IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 24, 2019, accepted November 7, 2019, date of publication November 22, 2019, date of current version December 10, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2955285

Epileptic Seizure Prediction With Multi-View Convolutional Neural Networks

CHIEN-LIANG LIU[1], (Member, IEEE), BIN XIAO[2], WEN-HOAR HSAIO[3], AND VINCENT S. TSENG[2,4]

[1]Department of Industrial Engineering and Management, National Chiao Tung University, Hsinchu 30010, Taiwan
[2]Department of Computer Science, National Chiao Tung University, Hsinchu 30010, Taiwan
[3]Information Management Center, National Chung-Shan Institute of Science and Technology, Taoyuan City 32545, Taiwan
[4]Center for Emergent Functional Matter Science, National Chiao Tung University, Hsinchu 30010, Taiwan

Corresponding author: Chien-Liang Liu (clliu@mail.nctu.edu.tw)

This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant MOST 107-2221-E-009-109-MY2 and Grant MOST 107-2218-E-009-005.

ABSTRACT The unpredictability of seizures is often considered by patients to be the most problematic aspect of epilepsy, so this work aims to develop an accurate epilepsy seizure predictor, making it possible to enable devices to warn patients of impeding seizures. To develop a model for seizure prediction, most studies relied on Electroencephalograms (EEGs) to capture physiological measurements of epilepsy. This work uses the two domains of EEGs, including frequency domain and time domain, to provide two different views for the same data source. Subsequently, this work proposes a multi-view convolutional neural network framework to predict the occurrence of epilepsy seizures with the goal of acquiring a shared representation of time-domain and frequency-domain features. By conducting experiments on Kaggle data set, we demonstrated that the proposed method outperforms all methods listed in the Kaggle leader board. Additionally, our proposed model achieves average area under the curve (AUCs) of 0.82 and 0.89 on two subjects of CHB-MIT scalp EEG data set. This work serves as an effective paradigm for applying deep learning approaches to the crucial topic of risk prediction in health domains.

INDEX TERMS Electroencephalograms (EEG), seizure prediction, convolutional neural network (CNN), multi-view CNN, representation learning.

I. INTRODUCTION

Approximately 50 million individuals worldwide have been diagnosed as having epilepsy [8], but existing treatments such as surgery and anticonvulsants pose severe side effects to patients. Epilepsy, which is characterized by unpredictable seizures, adversely affects patient mental health, often resulting in anxiety, depression, or cognitive impairment [26]. Therefore, an accurate prediction model that provides an alert prior to the occurrence of seizure can considerably improve patient quality of life. Most studies have relied on electroencephalograms (EEGs) [5], [16], [39], which are recordings of the electrical potential of the brain, to capture physiological measurements of epilepsy. Note that the EEG signals could be presented in frequency domain and time domain. Thus, the two domains provide two different views for the same data source, in which the time domain measures how the signals change over time, while frequency domain describes signal energy distribution with respect to frequency components. Based on various recording methods, we can categorize EEG signals into two types, namely scalp EEG (sEEG) and intracranial EEG (iEEG). Scalp EEG signals are noninvasive and commonly collected with electrodes situated on the scalp, while iEEG signals are collected by placing electrodes on the exposed surface of the brain.

The last decade has witnessed the great success of machine learning as it could automatically learn a model from data to perform a specific task effectively without using explicit instructions. To develop a machine learning model, data should be characterized by feature vectors in machine learning, and they are critical to machine learning algorithms, since subsequent machine learning algorithms can benefit from the representations that can reveal important properties of the data. Consequently, many researchers have applied various signal-processing methods to extract features from EEG signals, such as wavelet analysis [18], [24], [36],

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

Fast Fourier Transform (FFT) [45], Independent Component Analysis (ICA) [43] and Principal Component Analysis (PCA) [43]. However, the EEG signals generally comprise several channels, so one can use a subset of channels based on some criteria or all the channels. When multiple channels are selected, correlations between channels are often considered to extract features. Various channel selection methods have been proposed to extract features over last decades [2].

A convolutional neural network (CNN) [30], [31] is a type of feed-forward neural network inspired by the structure of the animal visual cortex. One critical feature of a CNN is that it comprises multiple layers of small neuron collections, referred to as receptive fields, that process input data. Notably, receptive fields in CNN are called kernels or filters, and they are used interchangeably in this work. Deep CNNs offer the benefit of being able to acquire high-level representations of data, which are difficult to extract using shallow methods. A typical CNN architecture comprises one or more convolutional layers and then followed by one or more fully connected layers as used in a multi-layer neural network. To reduce the number of free parameters and improve generalization, a convolution operation on small regions of input is introduced, so CNNs are easier to train and have fewer parameters than fully connected networks with the same number of hidden units. Recently, CNNs have been successfully applied to various application domains, including computer vision [28], signal processing [1], [14] and natural language processing [25], [27]. Therefore, this work proposes to develop a prediction model based on CNN to predict epileptic seizures.

The input data of seizure prediction are EEGs, which are produced in signal format; thus, fast Fourier transform (FFT) can conventionally be used to convert EEG signals into representations in the frequency domain. An EEG can be represented in the time domain in addition to the frequency domain. The frequency and time domains can be regarded as two perspectives of EEG signals, in which the time domain measures the change in signals over time and the frequency domain describes signal energy distribution with respect to frequency components. This study extends research on CNNs to propose a multi-view CNN framework for simultaneously receiving inputs in the frequency and time domains. The goal of this research is to use the proposed multi-view CNN to acquire a shared representation of time-domain and frequency-domain features; the final layer can accurately predict seizures because of the excellent representations acquired from previous layers.

The contributions of this work are listed as follows. First, we propose a multi-view CNN framework for seizure prediction, in which time-domain and frequency-domain features are the inputs. Notably, the proposed model is a framework as it allows the practitioners to incorporate additional inputs into the model. Second, we conduct experiments on two data sets, and one of them is obtained from Kaggle[1] competition.

[1] Kaggle: http://www.kaggle.com

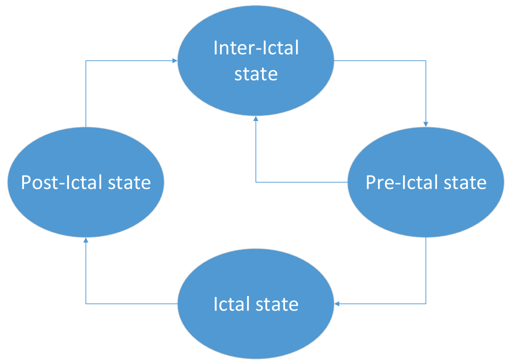

FIGURE 1. Epilepsy state flow.

The experimental results indicate that the private score of the proposed multi-view CNN outperforms all the methods listed in the leader board. Additionally, our proposed model achieves average area under the curve (AUCs) of 0.82 and 0.89 on two subjects of CHB-MIT scalp EEG data set. Notably, the proposed work could be applied to sEEG and iEEG, and yield better performances than state-of-the-art methods such as support vector machines (SVM) [12], [46]. Finally, we provide detailed discussion and analysis about the proposed method in this work.

The remainder of this paper is organized as follows. We briefly summarize the background in Section II. Section III presents the proposed multi-view CNN framework. Section IV shows the experimental results. We discuss the results and analyze the proposed method in Section V. Finally, Section VI draws the conclusions and possible future directions.

II. RELATED WORK

There is no denying that prediction of epileptic seizures before the occurrence of seizure is quite valuable for preventing the seizure and noxious treatment. Additionally, four phases are mainly involved in an epilepsy cycle as presented in Figure 1 [17], in which pre-ictal is the stage before the seizure, ictal is the stage during the seizure, post-ictal is the stage after the seizure and inter-ictal is the stage between two consecutive seizures. When patients are in ictal stage, they would have obviously physical symptoms, so one can easily distinguish this stage from other stages. Meanwhile, post-ictal stage is the stage where patients recovered from the seizure, and usually lasts several hours. When patients are in post-ictal stage, they would behave like normal people.

Notably, ictal and post-ictal stages are normally not considered in seizure prediction problem as they can be easily identified. As a result, the goal of seizure prediction is to classify inter-ictal stage and pre-ictal stage, because preemptive therapy intended to prevent the transition to the ictal stage may be possible in the pre-ictal stage. Thus, if one can correctly identify pre-ictal stage, the medical staffs can give appropriate treatment to prevent patients from ictal stage. In contrast, once the transition from pre-ictal to ictal phase has taken place, preemption is no longer an option.

Consequently, seizure prediction problem can be formulated as a binary classification problem, in which label zero denotes inter-ictal stage and label one represents pre-ictal stage. Note that this work focuses on seizure prediction rather than seizure detection that focuses on the detection of seizures. It is apparent that seizure prediction is more challenging and useful than seizure detection, since seizure prediction allows the patients or medical staffs to take appropriate actions to avoid the occurrence of seizures.

Most of previous research studies on seizure prediction or detection involve two basic steps. The first step is to extract features from the EEG signals, so that one can use feature vectors to represent EEG signals. One of the most popular methods to process EEG signals is FFT [3], which converts a signal from its original domain to a representation in the frequency domain. Besides, the EEG signals belong to time-series format, so one can slice EEG signals by a time window, usually ranges from 5 seconds to 60 seconds [6], [7], [41]. Once feature extraction stage is completed, the second step is to apply classification algorithms along with the obtained features to train a predictive model.

Several research methods for seizure prediction or detection have been devised over the last decades [6], [7], [41]. For example, Shoeb et al. proposed an approach to detect seizures based on FFT [41], in which they spanned the frequency range 0.5-25Hz with a filter bank and then grouped the frequency into eight equal bands. Then, they used SVM with RBF kernel as classifier to identify whether a state is inter-ictal state. Similarly, Bandarabadi also grouped the signal frequency into sub-bands and used a SVM with RBF kernel as classifier [6]. Moreover, they calculated power spatial density for every windowed signal and made a feature selection based on amplitude distribution. Brinkmann et al. explored the spatio-temporal relationship of EEG signals using phase correlation [7] as the features, and utilized an extension version of SVM named LS-SVM as a classifier. Mirowski et al. analyzed the bivariate features of EEG synchronization, and compared many combinations of features and classifiers in their work [3]. Li *et al.* [33] reported a spike based approach, which applied a morphological filter to detect EEG spikes and used smoothed spike rate in EEG at index for impending seizures.

Besides the methods mentioned above, wavelet analysis, ICA, PCA and many other signal processing methods are also widely used by researchers [3]. Carney *et al.* [8] discussed many other methods to process EEG signals, and illustrated various measures such as multivariate measures, correlation structure, phase correlation, and autoregressive measures of synchrony. Parvez and Paul [37] also reported various processing methods in the field of EEG seizure detection and prediction from the view of signal processing.

Applications of using deep learning techniques for processing EEG signals and seizure prediction have been very limited so far. Alotaiby *et al.* [3] applied CNN as a classifier after they extracted bivariate features of EEG for seizure prediction. Seizure prediction is an important research topic, so Kaggle held a competition for seizure prediction, since if these seizure-permissive brain states can be identified, devices designed to warn patients of impeding seizures would be possible. Brinkmann et al. made a good summary of the top 10 solutions to the Kaggle seizure prediction contest [44]. Among these solutions, a CNN based solution was proposed, which was ranked at 10th place, by using spectral power and standard deviation of signals as features. Note that the proposed method in this work is also based on CNN, but it is apparent that the two methods are different in architectures and input features. It is noted that the CNN can use raw data, such as images and texts, as the inputs; however, data preprocessing may help train an accurate and robust model in some application domains. Hosseini *et al.* [21] indicated that identifying an appropriate feature for the given problem may still be valuable because such a feature could enhance the performance of CNN-based algorithms. For example, Yuan *et al.* [47] used short-time Fourier transform (STFT) to represent time-frequency information as the input of their proposed deep-learning architecture as brain abnormality on EEGs is reflected by frequency changes and increased amplitudes [19].

Besides seizure prediction, CNN has been applied to brain related problems, such as brain-computer interface (BCI) [10] and emotion recognition [34]. In EEG signal processing, Cecotti [9] proposed a CNN architecture to classify EEG signals, in which FFT is added between two hidden layers. With this new design, signal can be switched from the time domain to the frequency domain inside the network.

III. LEARNING ALGORITHMS - MULTI-VIEW CONVOLUTIONAL NEURAL NETWORKS

This section introduces the proposed multi-view CNN framework. The proposed architecture is a framework that provides a foundation for incorporating various views or domains into the model, giving a base for the model to learn discriminative feature representations.

A. PROBLEM FORMULATION

Given a training set comprising m examples $(\mathbf{X}, \mathbf{y}) = \{(\mathbf{x}_i, y_i)\}$, where $i = 1, \cdots, m$, $\mathbf{x}_i \in \mathbb{R}^n$ is the ith training example and $y_i \in \{0, 1\}$ is the label of $\mathbf{x}_i$, and the goal is to find a mapping function or hypothesis h such that $h(\mathbf{x}_i) = y_i$. The seizure prediction problem is a binary classification problem, so this work uses cross entropy loss function as listed in Equation (1) in the classification model.

$$\mathcal{L}(\mathbf{y}, h(\mathbf{X})) = -\frac{1}{m}\sum_{i}^{m}[y_i \log h(\mathbf{x}_i) + (1-y_i)\log(1-h(\mathbf{x}_i))], \tag{1}$$

where $\mathbf{x}_i$ is the input, y_i is the desired output, and $h(\mathbf{x}_i)$ is the predicted output for $\mathbf{x}_i$.

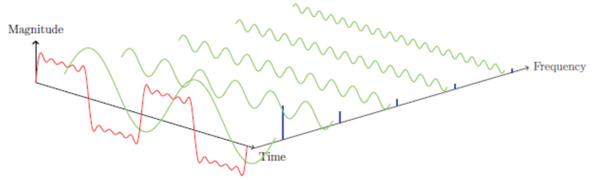

FIGURE 2. The relation between time domain and frequency domain.

B. PROPOSED MULTI-VIEW CNN ARCHITECTURE

Most studies have focused on establishing excellent data representations according to domain knowledge to enable the subsequent classifier to benefit from suitable representations. By contrast, the CNN uses a different approach by considering feature learning and classification in the same network and providing an elegant means of optimizing the parameters with error backpropagation.

The digital EEG signal is stored electronically, and FFT can be applied to convert the EEG signal from its original time domain to a representation in the frequency domain. Several relevant studies have used this approach to extract features from EEG signals [40], [45]. On the other hand, EEG signal is presented in time domain, and the correlation analysis could reveal important characteristic of the time-series data. Therefore, the EEG signal can be inspected from the frequency domain and analyzed from the time domain. For an identical EEG signal, the time domain and frequency domain can be regarded as two views of the same information source as presented in Figure 2.

Thus, this study proposes a mult-iview CNN framework to incorporate multiple views of EEG signals into the proposed architecture, as presented in Figure 3, in which time-domain and frequency-domain features are regarded as inputs. The architecture of the proposed framework involves convolutional layers, fully connected layers, and one softmax layer. Notably, the two inputs are combined with a fully connected layer, providing a foundation with which to map different views of the EEG signal into a shared view and automatically extract adequate feature representations.

In the experiments on intracranial EEG data set, we use five convolutional layers, and one fully connected layer in each stream following three fully connected layers after the features of two streams combined together as showed in Figure 3. The number of parameters in a typical CNN architecture is always enormous, so many techniques should be used to prevent model from overfitting. In this work, we use ℓ_2 regularizer and dropout [42] in convolutional layers. Additionally, we use Rectified Linear Unit (ReLU) [35] as the activation function for convolutional layers, and the ReLU function is presented in Equation (2).

$$f(x) = \max(0, x) \tag{2}$$

In the proposed multi-view CNN, the activation function for fully connected layer is different from that of convolutional layer, and the activation function for fully connected

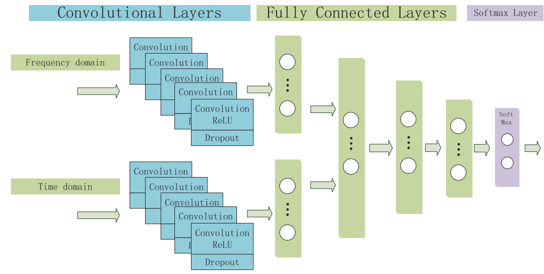

FIGURE 3. Multi-view CNN architecture.

layers is tanh as listed in Equation (3).

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{3}$$

Finally, we use a softmax function as presented in Equation (4) in the last layer, since the seizure prediction problem belongs to binary classification problem.

$$f_i(z) = \frac{e^{f_{y_i}}}{\sum_j e^{z_j}} \tag{4}$$

The proposed model follows the convention of CNN architecture to let the last layer act as a classifier. Besides the last layer, the layers preceding the last layer can be considered as feature learning algorithm, whose goal is to learn suitable representations for the last layer [23].

IV. EXPERIMENTAL RESULTS

To evaluate the effectiveness of the proposed multi-view CNN, we conduct experiments on two data sets, and the experimental results show that the proposed method outperforms all methods listed in the Kaggle leader board. Moreover, the experimental results for scalp EEG point out that our proposed method performs stably, and normally outperforms SVM methods.

A. DATA SETS

1) INTRACRANIAL EEG DATA SET

The American Epilepsy Society data set is an intracranial EEG data set comprising intracranial EEG signals for five dogs and two human patients. This data set was also used in a Kaggle contest [44], and training data are marked with two states, namely inter-ictal state and pre-ictal state. Pre-ictal data are the data from 1 hour before a seizure with a 5-minute seizure horizon. The purpose of the 5-minute gap is to enable medical staffs or patients to implement appropriate measures prior to seizure occurrence.

Inter-ictal data are data from canine measurements 1 week before or after a seizure. In the human data, inter-ictal data segments were restricted to be at least 4 hours before or after any seizure as the entire monitoring session may last less than one week. Test data are also available for model evaluation. Public scores in the leader board are calculated on a randomly sampled 40% subset of the test data segments, but official winners are determined based on the remaining 60% of the testing data. Notably, labels for the test data samples

TABLE 1. Summary of the American Epilepsy Society Data Set.

Data name	# of channels	# of pre-ictal segments	# of inter-ictal segments	# of test data	sampling rate
Dog_1	16	24	480	502	400
Dog_2	16	42	500	1000	400
Dog_3	16	72	1440	907	400
Dog_4	16	97	804	990	400
Dog_5	15	30	450	191	400
Patient_1	15	18	50	195	5000
Patient_2	24	18	42	150	5000

are unavailable, and prediction results must be submitted to Kaggle to obtain the outcomes, which only involve public and private scores. A summary of the data set is provided in Table 1.

2) SCALP EEG DATA SET

Scalp EEG signals are probably the most commonly used data source for seizure prediction because they can be obtained through noninvasive means by placing electrodes on the scalp. We used a data set from the CHB-MIT Scalp EEG Database [38] to evaluate the performance of our model.

The database includes data of 24 subjects, denoted by Chb01 through Chb24. Data are presented in European Data Format (EDF). Most EDF files contain EEG signals with a duration of 1 hour, but some are 4 hours long. Gaps of a few seconds exist between consecutive EDF files as a result of hardware limitations. We followed the same protocol used in Kaggle contest to preprocess the data. We used EEGLab [13] to slice data from 1 hour prior to a seizure with a 5-minute horizon as the pre-ictal data and 4 hours before and after seizures as the inter-ictal data. We also transformed these sliced data into 10-minute segments. A summary of the processed CHB-MIT scalp EEG data set is presented in Table 2.

B. DATA PREPROCESSING

In this work, we propose to pre-process the EEG data in the two domains, and then the two pre-processed data become the two inputs of our proposed multi-view CNN. In the frequency-domain stream, the input feature vector should be generated with frequency analysis tools, and we used FFT in the experiments. Conversely, the time-domain stream can be processed using many approaches, such as auto regression or correlation entropy. To verify the effectiveness of the data preprocessing, this work conducts experiments using the proposed model with raw EEG data as the input for comparison with the proposed approach.

Prior to feature extraction, we sliced the raw data with a time window of 30 seconds without overlap. PCA is applied to sliced data segments to reduce the dimension of the data and generate the feature vectors in the time domain. We used FFT to transform these sliced data segments into the frequency domain in addition to the time domain and categorized them into eight frequency bands, namely 0.1–4, 4–8, 8–12, 12–30, 30–50, 50–70, 70–100, and 100–180 Hz.

We used the mean value of the log amplitude in each band and their standard deviations as the feature values. Notably, the estimation is done for each individual channel. After completion of feature extraction of the time and frequency domains, we obtained feature vectors in both domains. Because slicing data without overlap may break the relations between consecutive segments in both the frequency and time domains, we used overlaps between consecutive 10-minute feature vectors in a 1-hour sequence following the data processed by PCA and FFT to obtain more training data for the relations between consecutive segments. Overlapping consecutive segments can also be considered a method of augmenting data. Data augmentation can typically help to produce more training data when the amount of data is limited and can prevent model overfitting [28].

C. ARCHITECTURE AND EXPERIMENTAL SETTINGS

The basic architecture is presented in Figure 3, which comprises five convolutional layers in each stream. Each convolutional layer in each stream is followed by a dropout layer, with the exception of the final convolutional layer. Convolutional layers are initialized with uniform distribution as described in [32]. A fully connected layer is added as the final layer in each stream, and both streams are then connected to each other by another fully connected layer. Next, two or more fully connected layers are supplemented with a softmax layer at the end because seizure prediction is a classification problem. Although these two streams do not necessarily possess the same architecture, we used the same architecture in our experiments.

Seizure prediction is a binary classification problem, explaining why we use binary cross entropy as the loss function. It is noted that binary cross entropy is one of the most commonly used loss functions in deep learning when dealing with binary classification problems. Besides, stochastic gradient descent is used to optimize the model. The activation function for convolutional layers is a rectified linear unit [35], whereas the activation function for fully connected layers is tanh. An ℓ_2 regularizer in addition to dropout layers is used to prevent model from overfitting. To enable the reproduction of these experimental results, we have shared the source code of the proposed multi-view CNN on GitHub.[2] Additionally, because of space limitations, detailed parameter settings and

[2]Source Code: https://github.com/buptxiaofeng/seizure-prediction

TABLE 2. Summary of the Processed CHB-MIT Scalp EEG Data Set.

Subject	Sampling Rate	# of Channels	# of Inter-ictal Segments	# of Pre-ictal Segments
Chb01	256Hz	23	84	28
Chb02	256Hz	23	72	6
Chb03	256Hz	23	150	6
Chb04	256Hz	24	664	4
Chb05	256Hz	23	84	24
Chb06	256Hz	23	77	12
Chb07	256Hz	23	279	6
Chb08	256Hz	23	0	24
Chb09	256Hz	23	238	6
Chb10	256Hz	23	132	6
Chb11	256Hz	23,28	0	0
Chb12	256Hz	28	0	0
Chb13	256Hz	22,25,28	0	0
Chb14	256Hz	28	18	18
Chb15	256Hz	31,38	0	54
Chb16	256Hz	22,28	30	18
Chb17	256Hz	22,28	0	0
Chb18	256Hz	22,28	138	12
Chb19	256Hz	22,28	132	0
Chb20	256Hz	28	95	0
Chb21	256Hz	28	126	6
Chb22	256Hz	28	96	12
Chb23	256Hz	23	0	0
Chb24	256Hz	23	0	0

the detailed architecture for Dog_1 based on the proposed multi-view CNN are available on GitHub.

D. EXPERIMENTAL RESULTS

We conducted experiments on two data sets. One of them is from American Epilepsy Society Seizure Prediction Challenge of Kaggle[3] and the other one is from CHB-MIT Scalp EEG Database of PhysioNet.[4]

1) INTRACRANIAL EEG

In the experiments on the Kaggle data set, we predominantly used the highest ranking solutions for comparison. The experimental results are presented in Table 3; the winner was determined according to the private score. The team ranked in first place did not open source their solution, and therefore their algorithm was unknown. The team in second place applied an ensemble learning approach to combine several state-of-the-art algorithms in the prediction model, including Lasso regularization of generalized linear models (GLM), random forest, and bagged support vector machine (SVM). In addition to ensemble learning, they used various features in their model. Both ensemble learning and feature engineering are crucial to performance and are therefore two prominent techniques in large-scale data science competitions. We provide the implementation proposed by Bandarabadi *et al.* [6] in addition to the solution listed by the Kaggle leader board.

Canonical correlation analysis (CCA) [22] is another technique that has been used for unsupervised data analysis when multiple views are available [15], [20], thereby providing a foundation for classifiers to learn features for multiple modalities. Andrew *et al.* [4] devised a deep learning algorithm termed deep canonical correlation analysis (DCCA) to learn the complex nonlinear transforms of two views of data and thereby producing highly correlated representations. The model proposed in this work also focuses on multi-view problems, providing an explanation for why we implemented DCCA when conducting a comparison with the proposed method. On the basis of the settings used in [4], we used a four-layer deep neural network with 200 units in the output layer; the inputs comprise the features presented in the time and frequency domains used in our proposed method. The hidden layer for the frequency domain has 2048 units, whereas the hidden layer for the time domain has 1024 units. Moreover, an RMSprop optimizer is used to reduce the amount of time required for convergence.

Because the DCCA emphasizes feature learning, we used an SVM as the final classifier. The experimental results indicated that the proposed method outperformed DCCA. The goal of DCCA is to maximize the total correlation among the parameters for different transformations, whereas that of the proposed method is to minimize cross-entropy loss and thereby update the parameters for the two input streams.

[3]American Epilepsy Society Seizure Prediction Challenge: https://www.kaggle.com/c/seizure-prediction

[4]CHB-MIT Scalp EEG Database: https://www.physionet.org/pn6/chbmit/

TABLE 3. Experimental Results for the Kaggle Competition (public and private scores are presented with AUC).

Approaches	Algorithm	Features	Ensemble	Public score	Private score
Proposed Method	Multi-view CNN	PCA , Mean log amplitude in frequency domain	None	0.83719	0.84227
1st place	Unknown	Unknown	Unknown	0.90316	0.83993
2nd place	Lasso GLM, Random Forest, Bagged SVM	Spectral entropy, correlation, fractal dimensions, Hjorth parameters, distribution statistics Spectral power, correlation, signal variance	weighted average of the form: (1/4 * Random Forest + 1/4 * Bagged SVM + 1/2 * Lasso GLM)	0.85951	0.81962
3rd place	SVM	Log spectral power, covariance	Platt scaling	0.83869	0.80079
10th place	CNN	Spectral power, signal standard deviation	Weighted average of rank scores	0.82455	0.78513
RSP+SVM [6]	Class Weighted SVM	relative spectral power	None	0.67931	0.67639
DCCA+SVM [4]	Linear SVM	PCA , Mean log amplitude in frequency domain	None	0.69560	0.65153

TABLE 4. Experimental results for scalp EEG.

Patient Name	Method	Feature	AUC	SS	SP
Chb01	Multi-view CNN	PCA and Mean log amplitude in frequency domain	0.82±0.05	0.93±0.07	0.71±0.13
	SVM-1 [6]	relative spectral power	0.83±0.06	0.97±0.01	0.69±0.11
	SVM-2	PCA and Mean log amplitude in frequency domain	0.61±0.07	0.94±0.01	0.27±0.14
Chb05	Multi-view CNN	PCA Mean log amplitude in frequency domain	0.89±0.04	0.90±0.02	0.88±0.10
	SVM-1 [6]	relative spectral power	0.69±0.07	0.97±0.01	0.41±0.10
	SVM-2	PCA and Mean log amplitude in frequency domain	0.66±0.08	0.88±0.07	0.45±0.16

Cross-entropy loss is highly correlated with the final performance measurement, which explains why the proposed method achieves better performance than DCCA.

2) SCALP EEG

Seizure prediction is characterized by imbalanced data, in which the volume of data in one class is overwhelmed by that in the other class. Accuracy is inappropriate for use as a performance evaluation criterion in an imbalanced data set, and receiver operating characteristic (ROC) curves have emerged as a popular alternative [11]. Reducing ROC performance to a single scalar value representing expected performance is recommended for comparing classifiers. The area under the curve (AUC) provides an alternative for evaluating classifier performance. Thus, regarding the scalp EEG experiments, this study compared the proposed method with two SVM methods using different features and used three evaluation metrics in the experiments: AUC, sensitivity (SS), and specificity (SP). The first SVM method, which was proposed by Bandarabadi *et al.* [6], is based on spectral power features, whereas the second SVM uses the same features as those used by the proposed multi-view CNN.

Although the scalp database involves 24 subjects, few of the subjects comprise sufficient inter-ictal and pre-ictal segments. Consequently, this study conducted experiments with data for Chb01 and Chb05. The proposed multi-view CNN with five convolutional layers is subject to the problem of overfitting because of an insufficient amount of training data in this data set for a complicated model. Therefore, we used a simple version of the multi-view CNN architecture that comprises only two convolutional layers. The experimental results are presented in Table 4. Although the number of training samples is limited in this data set, our simple version of the multi-view CNN architecture remains competitive for the scalp database. In contrast, SVM methods performs unstably in the experiments. Although SVM-1 performs well for Chb01, but the performance of AUC is poor for Chb02. We conjecture the possible reason is that the extracted features are not discriminative for the classifier to separate the data points well on the new space. As compared with deep learning that learns feature representations from data, SVMs use kernel functions to map data points to a high dimensional space, but the selection of kernel function is always influenced by the learning problems.

The limitation of this experiment is that we only used two subjects to evaluate performances as only the two subjects comprise sufficient segments for model training. Therefore, one of the future directions is to collect the data with more subjects to evaluate model performances. Moreover, we also conduct experiments to evaluate the time required for prediction in the proposed model. The proposed model is implemented with Keras using TensorFlow as the backend. The prediction time of the proposed model is approximately 1 second, which is acceptable for predicting seizures with a wearable device as the purpose of this work is to predict the occurrence of time point that is 5 minutes before the seizure onset. Details of the experiment are provided in the GitHub repository.

V. DISCUSSION

This section discusses the experimental results, and we also provide detailed analysis about our proposed method.

A. SEIZURE PREDICTION

Table 3 indicates that the proposed method outperforms other solutions in terms of the private score for the intracranial EEG data set. Notably, the public score is calculated based on a fraction of the test data set and displayed on the leader board during the competition; moreover, it is not the final score of the competition. After the competition has been completed, the final ranking is determined by the private score. One reason for using public and private scores is to ensure that participants do not overfit their models by only focusing on improving the public scores. Although the public score of our solution is not the best, the difference between the public score and private score is determined to be the smallest for our solution, indicating that it is not subject to an overfitting problem.

Ensemble learning has been commonly used in data science competitions because it entails the use of multiple learning algorithms to obtain favorable predictive performance. Ensembles tend to yield superior empirical results when significant diversity exists among the models [29], which explains why most of the solutions in Table 3 used the ensemble learning method. By contrast, the proposed method uses only one method, but it performs exceptionally well on this data set. One reason for this is that the proposed method considers feature learning with inputs from the time and frequency domains and can therefore learn a discriminative feature representation in the architecture.

Among the comparison methods listed in Table 3, the solution ranked in tenth place used a traditional CNN method. One key difference between the proposed multi-view CNN and traditional CNN other than architectural difference is that our method considers two views as inputs and establishes a shared representation to acquire discriminative representations. The private score indicates that the proposed architecture can significantly improve prediction performance.

In summary, the proposed method outperforms the other methods in the leader board. More importantly, the proposed approach is based on a single model, whereas most methods in the leader board rely on ensemble learning technique to combine several methods in the model. Deployment of a model with single method is expected to be much easier than that of a model with multiple methods as the practitioners need to train and tune multiple methods to obtain good model performance.

The experimental results on scalp EEG listed in Table 4 indicate that the method proposed by Bandarabadi *et al.* [6] exhibits excellent performance on the scalp EEG; however, it demonstrates unsatisfactory performance on the Kaggle data set, as indicated in Table 3. Regarding experiments on the CHB-MIT data set, the small number of test samples may have led to biased results, and further validation is required.

B. EFFECTIVENESS OF DATA PREPROCESSING

To verify the effectiveness of data preprocessing, we conduct another experiment using a CNN without feature extraction. The architecture is similar to that of the proposed model. We attempted to tune the model, but the AUC score is approximately 0.5, indicating poor performance.

In addition to the performance problem, much more time is required for training when no feature extraction is used because of the high feature dimension. The time series data always involve a long sequence, so the size of each data instance is large, making it difficult to train a model. For example, the sampling rate of Dog_1 is 400 Hz, and each segment is a 10-minute EEG signal with 16 channels. Thus, the size of the input matrix is 16×240000. By contrast, applying the feature extraction method described in the previous section to both the time domain and frequency domains produces two input matrices with sizes of only $16 \times 16 \times 17$ and $16 \times 9 \times 17$, respectively.

C. COMPARISON WITH SVM-BASED SOLUTIONS

SVMs are the dominant classifiers used in related research; therefore, we conduct experiments to compare SVM-based solutions with those of the proposed method. The first method of comparison concatenates the features of the time and frequency domains. The second method uses features in the time domain through PCA, as mentioned. The final method uses the features in the frequency domain as those in the frequency stream of the proposed multi-view CNN model.

The experimental results, presented in Figure 4, indicate that the proposed method outperforms the other three methods. A typical approach to processing signals is using FFT to obtain the features in the frequency domain; Figure 4 shows that frequency-domain features provide more discriminative information than time-domain features. More notably, the SVM fails to perform well with the concatenation of the two features. By contrast, the proposed multi-view CNN can benefit from the two views by regarding them as two inputs, which are then blended using a shared representation.

D. MULTI-VIEW VERSUS SINGLE-VIEW CNNS

Because this study proposes a multi-view approach to designing network architecture, we investigate whether the multi-view approach could improve prediction performance by comparing the proposed multi-view CNN with two

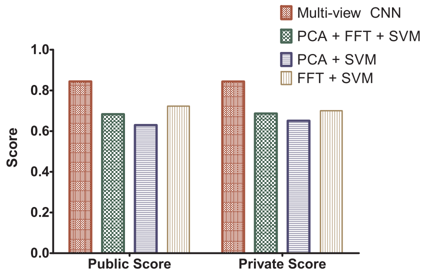

FIGURE 4. Comparison with SVM-based methods.

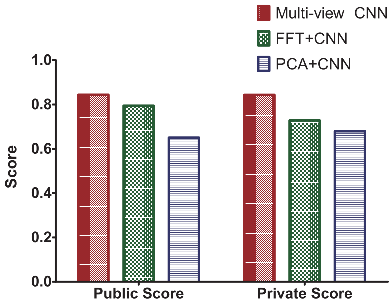

FIGURE 5. Performance comparison with single-view CNN and multi-view CNN.

single-view CNNs. To ensure objectivity of the experiments, the settings of the single-view CNNs are the same as those used for the the proposed multi-view CNN, except that only one information stream is used as the input. Figure 5 presents the experimental results, indicating that the proposed multi-view CNN significantly outperformed the two single-view CNNs. Notably, the single-view CNN with frequency-domain features outperformed that with time-domain features. This result is similar to the experimental results presented in Figure 4; that is, frequency-domain features provided more discriminative information than did time-domain features. Frequency-domain and time-domain features provide different views for the same EEG signal, and the experimental results indicate that using an appropriate method to process the features from the two views is necessary. Otherwise, the combination of the two features results in performance degradation, as presented in Figure 4.

In summary, the private scores in the Kaggle competition indicate that most methods require the use of ensemble learning to combine different algorithms in a model, whereas the proposed approach uses only a single model. It is expected that maintaining a single model is much easier than using a combination of algorithms when converting into real-world applications. Moreover, the flexibility of the proposed framework is demonstrated by its applicability to iEEG and scalp EEG signals.

VI. CONCLUSION AND FUTURE WORK

This work proposes a multi-view CNN framework to predict the occurrences of seizures. Central to the proposed framework is to consider time domain and frequency domain in two different inputs and use a shared layer to learn discriminative feature representations. The experimental results indicate that the proposed method outperforms the methods in the leader board of Kaggle competition. Although this work focuses on seizure prediction problem, EEG signals are widely used in the area of Brain Computer Interface and many other fields. Therefore, one of the future directions is to apply the proposed method to these areas. Moreover, we will collect the data with more subjects to evaluate model performances on scalp EEG. Another direction is to devise a new architecture to analyze Electrocardiography (ECG), which is the process of recording the electrical activity of the heart over a period of time using electrodes placed on the skin.

REFERENCES

[1] O. Abdel-Hamid, L. Deng, and D. Yu, "Exploring convolutional neural network structures and optimization techniques for speech recognition," *Interspeech*, vol. 11, pp. 3366–3370, Aug. 2013.
[2] T. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, and I. Ahmad, "A review of channel selection algorithms for EEG signal processing," *EURASIP J. Adv. Signal Process.*, vol. 2015, no. 1, p. 66, Dec. 2015.
[3] T. N. Alotaiby, S. A. Alshebeili, T. Alshawi, I. Ahmad, and F. E. A. El-Samie, "EEG seizure detection and prediction algorithms: A survey," *EURASIP J. Adv. Signal Process.*, vol. 2014, no. 1, p. 183, 2014.
[4] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1247–1255.
[5] E. B. Assi, L. Gagliano, S. Rihana, D. K. Nguyen, and M. Sawan, "Bispectrum features and multilayer perceptron classifier to enhance seizure prediction," *Sci. Rep.*, vol. 8, no. 1, 2018, Art. no. 15491.
[6] M. Bandarabadi, C. A. Teixeira, J. Rasekhi, and A. Dourado, "Epileptic seizure prediction using relative spectral power features," *Clin. Neurophysiol.*, vol. 126, no. 2, pp. 237–248, 2015.
[7] B. H. Brinkmann, J. Wagenaar, D. Abbot, P. Adkins, S. C. Bosshard, M. Chen, Q. M. Tieng, J. He, F. Muñoz-Almaraz, and P. Botella-Rocamora, "Crowdsourcing reproducible seizure forecasting in human and canine epilepsy," *Brain*, vol. 139, no. 6, pp. 1713–1722, 2016.
[8] P. R. Carney, S. Myers, and J. D. Geyer, "Seizure prediction: Methods," *Epilepsy Behav.*, vol. 22, pp. S94–S101, Dec. 2011.
[9] H. Cecotti, "A time–frequency convolutional neural network for the offline classification of steady-state visual evoked potential responses," *Pattern Recognit. Lett.*, vol. 32, no. 8, pp. 1145–1153, 2011.
[10] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 433–445, Mar. 2011.
[11] N. V. Chawla, "Data mining for imbalanced datasets: An overview," in *Data Mining and Knowledge Discovery Handbook*. Boston, MA, USA: Springer, 2005, pp. 853–867.
[12] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[13] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.
[14] L. Deng, O. Abdel-Hamid, and D. Yu, "A deep convolutional neural network using heterogeneous pooling for trading acoustic invariance with phonetic confusion," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6669–6673.
[15] P. Dhillon, D. P. Foster, and L. H. Ungar, "Multi-view learning of word embeddings via CCA," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 199–207.
[16] K. Edakawa, T. Yanagisawa, H. Kishima, R. Fukuma, S. Oshino, H. M. Khoo, M. Kobayashi, M. Tanaka, and T. Yoshimine, "Detection of epileptic seizures using phase–amplitude coupling in intracranial electroencephalography," *Sci. Rep.*, vol. 6, May 2016, Art. no. 25422.
[17] P. J. Franaszczuk, G. K. Bergey, P. J. Durka, and H. M. Eisenberg, "Time–frequency analysis using the matching pursuit algorithm applied to seizures originating from the mesial temporal lobe," *Electroencephalogr. Clin. Neurophysiol.*, vol. 106, no. 6, pp. 513–521, Jun. 1998.
[18] A. B. Geva and D. H. Kerem, "Forecasting generalized epileptic seizures from the EEG signal by wavelet analysis and dynamic unsupervised fuzzy clustering," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 10, pp. 1205–1216, Oct. 1998.
[19] J. Gotman, D. Flanagan, J. Zhang, and B. Rosenblatt, "Automatic seizure detection in the newborn: Methods and initial evaluation," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 3, pp. 356–362, Sep. 1997.

[20] D. R. Hardoon, J. Mourão-Miranda, M. Brammer, and J. Shawe-Taylor, ''Unsupervised analysis of fMRI data using kernel canonical correlation,'' *NeuroImage*, vol. 37, no. 4, pp. 1250–1259, Oct. 2007.
[21] S. Hosseini, S. H. Lee, and N. I. Cho, ''Feeding hand-crafted features for enhancing the performance of convolutional neural networks,'' Jan. 2018, *arXiv:1801.07848*. [Online]. Available: https://arxiv.org/abs/1801.07848
[22] H. Hotelling, ''Relations between two sets of variates,'' *Biometrika*, vol. 28, nos. 3–4, pp. 321–377, 1936.
[23] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[24] P. Jahankhani, V. Kodogiannis, and K. Revett, ''EEG signal classification using wavelet feature extraction and neural networks,'' in *Proc. IEEE John Vincent Atanasoff Int. Symp. Mod. Comput. (JVA)*, Oct. 2006, pp. 120–124.
[25] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, ''A convolutional neural network for modelling sentences,'' Apr. 2014, *arXiv:1404.2188*. [Online]. Available: https://arxiv.org/abs/1404.2188
[26] M. P. Kerr, ''The impact of epilepsy on patients' lives,'' *Acta Neurologica Scandinavica*, vol. 126, no. s194, pp. 1–9, 2012.
[27] Y. Kim, ''Convolutional neural networks for sentence classification,'' Aug. 2014, *arXiv:1408.5882*. [Online]. Available: https://arxiv.org/abs/1408.5882
[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ''ImageNet classification with deep convolutional neural networks,'' in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[29] L. I. Kuncheva and C. J. Whitaker, ''Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy,'' *Mach. Learn.*, vol. 51, no. 2, pp. 181–207, May 2003.
[30] Y. LeCun, Y. Bengio, and G. Hinton, ''Deep learning,'' *Nature*, vol. 521, pp. 436–444, May 2015.
[31] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, ''Gradient-based learning applied to document recognition,'' *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[32] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, ''Efficient BackProp,'' in *Neural Networks: Tricks of the Trade*. Berlin, Germany: Springer, 2012, pp. 9–48.
[33] S. Li, W. Zhou, Q. Yuan, and Y. Liu, ''Seizure prediction using spike rate of intracranial EEG,'' *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 6, pp. 880–886, Nov. 2013.
[34] Q. Mao, M. Dong, Z. Huang, and Y. Zhan, ''Learning salient features for speech emotion recognition using convolutional neural networks,'' *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2203–2213, Dec. 2014.
[35] V. Nair and G. E. Hinton, ''Rectified linear units improve restricted Boltzmann machines,'' in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, Jun. 2010, pp. 807–814.
[36] H. Ocak, ''Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy,'' *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2027–2036, 2009.
[37] M. Z. Parvez and M. Paul, ''Epileptic seizure prediction by exploiting spatiotemporal relationship of EEG signals using phase correlation,'' *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 1, pp. 158–168, Jan. 2016.
[38] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, ''PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals,'' *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
[39] J. Pyrzowski, M. Siemiński, A. Sarnowska, J. Jedrzejczak, and W. M. Nyka, ''Interval analysis of interictal EEG: Pathology of the alpha rhythm in focal epilepsy,'' *Sci. Rep.*, vol. 5, Nov. 2015, Art. no. 16230.
[40] A. H. Shoeb, ''Application of machine learning to epileptic seizure onset detection and treatment,'' Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 2009.
[41] A. Shoeb and J. Guttag, ''Application of machine learning to epileptic seizure detection,'' in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, Jun. 2010, pp. 975–982.
[42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, ''Dropout: A simple way to prevent neural networks from overfitting,'' *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[43] A. Subasi and M. Ismail Gursoy, ''EEG signal classification using PCA, ICA, LDA and support vector machines,'' *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8659–8666, Dec. 2010.
[44] *American Epilepsy Society Seizure Prediction Challenge*, Kaggle, San Francisco, CA, USA, 2014.
[45] T. N. Thieu and H.-J. Yang, ''Diagnosis of epilepsy in patients based on the classification of EEG signals using fast Fourier transform,'' in *Proc. Int. Conf. Ind., Eng. Other Appl. Appl. Intell. Syst.* Cham, Switzerland: Springer, 2015, pp. 493–500.
[46] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 1995.
[47] Y. Yuan, G. Xun, K. Jia, and A. Zhang, ''A multi-view deep learning method for epileptic seizure detection using short-time Fourier transform,'' in *Proc. 8th ACM Int. Conf. Bioinf., Comput. Biol., Health Inform.*, Aug. 2017, pp. 213–222.

CHIEN-LIANG LIU received the M.S. and Ph.D. degrees from the Department of Computer Science, National Chiao Tung University, Taiwan, in 2000 and 2005, respectively. He is currently an Associate Professor with the Department of Industrial Engineering and Management, National Chiao Tung University. His research interests include machine learning, data mining, deep learning, and big data analytics.

BIN XIAO received the M.S. degree in computer science from National Chiao Tung University, Taiwan, in 2016. He is currently an Engineer with Imsight Medical Technology, Co., Ltd., Shenzhen, China. His research interests include machine learning and data mining.

WEN-HOAR HSAIO received the M.S. and Ph.D. degrees from the Department of Computer Science, National Chiao Tung University, Taiwan, in 1996 and 2015, respectively. He is currently an Engineer with the National Chung-Shan Institute of Science and Technology, Taiwan. His research interests include information retrieval, data mining, and machine learning.

VINCENT S. TSENG received the Ph.D. degree with major in computer science from National Chiao Tung University, Taiwan, in 1997. From 1998 to 1999, he was a Postdoctoral Research Fellow with the Computer Science Division, University of California at Berkeley. He also acted as the Director for the Institute of Medical Informatics, National Cheng Kung University, from 2008 to 2011. He is currently a Distinguished Professor with the Department of Computer Science and the Director with the Institute of Data Science and Engineering, National Chiao Tung University. He has published more than 350 research articles in referred journals and conferences. He held and filed 15 patents. He has a wide variety of research interests covering data mining, big data, biomedical informatics, mobile, and web technologies. He has also served as Chairs/Program Committee Members for a number of premier international conferences related to data mining and intelligent computing, including KDD, ICDM, SDM, PAKDD, ICDE, CIKM, and IJCAI. He was a recipient of the 2014 K. T. Li Breakthrough Award and the 2015 Outstanding Research Award by Ministry of Science and Technology, Taiwan (only one recipient nationwide in Taiwan annually for each of these two awards). He has served as the Chair for the IEEE CIS Tainan Chapter, from 2013 to 2015, and the President of Taiwanese Association for Artificial Intelligence, from 2011 to 2012. He has been on the Editorial Board of a number of journals, including the IEEE Transactions on Knowledge and Data Engineering, *ACM Transactions on Knowledge Discovery from Data*, and the IEEE Journal of Biomedical and Health Informatics. In recent years, he has also been serving to oversee the architect on Big Data technologies and applications for the governmental and industrial units in Taiwan. By Google Scholar, his publications have been cited by more than 5500 times with H-Index 39.