN)		89.60% (5.5%)	-	-
	decision tree (DT)		84.07% (6.6%)	-	-
	Naive Bayes (NB)		77.31% (8.5%)	-	-
[74]	Twin SVM	7 gestures (wrist flexion, finger flexion, wrist flexion toward little finger, wrist flexion toward thumb, fingers and wrist flexion together, fingers and wrist flexion toward little finger, fingers and wrist flexion toward thumb.)	86.0%	84.8%	88.1%
[75]	Kernel Regularized Least Squares (KRLS)	The second version of the NinaPro database	82.49%	-	-
[76]	Decision tree	six specifics hand and wrist motions	91+1.9%	-	-
[77]	least-squares kernel (LSK)	EMG recordings are collected from healthy subjects (CO), 8 rheumatoid arthritis (RA) and 10 hip osteoarthritis (OA)	91%	90%	90%

III. AI-Based Biomedical Processor Design

A. AI-based ECG Classification Processor Design

ECG sensors have been widely used for cardiac health monitoring. The traditional ECG sensors only record the ECG raw data and transmit it out for diagnosis. This causes large power consumption and delay for the data transmission. In the recent years, smart ECG sensors have been developed to perform local ECG diagnosis to avoid the transmission of raw data. This requires integration of ECG classification processor in the ECG sensor. These ECG classification processors implements AI-based algorithms in hardware to achieve high classification accuracy with real-time performance. Compared to normal ECG processing such as denoising, QRS detection and compression, ECG classification has much higher computational complexity and memory requirement. Therefore, reduction of power consumption and hardware overhead have become important aspects for the ECG classification processors.

In the past, SVM-based ECG classifications processor been investigated due to its high accuracy and simplicity for implementation. In [87] a low power machine learning assisted ECG processor for mobile health applications is proposed. As shown in Fig. 5, a 2-stage processing engine is designed to extract the features of the ECG signal, where the DWT is used to identify the QRS/P/T waves for heartbeat segmentation and a multivariate autoregressive (MAR) estimator is used for analyzing the characteristics of each heartbeat. Following the feature extraction engine, a switchable classification engine is designed which can switch between SVM and maximum likelihood classification (MLC) to support different applications by reusing the computation hardware, as shown in Fig. 6. To reduce the dynamic power consumption, near-threshold circuit design has been adopted. The standard cell library was characterized at a near-threshold voltage (i.e. 0.5V) and the cells with functional failures were removed. Level shifter cell is designed to bridge between the ultra-low voltage region and the nominal voltage region. A switchable clocking scheme is applied to further reduce the power consumption, where a KHz-range crystal-less clock generator is used for the data sampling phase, and a MHz-range oscillators using critical-path replica is used for the computation phase. To reduce the leakage power consumption, a dynamic standby controller is

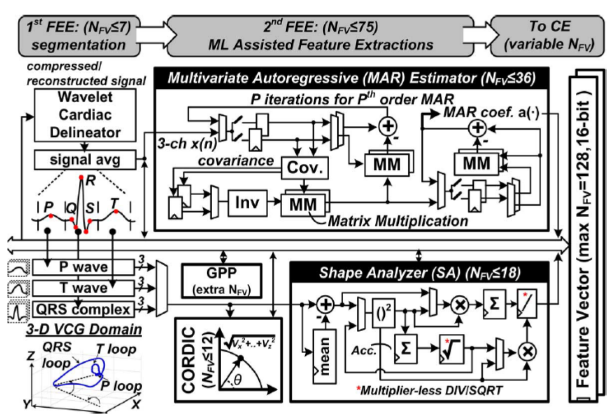


Fig. 5 The architecture of the two-stage processing engine for feature extraction. [87]

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBCAS.2020.2974154, IEEE Transactions on Biomedical Circuits and Systems

designed to minimize the leakage current in the sleep mode and enable fast wakeup.

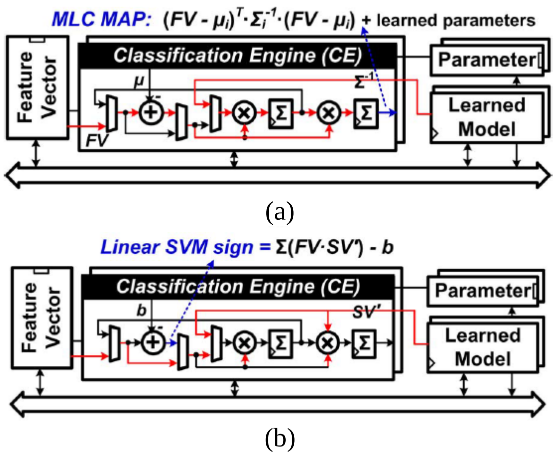


Fig. 6 Switchable classification engine with computation hardware reuse (a) MLC classification mode (b) SVM classification mode. [87]

In [88] a low power ECG processor with weak-strong hybrid classifier for arrhythmia detection is proposed. The hardware architecture of the design is shown in Fig. 7. The design includes a weak linear classifier (WLC) and a strong SVM classifier. The WLC is used to detect the beats with distinct feature by performing threshold comparisons based on beat interval and QRS area ratio information. The heart beats that are not detected by WLC will activate the SVM classifier which is more powerful but energy consuming. In this way, the power consumption of classification is greatly reduced while the accuracy is maintained. Before SVM, principal component analysis (PCA) is used for feature dimension reduction to lower the complexity of the subsequent SVM classifier. Further, a sparse matrix computing architecture is proposed to reduce the computational complexity of PCA, as shown in Fig. 8. A sparse region is identified by comparing the contribution of different features. When the coefficient in sparse region is lower than half of a preset threshold, it is forced to zero. When the coefficient is between half of the threshold and the threshold, the coefficient is rounded to the most significant bit and the rest bits are forced to zeros. This greatly reduce the computational complexity of PCA and thus power consumption.

Compared to SVM, neural network is promising to achieve higher classification accuracy at the price of higher computational complexity which has to be addressed in the processor design. In [89] a FPGA-based ECG anomaly detection processor is proposed. In this design, PCA is used for feature reduction followed by a neural network based classifier to achieve high accuracy. Linear approximation for activation functions is used to reduce hardware resources. This design achieves an accuracy of 99.82% on the MIT-BIH database.

In [90-91] a low power ECG classification processor to classify three types of abnormal heart beats is proposed. This design utilizes a sparse neural associative memory (SNAM) derived from low connectivity neural network Willshaw-

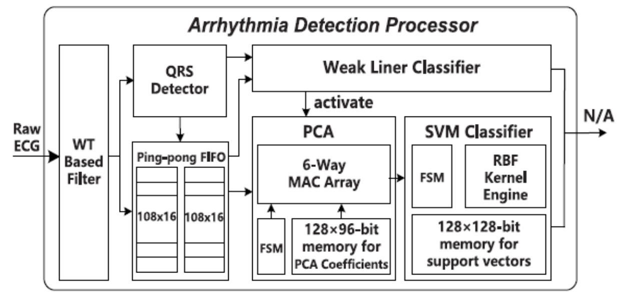


Fig. 7 ECG processor with weak-strong hybrid classifier for arrhythmia detection. [88]

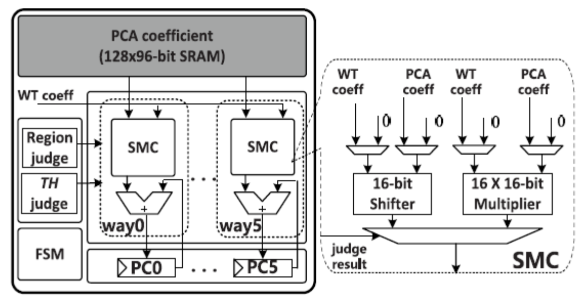


Fig. 8 Sparse matrix computing architecture for reducing the computational complexity of PCA. [88]

Palm for classification to reduce the power consumption. It behaves as an auto-associative memory which can retrieve the stored information from its noisy or incomplete version. As shown in Fig. 9, the SNAM consists of a group of partially interconnected nodes (i.e. neurons). Each node has multiple inputs and one output connecting to the input of several other nodes. The design is implemented using a winner-takes-all activation rule, which allows for the implementation of both SNAM and the encoded neural

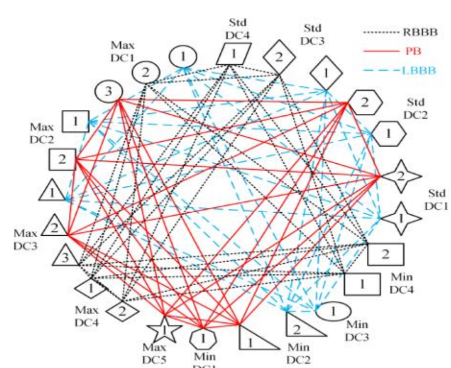


Fig. 9 Implemented low connectivity neural network for ECG classification. [90]

networks. The connections between nodes are reconfigurable

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBCAS.2020.2974154, IEEE Transactions on Biomedical Circuits and Systems

through an external memory for architecture flexibility. The design achieves a detection accuracy of 93.6% while consuming only 234 pJ per classification.

Other than diagnosis, the ECG classification processors can also be used for authentication applications. In [92] a low power ECG classification system for the biometric authentication is presented. The motivation behind the work is that as wearable devices become more and more popular, the ECG signal together with other biometric markers such as fingerprint is a promising approach for authentication. In this design, a streamlined method is proposed to utilize neural network to detect the QRS complex of the ECG signal as well as perform user authentication. Tested on in total 90 individuals, the developed system is able to achieve an equal error rate of only 0.0582% for user identification. The design is implemented on a FPGA and consume 31.75 mW when running at 12.5 MHz for a user authentication application.

In [93] an ECG processor for both cardiac monitoring and biometric authentication is proposed, as shown in Fig. 10. The hardware architecture is shown in Fig. 11. The monitoring and authentication share several modules including finite-impulse-response (FIR) filtering, outlier detection/removal, normalization, R-peak, detection, NN-based feature extraction, and cosine similarity evaluation. In the cardiac monitoring mode, R-peak detection and outlier detection modules are used to detect irregular ECG rhythm and abnormal ECG shape, respectively. During the design, a neural network based ECG feature extraction engine is design and Lasso regression is utilized to sparsity weight matrices to reduce the power and area consumption of the design. The ECG processor consumes only 1.06 μW at 0.55 V for ECG authentication with equal error rates of 1.70%.

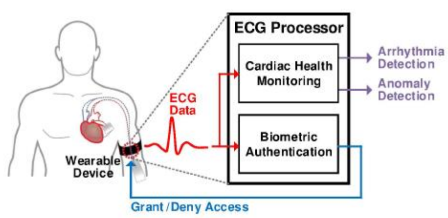


Fig. 10 Integrated ECG processor for both cardiac monitoring and biometric authentication. [93]

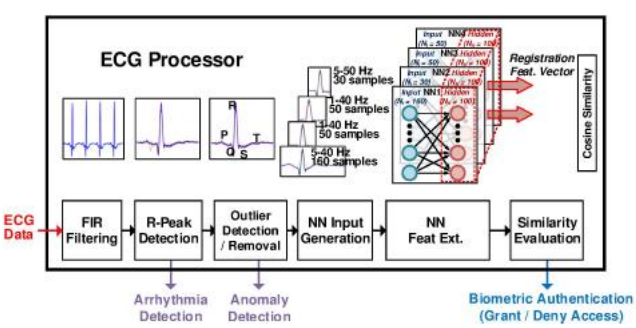


Fig. 11 Computation flow of the ECG processor. [93]

An accurate and low power ECG processor is designed for authentication and is presented in [94]. In this design, a new cost function is proposed to maximizes inter-individual distance and minimizes intra-individual distance over time so as to reduce the error rate. Structured compression with low-precision representation of weights are adopted to reduce the memory size required by the neural network computation, as shown in Fig. 12. The neural network trained with coarse-grain sparsity technique is implemented using a pipelined datapath to achieve real-time performance. The design consumes 59.4 μW at 1.2 V and the equal error rate is 1% for an in-house large database containing 741 subjects.

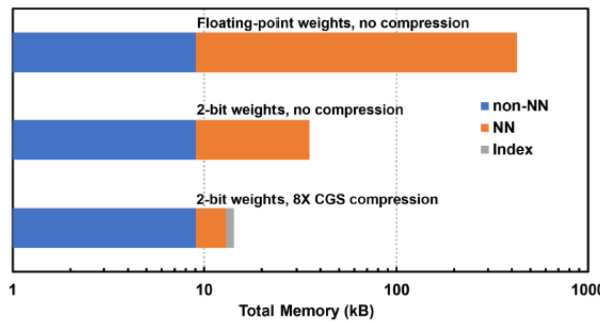


Fig. 12 Memory reduction by joint structured compression and low-precision optimization. [94]

Table VI Performance Summary of the AI-based ECG classification processors discussed in Section III.A.

Ref	Application	Platform & Resource	Method	Accuracy	Power
[87]	ECG Arrhythmia	ASIC 4.99mm² @90nm	SVM	95.8%	48.6μW @ 0.5V
[88]	ECG Arrhythmia	ASIC 0.12mm² @ 40nm	WLC and SVM	98.2%	1.98μW (WLC) @1.1V 3.76μW (SVM) @1.1V
[89]	ECG Arrhythmia	FPGA DSP: 42 FF: 9295 LUT: 15163	Neural Network	99.8%	-
[90] [91]	ECG Arrhythmia	ASIC 0.21mm² @ 65nm	SNAM	93.6%	234pJ/classification @1V
[92]	Biometric Authentication	FPGA FF: 4335 LUT: 3734	Neural Network	99.93%	31.75 mW
[93]	ECG Arrhythmia &	ASIC 5.88mm² @65nm	Neural Network	Arrhythmia Sensitivity: 93.13%	Arrhythmia 0.83μW @0.51V

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBCAS.2020.2974154, IEEE Transactions on Biomedical Circuits and Systems

	Biometric Authentication			Specificity: 89.78% Authentication 98.3%	Authentication 1.06μW @0.55V
[94]	Biometric Authentication	ASIC 0.59mm² @65nm	Neural Network	99%	59.4μW @1.2V

B. AI-based EEG Classification Processor Design

In the field of EEG classification, two important research topics include epilepsy seizure detection and emotion recognition.

By designing real-time and energy-efficient processor to perform on-device seizure detection, it is possible to alert people surrounding or directly prevent the seizure occurring through immediate stimulation. Accuracy is important for seizure detection to ensure safety. In [95] a high accuracy seizure detection processor is proposed. In this design, the hilbert transform (HT) is used as feature extractor for the input EEG signal. This choice is based on the report from [96][97] that HT is one of the most powerful approaches for extracting the features of nonlinear and nonstationary signals like EEG. To reduce the feature dimension so as to increase the processing speed and reduce the resource of hardware, efficient features are selected by applying mean power frequency on the feature vectors. Following that, a simple MLP neural network is utilized for classifying the EEG signal into normal and abnormal cases. The system architecture of the design is shown in Fig. 13. In the design, The HT was implemented based on the architecture proposed in [98], in which the HT is implemented as FIR filter. The design consumes 159.7 mW and achieves an accuracy of 100% on a user datasets for the seizure detection.

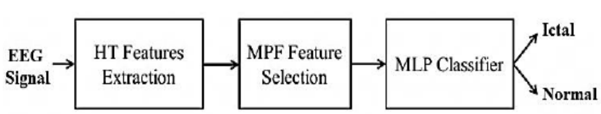


Fig. 13 System architecture of the seizure detection processor in. [95]

Other than accuracy, latency/power consumption/area are also important for seizure detection processors. The classifier plays a critical role in the optimization of these metrics. In [99] hardware architectures of 4 classifiers commonly used for seizure detection are proposed, including k-nearest neighbor (KNN), logistic regression (LR), Naive Bayes (NB), SVM, as shown in Fig. 14. The architecture of the feature extraction engine is also shown. Several techniques have been proposed to optimize the performance, power consumption and area. For SVM engine, the pipeline architecture has been adopted to improve the performance of dot product module. For KNN engine, parallel distance calculation has been used to balance between area, power and latency. Considering the low complexity of the LR and NB algorithm, serial processing architecture is adopted to reduce the area. In addition, as shown in Fig. 15, fixed-point data width optimization has been investigated to reduce the area overhead of the design while ensuring the accuracy. The experimental results show that the LR is the best in terms of accuracy, area overhead and power consumption.

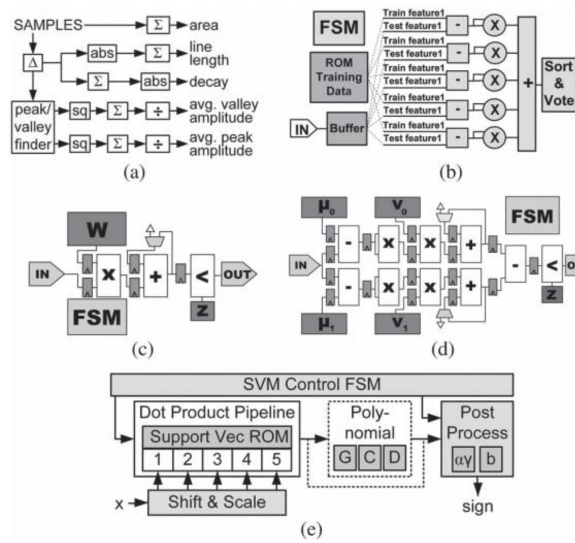


Fig. 14 Hardware architecture for the feature extraction and classification engines. (a) Feature Extraction; (b) KNN; (c) LR; (d) NB; and (e) SVM. [99]

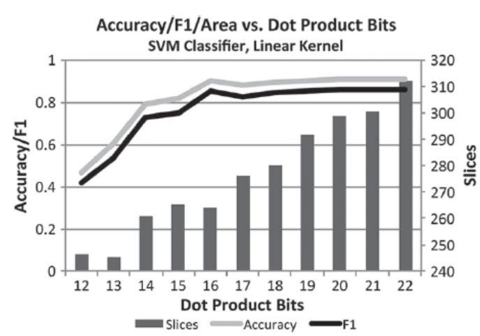


Fig. 15 Accuracy, F1 score, and area (FPGA slices) versus dot product word width for SVM classifier. [99]

Implemented with a 65 nm technology, the LR processor dissipates only 19 nJ per classification at 484 Hz.

Recently, neural network has been used in seizure detection and shows more powerful classification capability than conventional machine learning methods such as SVM and KNN. In [100] a low-power and low-cost seizure detection processor is proposed. In this design, the characteristics of EEG signal are extracted and a neural network is used for the seizure detection. Usually parallel

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBCAS.2020.2974154, IEEE Transactions on Biomedical Circuits and Systems

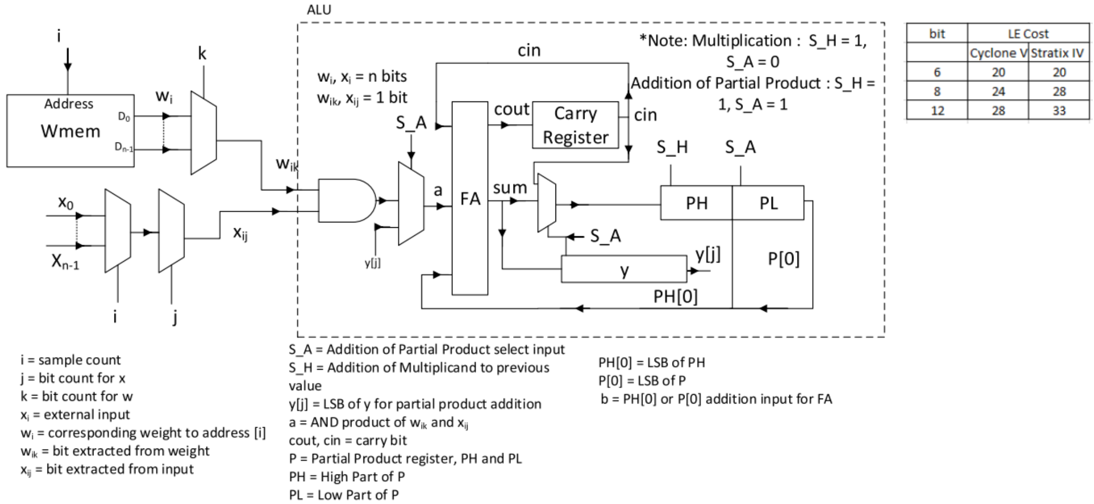


Fig. 16 Data processing unit proposed in. [100]

hardware architecture is used for neural network computation to improve the performance. However, in this design, the high performance is not a priority but low power and low cost is more emphasized, therefore a bit-serial architecture based data processing unit (DPU) is proposed for the neural network computation to reduce the power and cost. The proposed DPU is shown in Fig. 16. The Wmem is a SRAM used for storing the weights of the neural network. The ALU utilizes a custom multiplier for bit-serial processing. The DPU can be used in a vector arrangement and can be controlled by a simple state machine to perform the same operations.

In [101] a seizure detection processor using neural network as a classifier is proposed. Six features are extracted from the EEG signal, including the amplitude, frequency, approximate entropy and variance. The hardware design can be divided into two parts: feature extraction engine on FPGA and neural network engine on ASIC. The neural network part can be trained by itself to fit for different mice by self-adjusting the network parameters. However, the details of the self-adjusting process and hardware are not disclosed in the conference paper and may be included in the journal version.

In addition to epilepsy seizure detection, emotion recognition is another important research topic in the field of EEG classification. In [107], an emotion recogniton processor for detecting autism spectrum disorder (ASD) in children is developed. The design intgrates a hardware-efficient feature extraction engine and a patient-specific SVM engine. Several hardware-friendly features are utlilized including the power spectrial density (PSD), the absoluted difference of inter-hemispheric power asymmetry (IHPD) and the scaled inter-hemispheric power asymmetry ratio (SIHPR). In the SIHPR feature extraction engine, LUT-based divider is proposed to achieve high area and power efficiency. Implemented on a 65nm CMOS process, the design classifies valence and arousal with an accuracy of 63 % and 60% while comsuming 12.7 μW with a latency of 0.8 seconds for each classification.

In [108][109], a real-time high accuracy emotion recogniton system with optimized CNN processor chip has been proposed. In the design a subject-dependent methodology with on-line learning has been proposed to address the issue that defferent subjects undergoing the same emotional stimuli will have significant response on different EEG channels. A CNN processor chip with on-chip learning engine has been designed to assist this methodology while achieving real-time performance with low power consumption. Implmented on a 28nm CMOS process, the proposed design achieves an average classification accuracy of 70.51% for 4 class emotion recognition with a latency of 450 ms while consumes 71.6 mW and 29.5 mW in the training mode and testing mode, repsectively.

Table VII Performance Summary of the AI-based EEG classification processors discussed in Section III.B

Ref	Application	Platform & Resource	Method	Accuracy	Power
[95]	Epilepsy Seizure Detection	FPGA FF: 357 4LUT: 5355	Neural Network	100%	159.7mW
[99]	Epilepsy Seizure Detection	ASIC $0.008mm^2$ @65nm	KNN, LR, NB, SVM	91% F1 Score (LR)	37nW @1V
[100]	Epilepsy Seizure Detection	FPGA -	Neural Network	90%	-
[101]	Epilepsy Seizure Detectio	ASIC $1.98mm^2$	Neural Network	89.88 %	1.589mW @1.8V

	n	@ 0.18μm			
[107]	Emotion Recognition	ASIC -	SVM	63% (valence) 60% (arousal)	12.7μW
[108]	Emotion Recognition	AISC $3.35mm^2$ @28nm	Neural Network	70.51 %	71.6mW (training) 29.5mW (testing)

C. *AI-based EMG Classification Processor Design*

In addition to ECG and EEG, EMG processors have attracted lots of attentions in the recent years as it is commonly used for gesture recognition applications. As the gesture recognition is usually based on wearable devices, the design focus also includes co-optimization between classification accuracy and power consumption.

Microcontrollers have been used to realize low power EMG classification for gesture control. In [102] a low power embedded system for EMG acquisition and gesture recognition is proposed. This work focuses on multi-level design optimization, including software and hardware. Fig. 17 shows the system architecture of the proposed design. In addition to EMG sensor interface, inertial and pressure sensors have been used to for sensor fusion to improve the accuracy of gesture recognition and motion tracking. In the system, an ARM Corext-M4 based MCU (STM32F407) is used for real-time signal processing and EMG classification. A previously designed analog frontend ASIC Cerebro (as shown in Fig. 18) is used for data acquisition and interfaces with the MCU via SPI. The Cerebro has 8 differential data acquisition channels that are multiplexed for data sampling with a typical sampling frequency of 1KHz. Each channel includes a variable-gain instrumentation amplifier (IA) followed by an active RC low-pass filter. The Cerebro also includes a low-impedance patient ground (PGND) to set the input common mode and an auxiliary ADC to measure the internal temperature or sample an additional input. In the system, a Bluetooth module is used to transmit out the gesture recognition result or the acquired EMG data. A SVM classifier is implemented on the MCU for the EMG classification and gesture recognition, which achieves real-time performance and low power consumption. The platform achieves a classification rate of 90% while consuming 86

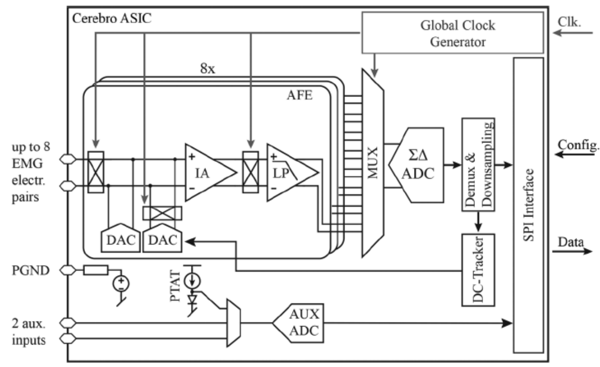
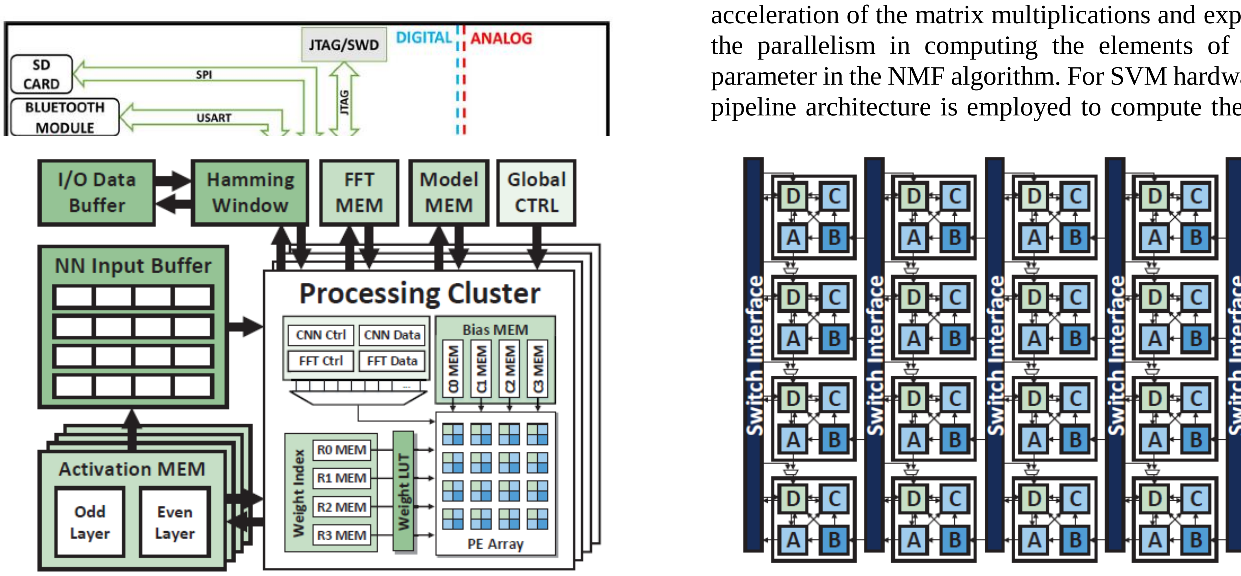


Fig. 18 System architecture of the Cerebro ASIC. [102]

mW for classification of 7 gestures at 1 KHz and 29.7 mW for the recognition at 25 Hz.

Compared to microcontrollers, FPGA can be used to implement algorithms with higher complexity so as to achieve better accuracy. In [103] a FPGA-based EMG classification processor is proposed. In this design non-negative matrix factorization (NMF) algorithm is used for feature extraction and SVM algorithm is used as classifier. Both the algorithms are implemented on a FPGA. For the NMF hardware design, the major optimization focuses on the acceleration of the matrix multiplications and exploration of the parallelism in computing the elements of H and W parameter in the NMF algorithm. For SVM hardware design, pipeline architecture is employed to compute the results of



(a) (b)

Fig. 19 Hardware architecture of the acoustic processor for intelligent hearing aid application. [105]

all the learners and then the results are compared with a lookup table storing the coding matrix. This design achieves an accuracy of 98% on 5 types of movements, including hand close, thumb close, thumb-index, middle-ring, and middle-ringlittle. Compared with the design implemented on embedded system, the processing speed is improved by 24×.

For achieving ultra-low power consumption, ASIC-based EMG classification processor has been investigated, which shows much lower power consumption than their counterparties based on microcontrollers and FPGAs. In [104] an ultra-low power EMG classification processor using spiking neural network (SNN) is proposed. This design is implemented on ASIC. The SNN processing engine implements a scalable multi-core neuromorphic computing architecture. It uses a asynchronous mixed-signal computing architecture with asynchronous inter-core and inter-chip communication. It contains 4 cores and each core comprises 256 neurons for a total of 1000 neurons per chip. Each neuro can support up to 64 synapses. Analog circuits are used to implement the neurons and synapses with biologically plausible temporal dynamics. These circuits operate in the sub-threshold domain so as to reduce the dynamic power consumption. The spikes are transmitted on chip following the address event representation (AER) communication protocol. The design achieving an accuracy of 84% for EMG classification, while consuming only 0.05 mW which is much lower than other existing designs.

Table VIII Performance Summary of the AI-based EMG classification processors discussed in Section III.C.

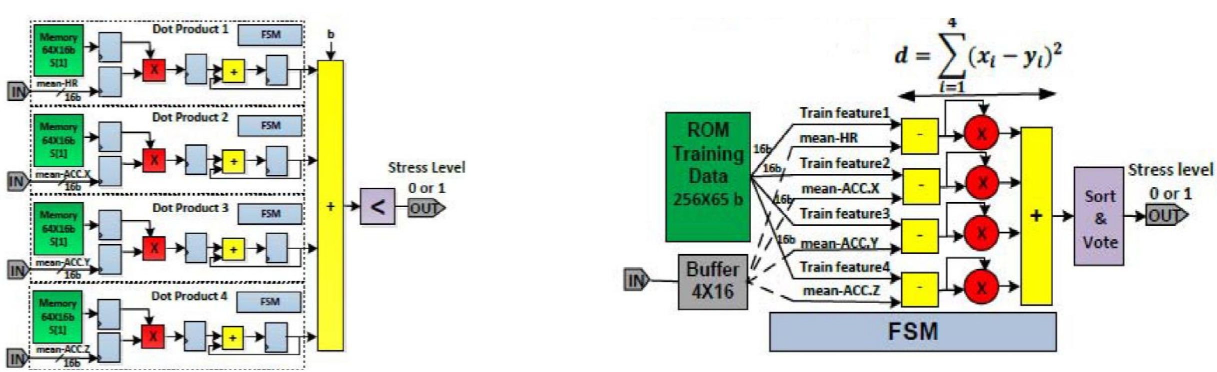
Ref	Application	Platform & Resource	Method	Accuracy	Power
[102]	EMG-based Gesture Recognition	MCU	SVM	90%	29.7 mW
[103]	EMG-based Gesture Recognition	FPGA (Xilinx Pynq-Z1 development board) DSP: 67% FF: 41% LUT: 73%	SVM	98%	-
[104]	EMG-based Gesture Recognition	ASIC 43.79 mm² @0.18μm	SNN	84%	0.05 mW @1.8V

D. *Other AI-based Biomedical Processors Design*

In addition to ECG, EEG and EMG processors, other AI-based biomedical processors have been investigated, such as acoustic processor for hearing aid application and stress detection for health monitoring application.

In [105] a low power acoustic processor with neural network and FFT hardware accelerators for intelligent hearing aid application is proposed. In the design, the input speech signals are transformed from time domain into the frequency domain using FFT, then the feature extraction is performed followed by convolutional neural network to generate a mask to remove the noise in the speech signals so as to improve the signal quality. As shown in Fig. 19 (a) and (b), the neural network and FFT hardware accelerators reuse the same PE array for realizing different operations. Each PE module contains a dual-MAC unit and a butterfly unit. For the FFT mode, the dual-MAC unit is used to calculate the input with twiddle factor in the real and imaginary components, and the butterfly unit is used for butterfly operation. For the CNN mode, the dual-MAC unit is used to calculate the input data with weights from 2 different filters, and the butterfly unit is used to accumulate the partial sums. In addition, CORDIC is included for calculating the required non-linear functions. In Fig. X (b), in each PE module, 'A' represents ALU, 'B' represents PE control, 'C' represents CORDIC and 'D' represents data registers. The design enhances the speech intelligibility by up to 41% in a low SNR condition and consumes only 2.17 mW while operating at 5 MHz.

Stress detection is one of the important approaches for health monitoring. In [106] a low power biomedical processor to detect stress using multi-modal physiological signals is proposed, including ECG, acceleration, respiration



(a) (b)

Fig. 20 Classification processors for stress detection (a) SVM classification processor (b) KNN classification processor. [106]

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBCAS.2020.2974154, IEEE Transactions on Biomedical Circuits and Systems

rate, and SpO2. KNN and SVM based classifiers with feature extraction were used to detect the stress. As shown in Fig. 20 (a) and (b), the SVM classification processor employs a reconfigurable parallel pipelined architecture to support different number of supporting vectors for various applications. The KNN classification processor utilizes a scalable semi-parallel architecture to support different number of features and different size of training data. A ROM is used to store the training data and s sorting block is used to find the K samples with the smallest distances. The proposed SVM and KNN processors implemented on ASIC with 65nm technology achieves classification accuracy of 96.7% and 95.8%, while consuming 39 mW and 77 mW power respectively.

Table IX Performance Summary of the AI-based biomedical processors discussed in Section III.D.

Ref	Application	Platform & Resource	Method	Accuracy	Power
[105]	Hearing Aid	ASIC 4.2mm² @40nm	Neural Network	Short-time objective intelligibility (STOI): improved by up to 41%	2.17 mW @0.6-0.9V
[106]	Stress Detection	ASIC SVM: 0.17mm² @65nm KNN: 0.3 mm² @65nm	SVM & KNN	SVM: 96.7% KNN: 95.8%	SVM: 39 mW @1V KNN: 77 mW @1V

IV. Outlook on AI-based Biomedical Circuits & Systems

In the past few years, the AI-based biomedical circuits & systems have seen remarkable progress. Nevertheless, the existing work still have many limitations.

First, regarding the applications, the designs mainly focus on the ECG/EEG/EMG applications. These applications have many similarities in terms of signal feature, processing approach and hardware architecture. So the design knowledge can be well shared. In the future, AI will be applied to more and more biomedical applications with different signal feature and processing approach, such as hearing aid, diabetic foot monitoring and endoscopy, where the processed signal changes from bioelectronics signal to sound, image or pressure distribution. These applications bring brand new design approach on both algorithm and hardware, triggering new signal processing flow and hardware architecture. They also bring design challenges including higher performance requirement, higher power consumption and larger storage, which needs to be addressed through cross-layer innovations and optimizations.

Second, regarding the algorithm design, the corresponding algorithms have been used in many fields and achieved remarkable performance. For AI algorithms, our goal is to learn a stable model with good performance in all aspects, but the actual situation is often not so ideal. In some applications, we can only get a weak classifier (weak supervisory model). Current research shows that the ensemble learning may be potential to improve the situation. In ensemble learning, several weak classifiers are combined to obtain a strong and better classifier (strong supervisory model). If one weak classifier makes wrong decision, the other classifiers can correct the error. That's why the learning algorithm, random forest and Gradient Boosting Decision Tree (GBDT) have received a lot of attention in recent years. Artificial intelligence algorithms are widely used because of the superior performance. At the same time, the limited interpretability of AI algorithms brings many doubts and concerns. This is related to whether the algorithm can be trusted for application. The future development of AI algorithm cannot avoid this problem. It is possible to study how to better design algorithms that take account of interpretability.

Third, regarding the hardware design, many designs integrated machine learning accelerators in the biomedical processors to perform classification for intelligent sensing. However, the selection of the classifier architecture is not thoroughly investigated. For different applications the requirements on the accuracy, performance, power consumption and area overhead differ which heavily affects the selection of the classifier architecture. Even for the same application, the requirement may vary over time. This calls for the design of scalable or hybrid classifiers which can be dynamically reconfigured among different classifier architecture or type to suit different tasks over time. Most of the existing designs follow a conventional processing flow where feature extraction is needed before the classification. As the end-to-end classification approach is introduced, the feature extraction and classification can be merged using neural network to automate the feature extraction and selection, which help achieve better classification accuracy. However, this comes at the price of computation complexity, which leads to significant increase in the power consumption and processing time. Therefore, to address this issue, low complexity algorithms and low power hardware architectures are needed. Furthermore, algorithm and hardware co-design can be investigated jointly which may prove more promising in achieving a good tradeoff between low computational complexity and power usage while still maintaining high classification accuracies. Another drawback of neural network based methods is that a large amount of training data is required to achieve high accuracy. To address this issue, a potential direction is to combine the conventional machine learning methods (e.g SVM) which is model driven and the neural network based deep learning methods which is data driven to reduce the reliance on the amount of training data. In addition, on-line training can also be utilized to reduce the amount of training data. This requires new hardware architectures and circuits supporting

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBCAS.2020.2974154, IEEE Transactions on Biomedical Circuits and Systems

efficient on-chip training. Another trend is the in-memory computing technologies using mixed-signal circuits which helps reduce the processing time, power consumption and area overhead. This includes in-memory computing technologies based on SRAM and emerging memory such as RRAM and PCRAM. Last but not the least, we anticipate that a close interaction between algorithm and hardware design would lead to continuous innovation and breakthrough in the AI-based biomedical applications.

V. CONCLUSIONS

Recently, AI has been widely applied to biomedical applications such as ECG monitoring, seizure detection, emotion recognition and hearing aid to help improve the accuracy and intelligence of health monitoring. This paper reviews the state of the arts AI-based biomedical processing algorithms and hardware design methods for various applications, including EEG, EEG, EMG, blood pressure, hearing aid, and other biomedical applications. The advantages and disadvantages of the existing design methods using machine learning and deep learning have been discussed and compared in terms of accuracy, performance, power consumption and hardware cost. The research trends of AI-based biomedical processing algorithms and processors have also been discussed, such as low complexity & interpretable algorithms, low power & efficient hardware architecture, AI hardware with online learning capability and AI hardware using mixed-signal circuits. The goal of this paper is to provide an overview on how the AI technology can benefit to biomedical applications, what challenges have been faced, what methods have been proposed on algorithm and hardware design levels, and what the future direction is.

REFERENCES

[1] S. Benatti et al., “A Versatile Embedded Platform for EMG Acquisition and Gesture Recognition,” IEEE Transactions on Biomedical Circuits and Systems, vol. 9, no. 5, pp. 620–630, Oct. 2015.

[2] A. Page, A. Kulkarni, and T. Mohsenin, “Utilizing deep neural nets for an embedded ECG-based biometric authentication system,” in 2015 IEEE Biomedical Circuits and Systems Conference (BioCAS), 2015, pp. 1–4.

[3] G. Franco, P. Cancian, L. Cerina, E. Besana, N. Beretta, and M. D. Santambrogio, “FPGA-based muscle synergy extraction for surface EMG gesture classification,” in 2017 IEEE Biomedical Circuits and Systems Conference (BioCAS), 2017, pp. 1–4.

[4] S. Yin et al., “A 1.06 μw smart ecg processor in 65 nm cmos for real-time biometric authentication and personal cardiac monitoring,” in 2017 Symposium on VLSI Circuits, 2017, pp. C102–C103.

[5] E. Donati, M. Payvand, N. Risi, R. B. Krause, and G. Indiveri, “Discrimination of EMG Signals Using a Neuromorphic Implementation of a Spiking Neural Network,” IEEE Transactions on Biomedical Circuits and Systems, pp. 1–1, 2019.

[6] A. H. Khandoker, M. Palaniswami and C. K. Karmakar, "Support Vector Machines for Automated Recognition of Obstructive Sleep Apnea Syndrome From ECG Recordings," in *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 1, pp. 37-48, Jan. 2009.

[7] Li, Gang, and Wan-Young Chung. "Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier," *Sensors*, vol. 13, no. 12, pp. 16494-16511, Dec. 2013.

[8] Q. Li, C. Rajagopalan and G. D. Clifford, "Ventricular Fibrillation and Tachycardia Classification Using a Machine Learning Approach," in *IEEE* Transactions *on Biomedical Engineering*, vol. 61, no. 6, pp. 1607-1613, June 2014.

[9] M. Bsoul, H. Minn and L. Tamil, "Apnea MedAssist: Real-time Sleep Apnea Monitor Using Single-Lead ECG," in *IEEE Transactions on Information* Technology *in Biomedicine*, vol. 15, no. 3, pp. 416-427, May 2011.

[10] R. J. Martis, U. R. Acharya, et al. "Application of principal component analysis to ECG signals for automated diagnosis of cardiac health," *Expert Systems* with *Applications*,vol. 39,no. 14 pp. 11792-11800, Oct. 2012.

[11] F. Melgani and Y. Bazi, "Classification of Electrocardiogram Signals With Support Vector Machines and Particle Swarm Optimization," in *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 5, pp. 667-677, Sept. 2008.

[12] A. Daamouche, L. Hamami , N. Alajlan, et al. "A wavelet optimization approach for ECG signal classification," *Biomedical Signal Processing and Control* ,vol. 7, no .4, pp. 342-349, July. 2012.

[13] S. Kiranyaz, T. Ince and M. Gabbouj, "Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks," in *IEEE Transactions on Biomedical* Engineering, vol. 63, no. 3, pp. 664-675, March. 2016.

[14] U. R. Acharya, H. Fujita, O. S. Lih, et al. "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Information sciences*, vol. 405, pp. 81-90.Sept. 2017.

[15] U. R. Acharya, H. Fujita, O. S. Lih, et al. "Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network," *Knowledge-Based Systems*, vol. 132, pp. 62-71, Sept. 2017.

[16] U. R. Acharya, H. Fujita, O. S. Lih, et al."Automated identification of shockable and non-shockable life-threatening ventricular arrhythmias using convolutional neural network," *Future Generation Computer Systems*, vol. 79, pp. 952-959, Feb. 2018.

[17] U. R. Acharya, H. Fujita, O. S. Lih, et al."Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals," *Information Sciences*, vol. 415, pp. 190-198, Nov. 2017.

[18] U. R. Acharya, H. Fujita, O. S. Lih, et al. "A deep convolutional neural network model to classify heartbeats," *Computers in biology and medicine*, vol. 89, pp. 389-396, Oct. 2017.

[19] Hassan Hamsa Haseena, Abraham Mathew and Joseph K Paul, “Fuzzy clustered probabilistic and multi layered feed forward neural networks for electrocardiogram arrhythmia classification,” *Journal of Medical Systems*, vol. 35, no. 2, pp. 179-188, Apr. 2011.

[20] Jeen-Shing Wang, Wei-Chun Chiang and Yu-Liang Hsu, “ECG arrhythmia classification using a probabilistic neural network with a feature reduction method,” in *7th International Conference on Intelligent Computing*, Zhengzhou, China, 2011, pp. 38-45.

[21] Ubeyli Elif Derya, “Combining recurrent neural networks with eigenvector methods for classification of ECG beats,” *Digital Signal Processing*, vol. 19, no. 2, pp. 302-329, Mar. 2009.

[22] Ubeyli Elif Derya, “Recurrent neural networks employing Lyapunov exponents for analysis of ECG signals,” *Expert Sytems with Applications*, vol. 37. no.2, pp. 1192-1199, Mar. 2010.

[23] Yildirim Ozal, “A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification,” *Computers in biology and medicine*, vol. 96, pp. 189-202, May 2018.

[24] Chauhan Sucheta, Vig Lovekesh and Ahmad Shandar, “ECG anomaly class identification using LSTM and error profile modeling,” *Computers in biology and medicine*, vol.109, pp. 14-21, Jun. 2019.

[25] Thomas Manu, Das Manab Kr and Ari Samit, “Automatic ECG arrhythmia classification using dual tree complex wavelet based features,” *Aeu*-International *Journal of electronics and communications*, vol. 69. no. 4, pp. 715-721, 2015.

[26] Hong-Qiang Li, Dan-Yang Yuan, Xiao-Dong Ma, Dian-Yin Cui and Lu Cao, “Genetic algorithm for the optimization of features and neural networks in ECG signals classification,” *SCIENTIFIC REPORTS*, vol. 7. Jan. 2017.

[27] Lin Shu, Jin-Yan Xie and Ming-Yue Yand, “A Review of Emotion Recognition Using Physiological Signals,” *SENSORS*, vol. 18, no. 7, Jul. 2018.

[28] Min-Ki Kim, Miyoung Kim and Eunmi Oh, “A Review on the Computational Methods for Emotional State Estimation from the Human EEG,” *Computational and Mathematical Methods in Medicine*, 2013.

[29] Xiao-Wei Wang, Dan Nie and Bao-Liang Lu, “EEG-Based Emotion Recognition Using Frequency Domain Features and Support Vector Machines,” in *18th ICONIP*, Shanhai, China, 2011, pp. 734-.

[30] M. Park, H. Oh, H. Jeong and J. Sohn, "Eeg-based emotion recogntion during emotionally evocative films," *2013 International Winter Workshop on Brain-Computer Interface (BCI)*, Gangwo, 2013, pp. 56-57.

[31] R. Duan, J. Zhu and B. Lu, "Differential entropy feature for EEG-based emotion *classification*," *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, San Diego, CA, 2013, pp. 81-84.

[32] W. Zheng and B. Lu, "Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks," in *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162-175, Sept. 2015.

[33] Suwicha J, Setha P N, and Pasin I, "EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation," in *The Scientific World Journal*, vol. 2014, pp. 1-10, Sept. 2014.

[34] M. Murugappan, M. Rizon, R. Nagarajan, S. Yaacob, D. Hazry and I. Zunaidi, "Time-Frequency *Analysis* of EEG Signals for Human Emotion Detection," *2008 4th Kuala Lumpur International Conference on Biomedical Engineering*, Kuala Lumpur, 2008, pp. 262-265.

[35] Koelstra S et al., "Single Trial Classification of EEG and Peripheral Physiological Signals *for* Recognition of Emotions Induced by Music Videos," *2010 International Conference on Brain Informatics*, Berlin, Heidelberg, 2010, pp. 89-100.

[36] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion Recognition From EEG Using Higher Order Crossings," *in IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 186-197, March 2010.

[37] Amin H U, Malik A S, Kamel N and Hussain M, " A Novel Approach Based on Data *Redundancy* for Feature Extraction of EEG Signals," in *Brain Topography*, vol. 29, no. 2, pp. 207-217, March 2016.

[38] Wang X W, Nie D and Lu B L, " Emotional State Classification from EEG *Data* Using Machine Learning Approach," in *Neurocomputing*, vol. 129, no. 4, pp. 94-106, April 2014.

[39] Pane E S, Wibawa A D and Purnomo M H, "Improving the accuracy of EEG emotion recognition by combining valence lateralization and ensemble learning with tuning parameters," in *Cognitive processing*, pp. 1-13, 2019.

[40] A. T. Tzallas, M. G. Tsipouras and D. I. Fotiadis, "Epileptic Seizure Detection in EEGs Using Time–Frequency Analysis," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 703-710, Sept. 2009.

[41] R. Panda, P. S. Khobragade, P. D. Jambhule, S. N. Jengthe, P. R. Pal and T. K. Gandhi, "Classification of EEG signal using wavelet transform and support vector machine for epileptic seizure diction," *2010 International Conference on Systems in Medicine and Biology*, Kharagpur, 2010, pp. 405-408.

[42] Y. Liu, W. Zhou, Q. Yuan and S. Chen, "Automatic Seizure Detection Using Wavelet Transform and SVM in Long-Term Intracranial EEG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 6, pp. 749-755, Nov. 2012.

[43] Kannathal N , Choo M L , Acharya U R , et al, "Entropies for detection of epilepsy in EEG," *Computer Methods & Programs in Biomedicine*, vol.80, no.3, pp.187-194, 2005.

[44] M. Z. Parvez and M. Paul, "Epileptic seizure detection by exploiting temporal correlation of electroencephalogram signals," *IET Signal Processing*, vol. 9, no. 6, pp. 467-475, Aug 2015.

[45] Behara D S T, Kumar A, Swami P, et al, "Detection of epileptic seizure patterns in EEG through fragmented feature extraction," *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE, 2016: 2539-2542.

[46] S. M. S. Alam and M. I. H. Bhuiyan, "Detection of Seizure and Epilepsy Using Higher Order Statistics in the EMD Domain," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 2, pp. 312-318, March 2013.

[47] Guzmán J C, Miramontes I, Melin P, et al., " Optimal genetic design of type-1 and interval type-2 fuzzy systems for blood pressure level classification," in Axioms, vol. 8, no. 1, pp. 1-35, January 2019.

[48] Melin Patricia, Miramontes Ivette and Prado-Arechiga German, " A hybrid model based on modular neural networks and fuzzy systems for classification of blood pressure and hypertension risk diagnosis," in *Expert systems with applications*, vol. 107, pp. 146-164, October 2018.

[49] Guzman J C, Melin P and Prado-Arechiga G, " Design of an Optimized Fuzzy Classifier for the Diagnosis of Blood Pressure with a New Computational Method for Expert Rule Optimization," in *Algorithms*, vol. 10, no. 79, pp. 1-27, July 2017.

[50] S. Das, P. K. Ghosh and S. Kar, "Hypertension diagnosis: A comparative study using fuzzy expert system and neuro fuzzy system," *2013 IEEE* International *Conference on Fuzzy Systems (FUZZ-IEEE)*, Hyderabad, 2013, pp. 1-7.

[51] A. Akutekwe and H. Seker, "A hybrid dynamic Bayesian network approach for modelling temporal associations of gene expressions for hypertension diagnosis," *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Chicago, IL, 2014, pp. 804-807.

[52] C. Barbieri, I. Cattinelli, L. Neri, F. Mari, R. Ramos, D. Brancaccio, B. Canaud and S. Stuard, "Development of an Artificial Intelligence Model to Guide the Management of Blood Pressure, Fluid Volume, and Dialysis Dose in End-Stage Kidney Disease Patients: Proof of Concept and First Clinical Assessment, " *International Journal of Medical Informatics*, vol. 5, no. 1, pp. 28-33, Feb. 2019.

[53] T. H. Wu, G. K. Pang and E. W. Kwong, "Predicting Systolic Blood Pressure Using Machine Learning," *7th International Conference on Information and Automation for Sustainability*, Colombo, 2014, pp. 1-6.

[54] S. Lee and J. Chang, "Oscillometric Blood Pressure Estimation Based on *Deep* Learning," in *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 461-472, April 2017.

[55] F. Pan, P. He, F. Chen, J. Zhang, H. Wang and D. Zheng, "A novel *deep* learning based automatic auscultatory method to measure blood pressure, " *International Journal of Medical Informatics*, vol. 128, pp. 71-78,August 2019.

[56] D. Wang and J. Chen, "Supervised Speech Separation Based on Deep Learning: An Overview," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 10, pp. 1702-1726, Oct. 2018.

[57] Y. Wang and D. Wang, "Towards Scaling Up Classification-Based Speech Separation," in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 7, pp. 1381-1390, July 2013.

[58] Y. Xu, J. Du, L. Dai and C. Lee, "An Experimental Study on Speech Enhancement Based on Deep Neural Networks," in *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 65-68, Jan. 2014.

[59] J. Du, Y. Tu, L. Dai and C. Lee, "A Regression Approach to Single-Channel Speech Separation Via High-Resolution Deep Neural Networks," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 8, pp. 1424-1437, Aug. 2016.

[60] Y. Zhao, Z. Wang and D. Wang, "A two-stage algorithm for noisy and reverberant speech enhancement," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, 2017, pp. 5580-5584.

[61] X. Zhang and D. Wang, "A Deep Ensemble Learning Method for Monaural Speech Separation," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 5, pp. 967-977, May 2016.

[62] J. Chen, Y. Wang, S. E. Yoho, D. L. Wang, and E. W. Healy, "Largescale training to increase speech intelligibility for hearing-imparied listeners in novel noises, " *J. Acoust. Soc. Amer*, vol. 139, no. 5, pp. 2604–2612, May. 2016.

[63] J. Chen and D. L. Wang, "Long short-term memory for speaker generalization in supervised speech separation," *J. Acoust. Soc. Amer*, vol. 141, no. 6, pp. 4705-4714, June. 2017.

[64] Y. Jiang, D. Wang, R. Liu and Z. Feng, "Binaural Classification for Reverberant Speech Segregation Using Deep Neural Networks," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 2112-2121, Dec. 2014.

[65] N. Fan, J. Du and L. Dai, "A regression approach to binaural speech segregation via deep neural network," *2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, Tianjin, 2016, pp. 1-5.

[66] X. Zhang and D. L. Wang, “Deep learning based binaural speech separation in reverberant environments*,” IEEE/ACM Trans. Audio Speech Lang. Process*., vol. 25, no. 5, pp. 1075–1084, May 2017.

[67] H. Erdogan, J. R. Hershey, S. Watanabe, M. Mandel, and J. L. Roux, “Improved MVDR beamforming using single-channel mask prediction networks,” in *Proc. INTERSPEECH*, 2016, pp. 1981–1985.

[68] J. Heymann, L. Drude and R. Haeb-Umbach, "Neural network based spectral mask estimation for acoustic beamforming," *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, 2016, pp. 196-200.

[69] X. Zhang, Z. Wang and D. Wang, "A speech enhancement algorithm by iterating single- and multi-microphone processing and its application to robust ASR," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, 2017, pp. 276-280.

[70] Subasi, and Abdulhamit, "Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders," *Computers in Biology and Medicine*, vol. 43, no. 5, pp. 576-586, February 2013.

[71] Gokgoz, Ercan; Subasi, and Abdulhamit, "Comparison of decision tree algorithms for EMG signal classification using DWT," *Biomedical signal processing and control*, vol. 18, pp. 138-144, 2015.

[72] E. J. Scheme, K. B. Englehart and B. S. Hudgins, "Selective Classification for Improved Robustness of Myoelectric Control Under Nonideal Conditions," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 6, pp. 1698-1705, June 2011.

[73] Angkoon, Phinyomark,Franck, Quaine,Sylvie, Charbonnier,Christine, Serviere,Franck, Tarpin-Bernard, and Yann, Laurillau, "Feature extraction of the first difference of EMG time series for EMG pattern recognition," *Computer Methods and Programs in Biomedicine*, vol. 117, no. 2, pp. 247-256, November 2014.

[74] G. R. Naik, D. K. Kumar and Jayadeva, "Twin SVM for Gesture Classification Using the Surface Electromyogram," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 301-308, March 2010.

[75] Gijsberts, Arjan, Atzori, Manfredo, Castellini, Claudio, Müller, Henning, and Caputo, Barbara, "Movement Error Rate for Evaluation of Machine Learning Methods for sEMG-Based Hand Movement Classification," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 22, no. 4, pp. 735-744, January 2014.

[76] Geethanjali, P.,and K. K. Ray, "Identification of motion from multi-channel EMG signals for control of prosthetic hand," *Australasian Physical & Engineering Sciences in Medicine* , vol. 34, no. 3, pp. 419-427, June 2011.

[77] S. S. Nair, R. M. French, D. Laroche and E. Thomas, "The Application of Machine Learning Algorithms to the Analysis of Electromyographic Patterns From Arthritic Patients," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 2, pp. 174-184, April 2010.

[78] Abràmoff, Michael D., et al. "Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices." *Npj Digital Medicine* vol. 1, no. 1, pp. 39, August 2018.

[79] X. Tan, C. Chng, Y. Su, K. Lim and C. Chui, "Robot-Assisted Training in Laparoscopy Using Deep Reinforcement Learning," in *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 485-492, April 2019.

[80] Haloi, M. "Towards Ophthalmologist Level Accurate Deep Learning System for OCT Screening and Diagnosis," *arXiv preprint arXiv: 1812.07105,* Dec. 2018. [Online]. Available: https://arxiv.org/abs/1812.07105

[81] S. A. Siddiqui, Y. Zhang, J. Lloret, H. Song and Z. Obradovic, "Pain-Free Blood Glucose Monitoring Using Wearable Sensors: Recent Advancements and Future Prospects," in *IEEE Reviews in Biomedical Engineering*, vol. 11, pp. 21-35, 2018.

[82] Q. Zhang, Y. Liu, H. Han, J. Shi and W. Wang, "Artificial Intelligence Based Diagnosis for Cervical Lymph Node Malignancy Using the Point-Wise Gated Boltzmann Machine," in *IEEE Access*, vol. 6, pp. 60605-60612, 2018.

[83] C. Lin, Y. Chang, C. K. Lin, L. Tsai and J. Chen, "Development of an AI-based non-invasive Pulse AudioGram monitoring device for arrhythmia screening," *2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT)*, Bethesda, MD, 2017, pp. 40-43.

[84] X. Liang, P. Hu, L. Zhang, J. Sun and G. Yin, "MCFNet: Multi-Layer Concatenation Fusion Network for Medical Images Fusion," in *IEEE Sensors Journal*, vol. 19, no. 16, pp. 7107-7119, 15 Aug.15, 2019.

[85] D. J. Hemanth, C. K. S. Vijila, et al., "Distance metric-based time-efficient fuzzy algorithm for abnormal magnetic resonance brain image segmentation," *Neural Computing and Applications*, vol. 22, no. 5, pp. 1013-1022, Apr. 2013.

[86] Q. Jin, Z. Meng, et al. "DUNet: A deformable network for retinal vessel segmentation," *KNOWLEDGE-BASED SYSTEMS*, vol. 178, pp. 149-162, Aug. 2019.

[87] S. Hsu, Y. Ho, P. Chang, C. Su, and C. Lee, "A 48.6-to-105.2 μW Machine Learning Assisted Cardiac Sensor SoC for Mobile Healthcare Applications," IEEE Journal of Solid-State Circuits, vol. 49, no. 4, pp. 801–811, Apr. 2014.

[88] Z. Chen et al., "An Energy-Efficient ECG Processor with Weak-Strong Hybrid Classifier for Arrhythmia Detection," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 65, no. 7, pp. 948–952, Jul. 2018.

[89] M. Wess, P. D. S. Manoj, and A. Jantsch, "Neural network based ECG anomaly detection on FPGA and trade-off analysis," in 2017 IEEE International Symposium on Circuits and Systems (ISCAS), 2017, pp. 1–4.

[90] P. Chollet, R. Pallas, C. Lahuec, M. Arzel, and F. Seguin, "A sub-nJ CMOS ECG classifier for wireless smart sensor," in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2017, pp. 3840–3843.

[91] B. Larras, P. Chollet, C. Lahuec, F. Seguin, and M. Arzel, "A fully flexible circuit implementation of clique-based neural networks in 65-nm CMOS," in 2018 IEEE International Symposium on Circuits and Systems (ISCAS), 2018, pp. 1–4.

[92] A. Page, A. Kulkarni, and T. Mohsenin, "Utilizing deep neural nets for an embedded ECG-based biometric authentication system," in 2015 IEEE Biomedical Circuits and Systems Conference (BioCAS), 2015, pp. 1–4.

[93] S. Yin et al., "A 1.06 μw smart ecg processor in 65 nm cmos for real-time biometric authentication and personal cardiac monitoring," in 2017 Symposium on VLSI Circuits, 2017, pp. C102–C103.

[94] S. K. Cherupally et al., "ECG Authentication Neural Network Hardware Design with Collective Optimization of Low Precision and Structured Compression," in 2019 IEEE International Symposium on Circuits and Systems (ISCAS), 2019, pp. 1–5.

[95] H. G. Daoud, A. M. Abdelhameed, and M. Bayoumi, "FPGA Implementation of High Accuracy Automatic Epileptic Seizure Detection System," in 2018 IEEE 61st International Midwest Symposium on Circuits and Systems (MWSCAS), 2018, pp. 407–410.

[96] S. Wang, B. Guo, C. Zhang, X. Bai, and Z. Wang, "EEG detection and de-noising based on convolution neural network and Hilbert-Huang transform," in 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2017, pp. 1–6.

[97] M. Berthelot, A. Witon, and L. Li, "Unconsciousness state identification using phase information extracted by Wavelet and Hilbert transform," in 2017 22nd International Conference on Digital Signal Processing (DSP), 2017, pp. 1–5.

[98] M. Kumm and M. S. Sanjari, "Digital hilbert transformers for FPGA-based phase-locked loops," in 2008 International Conference on Field Programmable Logic and Applications, 2008, pp. 251–256.

[99] A. Page, C. Sagedy, E. Smith, N. Attaran, T. Oates, and T. Mohsenin, "A Flexible Multichannel EEG Feature Extractor and Classifier for Seizure Detection," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 62, no. 2, pp. 109–113, Feb. 2015.

[100] S. M. Kueh and T. J. Kazmierski, "Low-Power and Low-Cost Dedicated Bit-Serial Hardware Neural Network for Epileptic Seizure Prediction System," IEEE Journal of Translational Engineering in Health and Medicine, vol. 6, pp. 1–9, 2018.

[101] C. Tsou, C. Liao, and S. Lee, "Epilepsy Identification System with Neural Network Hardware Implementation," in 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), 2019, pp. 163–166.

[102] S. Benatti et al., "A Versatile Embedded Platform for EMG Acquisition and Gesture Recognition," IEEE Transactions on Biomedical Circuits and Systems, vol. 9, no. 5, pp. 620–630, Oct. 2015.

[103] G. Franco, P. Cancian, L. Cerina, E. Besana, N. Beretta, and M. D. Santambrogio, "FPGA-based muscle synergy extraction for surface EMG gesture classification," in 2017 IEEE Biomedical Circuits and Systems Conference (BioCAS), 2017, pp. 1–4.

[104] E. Donati, M. Payvand, N. Risi, R. B. Krause, and G. Indiveri, "Discrimination of EMG Signals Using a Neuromorphic Implementation of a Spiking Neural Network," IEEE Transactions on Biomedical Circuits and Systems, pp. 1–1, 2019.

[105] Y. Lee, T. Chi, and C. Yang, "A 2.17mW Acoustic DSP Processor with CNN-FFT Accelerators for Intelligent Hearing Aided Devices,"

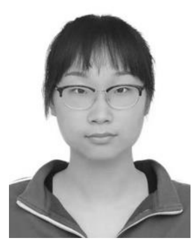
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBCAS.2020.2974154, IEEE Transactions on Biomedical Circuits and Systems

in 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), 2019, pp. 97–101.
[106] N. Attaran, A. Puranik, J. Brooks, and T. Mohsenin, “Embedded Low-Power Processor for Personalized Stress Detection,” IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 65, no. 12, pp. 2032–2036, Dec. 2018.
[107] A. R. Aslam and M. A. B. Altaf, "An 8 Channel Patient Specific Neuromorphic Processor for the Early Screening of Autistic Children through Emotion Detection," 2019 IEEE International Symposium on Circuits and Systems (ISCAS), Sapporo, Japan, 2019, pp. 1-5.
[108] Y. Huang, K. Wang, Y. Ho, C. He and W. Fang, "An Edge AI System-on-Chip Design with Customized Convolutional-Neural-Network Architecture for Real-time EEG-Based Affective Computing System," in 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS), 2019, pp. 1-4.
[109] W. Fang, K. Wang, N. Fahier, Y. Ho and Y. Huang, "Development and Validation of an EEG-Based Real-Time Emotion Recognition System Using Edge AI Computing Platform With Convolutional Neural Network System-on-Chip Design," IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 9, no. 4, pp. 645-657, Dec. 2019.

Ying Wei received the B. S and M. S. degrees from Xi’an Jiaotong University, China, in 2000 and 2003, respectively. She received the Ph.D. degree from National University of Singapore (NUS), Singapore, in 2008. She is currently working as a full Professor with the School of Control Science and Engineering, Shandong University, China. She is the director of International Cooperation Research Center of Intelligent Sensing and Information Processing in Shandong Province. She now serves as an associate editor of IEEE Transactions on Biomedical Circuits and Systems. Her research interests include implementation of high-speed digital systems and biomedical signal processing. In recent years she has taken charge of various projects including National Key Research and Development Project, National Natural Science Foundation, the Ministry of Science and Technology and the Ministry of Education.

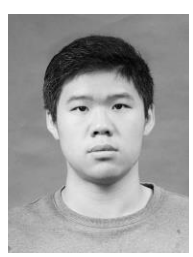
Jun Zhou (M’07–SM’14) received the dual B.S. degree in communication engineering and microelectronics from the University of Electronic Science and Technology of China in 2004, and the Ph.D. degree in microelectronics system design from Newcastle University, Newcastle upon Tyne, U.K., in 2008. He joined IMEC Netherlands in 2008 as a Research Scientist in the Ultra-Low Power Digital Signal Processor (ULPDSP) Group, where he worked on ultra-low power digital signal processor design in collaboration with Philips and NXP. In 2011, he joined the Institute of Microelectronics (IME), Agency for Science, Technology, and Research (A*STAR), Singapore, where he led projects and supervises Ph.D. students on energy-efficient digital signal processor design for intelligent sensing. In 2017, he joined the University of Electronic Science and Technology of China (UESTC) as a Professor and is currently leading the research group of Smart ICs and Systems for IoT. His major research interests is algorithm & processor co-design for intelligent sensing. He has published papers in prestigious conferences and journals, including ISSCC, JSSC, TBioCAS, TCAS-I, DAC and FCCM. He is serving as A-SSCC TPC sub-committee chair of Digital Circuits & Systems, Associate Editor of TBioCAS, Associate Editor of TVLSI, and Chair of Embedded AI Committee of Sichuan Institute of Electronics. He has also served as TPC/OC member for a number of IEEE conferences, including SoCC, ICCD, ISCAS, NEWCAS and EDSSC. He is a senior member of IEEE.

Yin Wang (S’19) received the B.S. degree in communication engineering from Nanjing University of Science and Technology, Nanjing, China, in 2017. She is currently pursuing master degree at the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China. Her research interests include area- and power-efficient neural network processor design. She is a student member of IEEE.



Yinggang Liu received the bachelor’s degree in electronic information engineering from Hunan Normal University, Changsha, China, in 2018. Then she started studying for the master’s degree in information and communication engineering from Shandong University. And now she has obtained the qualification of successive postgraduate and doctoral program, and will continue to work toward the Ph.D. degree in pattern recognition and intelligent system with Shandong University, Jinan, China. Her current research interests include audio signal processing and machine learning.

Qingsong Liu received the B.S. degree in electronic information engineering from the Sichuan University of China in 2019. He is currently pursuing master degree at the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China. His research interest is AI-based biomedical processor design.



Jiansheng Luo (S’19) received the B.Eng. degree in communication and information engineering from University of Electronic Science and Technology of China in 2018. He is currently pursuing master degree at the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China. His research interests include algorithm & processor co-design for intelligent sensing. He is a student member of IEEE.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBCAS.2020.2974154, IEEE Transactions on Biomedical Circuits and Systems

Chao Wang (M'10–SM'15) received his B.Eng. and Ph.D degrees, both in Electronics Engineering, from Huazhong University of Science and Technology (HUST), China in 2000, and Nanyang Technological University (NTU), Singapore in 2008, respectively. From 2005 to 2007, he was also with the Center for Signal Processing, NTU as a Research Engineer. Since then, he worked as an IC Design Engineer with STMicroelectronics (STM), Asia-Pacific Design Centre, Singapore, from 2008 to 2010. He participated in the development of STM Bayer/YUV CMOS image sensing and processing ICs/SoCs. From 2010 to 2017, he was a Research Scientist and a Project Leader with the Institute of Microelectronics (IME), Agency for Science, Technology and Research (A*STAR), Singapore, where he led projects on wearable medical devices for cardio-engineering applications, and MEMS Accelerometer ASICs for medical and navigation applications. Currently, he is a Professor with the Huazhong University of Science and Technology. His major research interests include energy-efficient digital signal processor design, ultra-low-voltage circuit design, MEMS sensor ASIC design, and heterogeneous 3D-IC design, especially for biomedical/healthcare, wireless sensor, IoT, Neuromorphic computing and robot applications. Dr. Wang is an IEEE Senior Member and currently the chair of IEEE CASS-EDS-SSCS Wuhan Joint Chapter. He had served as a committee member (2012-2018), the vice chair (2015-2016) and chair (2017-2018) of IEEE SSCS Singapore Chapter. He has also served as Technical Program Committee (TPC) members for a number of international devices, circuits and systems conferences. He used to serve as a Guest Editor for Journal Sensors, Hindawi in 2016 and IEEE Transaction on Biomedical Circuits and Systems in 2019. Dr. Wang is currently the Associate Editor of IEEE TCAS-I, IEEE CAS Magazine, Access Journal and Microelectronics Journal.

Fengbo Ren (S'10-M'15) received the B.Eng. degree from Zhejiang University, Hangzhou, China, in 2008 and the M.S. and Ph.D. degrees from the University of California, Los Angeles, in 2010 and 2014, respectively, all in electrical engineering. Since 2015, he has been a faculty of the School of Computing, Informatics, and Decision Systems Engineering at Arizona State University (ASU). His Ph.D. research involved in designing energy-efficient VLSI systems, accelerating compressive sensing signal reconstruction, and developing emerging memory technology. His current research interests are focused on hardware acceleration and parallel computing solutions for data analytics and information processing, with emphasis on bringing energy efficiency and signal intelligence into a wide spectrum of today's computing infrastructures, from data center server systems to wearable and Internet-of-things devices. He is a member of the Digital Signal Processing Technical Committee and VLSI Systems & Applications Technical Committee of the IEEE Circuits and Systems Society. Dr. Ren is a recipient of the Broadcom Fellowship in 2012, the prestigious National Science Foundation (NSF) Faculty Early Career Development (CAREER) Award in 2017, and the Google Faculty Research Award in 2018. He is also a recipient of the Top 5% Best Teacher Award from the Fulton Schools of Engineering at ASU from 2016 to 2019.

Li Huang received his B. Eng degree in Electronics and Information Engineering from Huazhong University of Science & Technology (HUST) and his Ph. D. degree in Electrical Engineering from both the National University of Singapore (NUS) and the Technische Universiteit Eindhoven (TU/e). After finishing his Ph. D. research activities, Dr. Huang joined Interuniversity Microelectronics Centre (imec) at Holst Centre in 2006, where he is now the senior business develop manager responsible for heading the iot business segment. He was also a visiting professor at Huazhong University of Science and Technology, and Shanghai Institute of Microsystem and Information Technology of Chinese Academy of Sciences. Before being the senior business develop manager, Dr. Huang led a R&D team developing system on chip (SoC) solutions for wireless body and personal area networks, supporting e.g. BLE and Zigbee standards. He has been involved in various industry and European projects on ULP design for different wireless communication systems. He has also been active in different wireless standards. He designed baseband algorithms for imec's ultra wideband (UWB) Integrated Circuit (IC) chip, which is known to be the first digital chip for the IEEE 802.15.4a standard. He also contributed several proposals for the IEEE 802.15.6 standard. He has published more than 100 papers/reports/books, one of which was among the top 10 of most downloaded papers in IEEE Transactions on Signal Processing in 2007. Besides, he is the founding president of the HUST alumni association in the Netherlands.