IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ARTIFICIAL INTELLIGENCE AND COGNITIVE COMPUTING FOR COMMUNICATION AND NETWORK

Received December 31, 2018, accepted January 16, 2019, date of publication January 29, 2019, date of current version February 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2895688

Multilevel Weighted Feature Fusion Using Convolutional Neural Networks for EEG Motor Imagery Classification

SYED UMAR AMIN[1], MANSOUR ALSULAIMAN[1], GHULAM MUHAMMAD[1], MOHAMED A. BENCHERIF[1], AND M. SHAMIM HOSSAIN[2], (Senior Member, IEEE)

[1]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[2]Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

Corresponding authors: Ghulam Muhammad (ghulam@ksu.edu.sa) and M. Shamim Hossain (mshossain@ksu.edu.sa)

This work was supported by the Deanship of Scientific Research at King Saud University, Riyadh, Saudi Arabia, through the research group under Grant RG-1436-023.

ABSTRACT Deep learning methods, such as convolution neural networks (CNNs), have achieved remarkable success in computer vision tasks. Hence, an increasing trend in using deep learning for electroencephalograph (EEG) analysis is evident. Extracting relevant information from CNN features is one of the key reasons behind the success of the CNN-based deep learning models. Some CNN models use convolutional features from different CNN layers with good effect. However, extraction and fusion of multilevel convolutional features remain unexplored for EEG applications. Moreover, cognitive computing and artificial intelligence experience increasing applications in all fields. Cognitive process is based on understanding human brain cognition through signals, such as EEG. Hence, deep learning can aid in developing cognitive systems and related applications by improving EEG decoding. The classification and recognition of EEG have consistently been challenging due to its characteristics of dynamic time series data and low signal-to-noise ratio. However, the information hidden in different convolution layers can aid in improving feature discrimination capability. In this paper, we use the EEG motor imagery data to uncover the benefits of extracting and fusing multilevel convolutional features from different CNN layers, which are abstract representations of the input at various levels. Our proposed CNN model can learn robust spectral and temporal features from the raw EEG data. We demonstrate that such multilevel feature fusion outperforms the models that use features only from the last layer. Our results are better than the state of the art for EEG decoding and classification.

INDEX TERMS EEG motor imagery classification, deep learning, convolution neural network, multilevel feature fusion.

I. INTRODUCTION

Electroencephalograph (EEG) is an efficient and relatively inexpensive technique for analyzing and studying brain electrical activities by placing electrodes on the skull surface (scalp EEG recording) or from inside the skull (intracranial EEG recording) [1]. The scalp EEG is recorded with a noninvasive technique using multiple electrodes; it also has high temporal resolution, which makes it a feasible technique for analysis and monitoring of the changes in brain electrical activity. This data recording is massive and contains synchronous activities of neurons in different areas of human brain.

Motor imagery (MI) is brain potentials related to MI tasks. MI data contain EEG recordings when subjects are asked to simply imagine moving a body part and not actually move it. Some oscillatory activities in the sensorimotor cortex region of the brain are present as a result of these MI tasks [2]. Machine learning techniques are often applied to classify these oscillatory activities for recognition of MI tasks. Many brain-computer interface (BCI)-related studies commonly use MI tasks, such as imagining the movement of both hands and feet [4], [5]. In this study, we use EEG MI datasets.

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

Recently, cognitive computing and artificial intelligence are being infused in all automated systems and technologies [3]. Cognitive process allows systems to think similarly to a human brain without requiring any human assistance for operations. This cognitive ability is inspired by the human brain. Automated systems that can learn human cognition from inputs, such as EEG not only can improve cognitive technology but also increase the intelligence of other automated applications. Hence, the development of MI decoding leads to improved cognitive computing technology. Researchers use EEG for various applications [6]–[8], such as brain-controlled robots, diagnosis of medical conditions disability and epilepsy, driving automated cars, and controlling drones. Building automated systems based on the understanding and decoding of MI EEG allows transfer of human cognitive intelligence into machines, which makes them powerful. Therefore, improving EEG decoding and classification is important to the field of cognitive computing and artificial intelligence.

Conventional machine learning techniques that use handcrafted features have been used in many important EEG-related research areas to extract meaningful information from EEG data. Such techniques have been used in many BCI systems for communicating with patients affected by strokes, for treating patients with epilepsy [7], controlling robots [6]. Machine learning has also been used for the medical analysis and interpretation of EEG signals [8]. Despite various applications of conventional machine learning techniques in EEG, they have failed to achieve acceptable performance and accuracy on EEG data. However, promising results have been obtained with the use of deep learning techniques, which shows that automatically extracted features perform better than handcrafted ones. These deep learning techniques have been used in different fields, such as computer vision and speech recognition [9], [10]. Convolution neural network (CNN) can extract features that are spatially robust [11]. Other models (such as recurrent neural network (RNN), which is particularly useful in applications that have temporal sequences, including video and speech recognition) are also available [12]. Some other models, such as autoencoders, are excellent for unsupervised learning [13].

Researchers have started to apply and investigate the potential of various deep learning models for EEG signal analysis [14]. Limited application might be due to the relatively small size of most EEG datasets compared with image datasets, thereby making it difficult to train deep networks that have several thousand parameters. Deep networks have been most successful in applications that have large dataset sizes. However, some studies show that deep belief network (DBN) and CNN are also useful in learning good features from EEG and functional magnetic resonance imaging data that have comparatively smaller dataset size [15]. These studies demonstrate that deep neural networks with reduced size and parameters can be applied in EEG datasets that usually have a smaller size. In fields such as computer vision, pretrained deep learning models have shown good performance [11], especially when only few training data are available. Hence, using deep learning model that is pretrained on related EEG data can be beneficial. This technique can also overcome the problem of small dataset size. However, the performance and accuracy of such deep network techniques are incomparable to other fields, such as image and speech recognition. Therefore, a large void and a need for major enhancement with respect to many aspects of applying deep learning for EEG MI classification should be considered.

Deep learning models, particularly CNN, have been successful for 2D signals, such as images; however, EEG data are recorded from the scalp surface using a set of electrodes. Hence, such models are time series data that are dynamic in nature. Properties, such as low signal-to-noise ratio, make the application of deep learning for EEG data slightly difficult.

Few studies [16]–[22] have recently used intermediate features from CNN layers to improve classification accuracy. Each CNN layer contains some relevant features that represent important information at their respective level of abstraction of the input data. Information from low- and high-level features represent the local and global structures of the input, respectively. EEG signals have spatial and temporal structures; thus, the local and global features extracted from CNN layers can be fused together to form a robust classification model. This type of CNN model for integrating multilayer features has yet to be explored for EEG signals.

Therefore, in this study, we use a pretrained CNN as feature extractor and propose a model for multilayer feature extraction and fusion for EEG MI data. We attempt to construct a robust feature representation for EEG signals using information hidden in CNN layers. We also show experimentally that our framework improves the performance of conventional CNN models using only the last layer features.

The remainder of this paper is organized as follows. Section II presents the related studies about EEG MI classification methods and feature fusion techniques. Sections III presents the proposed model. Section IV discusses the experimental details and results. Finally, Section V presents the conclusions.

II. RELATED STUDY

In this section, we briefly review existing methods and models for EEG MI classification. We also discuss the multilevel feature integration methods based on CNNs.

A. EEG MI CLASSIFICATION

Many feature extraction and classification methods have been used to recognize MI tasks. Filter bank common spatial patterns (FBCSPs) [5] are popular among MI classification techniques that use handcrafted features. FBCSP has been the previous state of the art for MI task recognition and has achieved excellent results [4], [5]. Feature extraction and dimension reduction techniques, such as principal component analysis (PCA) and independent component analysis, are also well-known techniques used by many researchers to improve

MI task recognition performance [23]–[25]. Support vector machine (SVM), linear discriminant analysis, and other methods have been used as classifiers in many MI classification studies [26]–[28].

In one study [15], Plis *et al.* show that using multiple restricted Boltzmann machine in performing supervised training allows DBN to learn more complex features and perform better than other techniques. Many recent studies have attempted to utilize the advantages of CNN for brain signal interpretation [27]–[30]. DBN is used extensively in several EEG studies for various applications [31]–[33]. In one study, CNN and RNN have been combined to achieve good performance for EEG time series data [27], [28]. DBN and SVM have also been compared [31] for two MI classification problems, in which DBN has shown better performance. CNN was also used in [32], for MI task classification. RNN and CNN were utilized in [33], wherein multidimensional features were proposed to find cognitive events from MI EEG data. Deep learning models, such as autoencoders have also been used for emotion recognition from EEG signals [34]. Many studies have converted EEG signals into images and applied different deep learning models that are excellent at classifying images. A new type of combined features was proposed in [35] to preserve the spatial, spectral, and temporal structures of EEG data. EEG signal from each of the electrodes was used to estimate the power spectrum for three selected frequency bands. The 3D electrode locations were then mapped into 2D to form EEG images. In [30], EEG time series data were converted into 2D images using short-time Fourier transform. Spectral features from mu and beta frequency bands were used and 1D CNN [35] and a stacked autoencoder (SAE) [36] were used to achieve good performance for MI dataset.

The aforementioned studies have attempted to utilize the capabilities and advantages of deep learning for EEG classification; however, the performance of the models is incomparable with other fields, such as image and speech recognition. Therefore, research on designing and applying deep learning models for EEG MI classification is necessary.

CNN has been applied successfully in many fields and has achieved outstanding results in computer vision and speech processing [9], [11] and has the ability to extract spatial and temporal features through convolution process. It is composed of many convolutional layers [9], where initial layers learn low-level spatial features and the high-order layers extracts global high-level features in a progressive manner. Low-level features correspond to edges, boundaries, or simple properties of the object, whereas high-level ones are learnt in deep layers of CNN, including complex shapes and orientations of objects. CNN can learn features automatically from raw data, thereby making it well suited for end-to-end learning. However, it requires large amount of training data to learn good features.

CNNs have been successful for 2D signals, such images. However, given that EEG data are recorded from scalp surface using a set of electrodes, they are time series data that are dynamic in nature. EEG recordings are prone to noise from artifacts, such as eye blinking and muscle movement, which do not have task-related information [37]. Hence, extracting good features from EEG signals is complex and difficult. Existing CNN architectures do not adapt to the dynamic characteristics of EEG signals.

Different deep learning and machine learning architectures have attempted to utilize several types of inputs from frequency or time domain or both. Some of the deep learning studies have used raw EEG data for end-to-end learning approach. They have successfully extracted useful information from EEG signals; meanwhile, CNNs have the best performance compared with these approaches. Given that the best performance for EEG MI data has been achieved using CNN, we use deep CNN model as a feature extractor.

B. FEATURE FUSION IN CNN

Many studies in the image processing domain [16]–[18] have proven that the convolutional features extracted from different CNN layers have different abstraction of information pertaining to the object to be detected. High-level features can help recognize object classes, whereas low-level ones can help define boundaries for recognized object. Unfortunately, although many techniques have been proposed in literature for different domains, no definite methods or techniques for extracting and fusing multilevel convolutional features are available. In [19], multilayer feature maps were used to form multiresolution images for salient object detection. The method learned to integrate the feature maps for each resolution and detected salient objects using integrated features. The features from low-level layers were used to form edge-aware or low-resolution feature maps, which help predict boundaries. Hariharan *et al.* [20] proposed a technique based on hyper column, which combines convolutional features from middle layers and fully connected classification layers.

Bhattacharjee and Das [21] proposed a multi-stream CNN using multilevel feature aggregation for human action classification from videos. They combined features at multiple stages of the model to extract spatial and temporal features in local and global contexts. Spatial and temporal streams were used at local context to recognize actions, and the fused information obtained was fed to stacked deep LSTM networks to extract global context features.

A system based on pretrained CNN was proposed in [16] for multilevel and multiscale feature aggregation for music input tagging. Local audio features were extracted using multiple CNNs. Then, audio features were captured from all layers of the pretrained CNNs at different time scales. These features were fused to form the global audio features for the complete audio clip and later fed to the classifier layer for tagging. Another research [22] used multilayer CNN features with bag-of-features technique in obtaining additional discriminative features to improve image classification.

Multilevel CNN features were also used for multimodal biometric identification in [17], in which multimodal fusion was utilized by using several modality-specific CNNs.

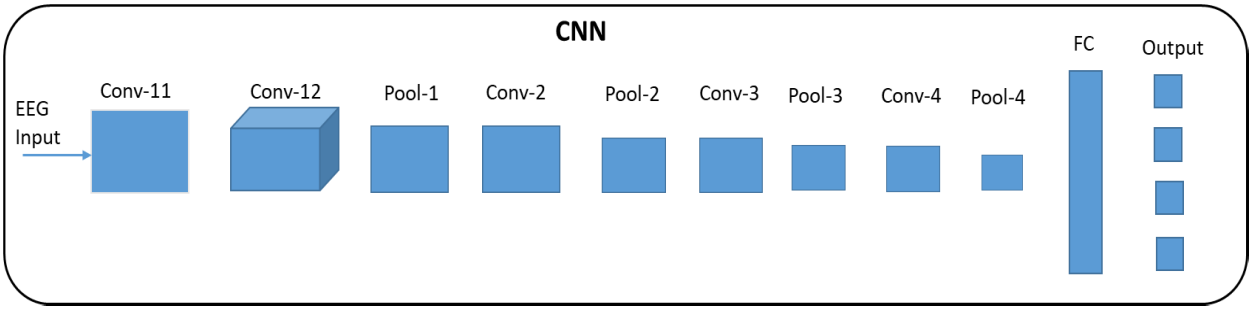

FIGURE 1. Deep CNN architecture.

The features were extracted at all CNN layers and were then compressed, fused together, and optimized for classification.

Another study [18] integrated multilayer CNN features for remote sensing scene classification to improve feature discrimination capability. Multilayer features were extracted from different convolutional and fully connected layers of pretrained CNN model. Fisher kernel coding was also applied to form a mid-level feature representation using convolutional features, which were then fused using PCA for classification.

All of the aforementioned approaches have obtained better performance in their specific domains by extracting and fusing multilevel CNN features. As previously discussed, EEG signals have dynamic nature due, with which it differs within the same subject and with other subjects. MI EEG signals also have highly subject-dependent characteristics and show dynamic behavior locally and globally for the same subject. These facts have encouraged us to utilize multilevel CNN feature extraction approach on EEG signal, which has never been applied before.

III. PROPOSED MULTILEVEL FEATURE FUSION MODEL

In this section, we describe the architecture of our proposed multilevel feature fusion model and provide detailed formulas of our information fusion learning method. Finally, we construct an EEG classifier on the basis of multilevel predictions of the proposed model.

Our proposed model consists of four components, namely, pretraining and transfer learning, multilevel feature extraction, weight-based feature fusion, and EEG MI classification. Figure 1 presents the model architecture.

The layers of the CNN model are based on popular CNN architecture in computer vision known as AlexNet [9]. This model has some blocks of convolutional and max pooling layers and some fully connected layers at the end. EEG signals are a multiple channel [32]; thus, we split the first convolution layer into two. Therefore, the first convolution is performed over time samples for each electrode, and the second is conducted on all electrodes or channels. The EEG input is stored as 2D array that has time across channels. Figure 2 shows the first convolution layer. Thereafter, we have the max pooling layer; the second, third, and fourth convolution max pooling block; and the dense softmax layer.

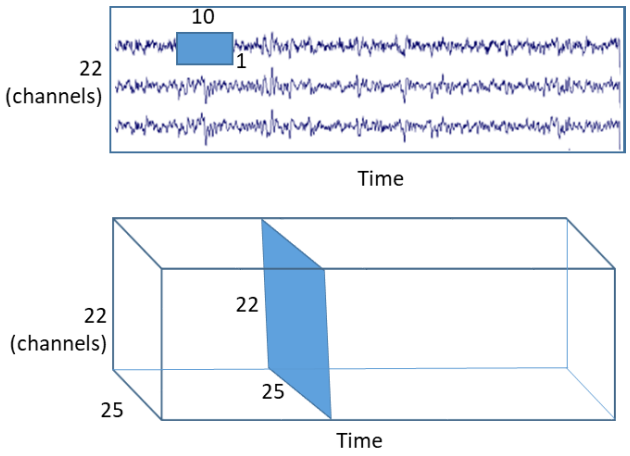

FIGURE 2. First convolution split into two parts (The first across time, and the second across all electrodes).

Different types of training strategies, normalization techniques, and activation functions have been experimented upon. We make use of the best strategies available from recent advances in deep learning and machine learning research on CNN. We use novel regularization strategies, namely, dropout and batch normalization. Exponential linear units prove to be fast and more accurate than rectified linear units.

Training is performed in batches, and batch normalization technique is used to improve the performance. Batch normalization is also used for convolutional outputs. To further increase the accuracy and reduce the chances of overfitting, we use the dropout technique because it aids in obtaining generalized results.

In this study, our deep learning model is evaluated on BCI Competition IV dataset 2a [38]. We compare the results of our study with those of the current state-of-the-art deep learning models in this field.

Some researchers have used electrode voltage over the flattened scalp surface and converted the EEG signal to topographical time series images [29]. However, research has proven that EEG signals are correlated over time series data [39]. Hence, in this study, we use raw EEG data as input to CNN, which can extract spatial and temporal features.

The EEG signal representation as 2D array that has time steps and channels also helps in EEG reducing data dimensionality.

We use pretrained CNN models because EEG BCI datasets for MI are usually small. Cropped input, pretraining, and transfer learning aid us to achieve good performance even on small training data and reduce the training time.

We have limited EEG MI data to train our deep CNN model; thus, we pretrain our model on an MI dataset called the High Gamma dataset [40]. This dataset is a huge MI dataset consisting of 128 electrode recordings from 20 subjects, having a total of 880 and 160 trials in the training and test sets, respectively. The four MI classes for movements are both hands and feet and rest. In comparison with the BCI dataset, High Gamma dataset has more training data; thus, by pretraining our CNN model on this dataset, we can eventually train our model on the BCI dataset without any fear for overtraining or memorization due to small size of the BCI dataset.

The CNN model has excellent capability to extract hierarchical structure of the input data. CNN represents high-order features as a set of low-order features by extracting spatial characteristics of the input signal.

A. CNN TRAINING

We divide the dataset, such that we randomly select eight subject recordings for training and one for subject-specific testing. In this manner, the system is tested without seeing the subject beforehand and is a completely new test case for the system. The testing performed in this manner is challenging, and the results are generalized.

These sets are retained until complete training and testing are over, after which we randomly select another training and testing sets. Finally, we calculate the results by averaging the values obtained in all phases. We use the supervised learning strategy, in which the model maps each input sample to the output classes. The softmax layer produces probability score for the target classes. We use mini-batch stochastic gradient descent to optimize the parameters for the network by using a backpropagation algorithm as a supervised learning algorithm [36]. The softmax classification function uses the output from the feature extraction function, such that the CNN network can optimize both functions simultaneously. In this manner, we can extract important features, remove noise artifacts from the EEG data, and solve the problem of overfitting to noise.

We use the BCI Competition IV dataset 2a (BCI dataset), which is an EEG MI dataset that has 22 scalp electrode positions. Nine subjects are involved in the recordings completed over two sessions. Each of the two sessions contain 288 trials, each having a 4 s recording of MI tasks related to the imagining of movements per subject. The imagined tasks consist of thinking to move both hands and feet and the tongue [38].

We use sliding windows on the input EEG signals; this cropped signal increases the training data to many folds [40]. Cropped training strategy helps because most of the EEG BCI datasets are not large, and CNN requires a huge amount of training data to generalize. This cropped training technique is successfully used for object recognition. Using this training technique increases the performance for EEG decoding.

In this study, we use a 2 s input window crop to generate a large number of training data. For each time step, we have one crop of each input sample in EEG data, which creates new training examples from the original set, thereby increasing the training set. Each window crop will use the same output label that we have for that entire event. Hence, our CNN uses features from all window segments of the event and can learn global features extracted from the entire event. This approach helps our model to learn general and not subject-specific features.

The BCI dataset has been recorded with a sampling rate of 256 Hz, thus, each 2 s input window has approximately 500 recorded input samples, which are given as input to the CNN model. The first convolution is performed over time samples for each channel. Table 1 presents the number, size of filters, and stride. The second convolution operation is performed over all 22 channels simultaneously. The rest of the convolution and max pooling operations are performed across all channels.

TABLE 1. Structure of CNN model.

Layers	Type
Conv-11	Convolution (10×1, 25 filters)
Conv-12	Convolution (1×22, 25 filters)
Pool-1	Max pooling (3×1, stride 3)
Conv-2	Convolution (10×1, 50 filters)
Pool-2	Max pooling (3×1, stride 3)
Conv-3	Convolution (10×1, 100 filters)
Pool-3	Max pooling (3×1, stride 3)
Conv-4	Convolution (10×1, 200 filters)
Pool-4	Max pooling (3×1, stride 3)
FC	Fully connected (1024)
Classifier	Softmax (4 classes)

The layers and filters are shown in Table 1. Completing training and testing for 90 epochs for each subject on the cross-subject data have taken us approximately 5 h to complete. However, completing subject-specific training has taken only 2 h because the number of iterations is less than the cross-subject approach.

B. MULTILEVEL FEATURE EXTRACTION

As previously described, the issue with EEG data is that they are dynamic in nature and have low signal-to-noise ratio; hence, extracting features manually and automatically has not led to high accuracy results even when using successful

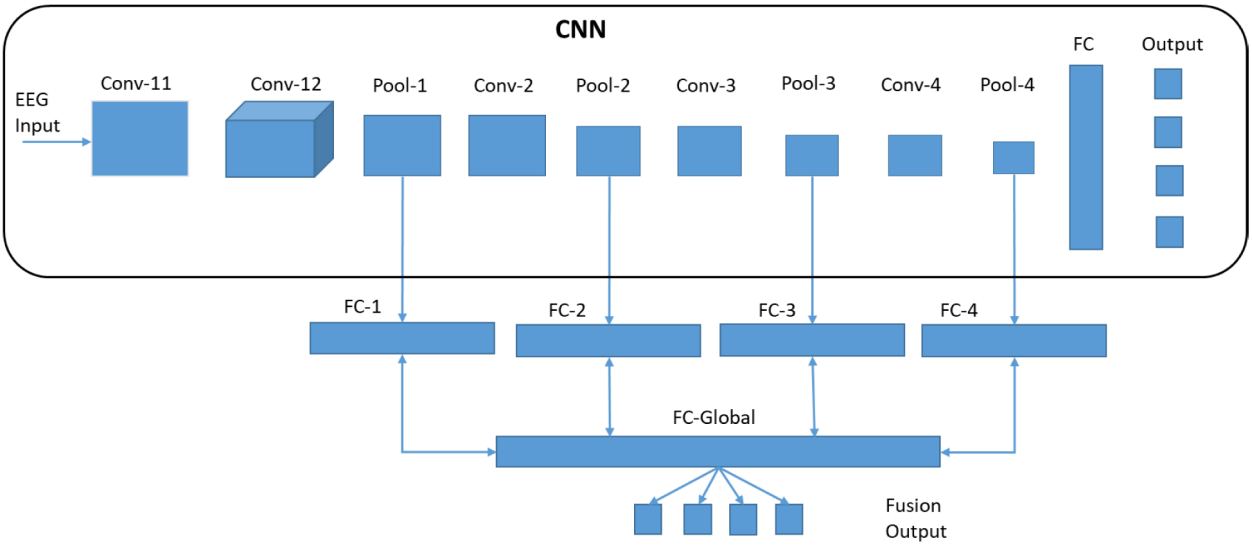

FIGURE 3. Multilayer feature fusion architecture.

models, such as CNN. Thus, we propose the extraction and fusion of multilayer features from a trained CNN model. By combining these features, we aim to uncover domain-specific knowledge and class-discriminative features that CNN layers have extracted at various levels.

Most of the studies have fused the networks at the fully connected layers or the softmax layers, but in this study feature representations from all convolution layers are fused using fully connected layers thereby preserving features at each convolution level. We extract these features after pooling layers to have compressed features without losing meaningful and important information. Instead of the simple concatenation of convolution features the dedicated fully connected layers have variable neurons to adjust the importance of features extracted at various levels. The CNN model is first pretrained before performing feature extraction. The model is then trained on BCI dataset and then the fully connected layers are added for feature extraction. The parameters and weights for trained model are frozen and the fusion model is trained with the extracted features acting as input to the fully connected layers FC-1, FC-2, FC-3 and FC-4. We extract trained features from the pooling layers which appear right after the convolution layers. Pooled features are compact form of convolution features and thus we have redundant features removed. This would also help in reducing the parameters as training the fully connected layers add to the computation overhead. At last we add another fully connected layer (FC-Global) for feature concatenation. By using dedicated fully connected layers we are able to extract local convolution features which represent the object at the abstract level and the global level features are extracted by applying concatenation through the FC-global layer. Hence the features from the layers FC-1, FC-2, FC-3, and FC-4 are concatenated before being supplied to the softmax classification layer. The fusion model is trained and tested with different lengths of pooling features and the best features are stored. The whole fusion model is then jointly-optimized.

TABLE 2. Structure of fusion model.

Input	Pool-1	Pool-2	Pool-3	Pool-4
Local fusion layers (size between 64 and 1024)	FC-1	FC-2	FC-3	FC-4
Global fusion layer	FC-Global (1024)			
Classifier	Softmax (4 classes)			

We use weight-based feature fusion to find the best size of extracted features for fusion. This feature extraction model is flexible and can be modified in various ways to include or remove different layers. Figure 3 shows the multi-level feature fusion architecture. The features extracted after the pooling operations have different sizes, hence we used fully connected layers to fuse these features.

C. WEIGHT-BASED FEATURE FUSION

We propose weight-based feature fusion that uses fully connected layers for fusing all extracted convolutional features. EEG channels are inputted across time samples, and the output after convolutions is a 2D feature map. We execute the multilayer feature extraction and fusion phase after the execution of the feature learning phase. The first phase of our deep CNN-based feature extraction model takes the raw cropped input EEG signal and produces feature maps for convolutional feature fusion. We use features from the convolution layers after the max pooling to reduce the feature size without losing relevant information extracted by the convolutional layers. Hence, the feature maps are extracted from Pool-1, Pool-2, Pool-3, Pool-4 (max pooling) layers.

TABLE 3. Comparison of some methods for the BCI dataset.

Methods	[5]	[30]	[33]	[40]	[41]	CNN (pre-trained)	Multi-fusion (proposed method)
Accuracy	68.0%	70.0%	69.0%	72.0%	73.4%	73.2%	**74.5%**

Each of these pooling layers represents convolution layers In the CNN architecture at different abstract level where initial layers represent simpler features and deeper layers represent complex features. Hence this architecture attempts to use different resolutions and abstractions of features at different layers.

In order to fuse the extracted convolution features, the output from each pooling layer is connected to the additional fully connected layers, each of which has a size of 64, 128, 256, 512, or 1024. All these sizes are power of 2, which was convenient as the final fully connected layer (FC-global) is of size 1024. The size of the additional fully connected layer is decided, when the fusion network is being optimized at runtime. Any layer whose features contribute less to the output, the corresponding fully connected layer is shrunk in size; moreover, the more important its features, the larger the size. The system starts with each of the fully connected layers having 1024 size and then gradually decreases the size of each of them trying all possible combinations within the predetermined values. Finally, the best possible values for the respective resolutions of FC layers are stored. Hence, the features are weighted on the basis of the best accuracy obtained by the fusion model. Using this weighted feature fusion we give variable importance to the convolutional features at different levels. Some of the researchers like [40] have achieved different accuracy with different number of convolution layers. Therefore we tried CNN with different number of convolutional layers and got different results, showing that convolution layers extract features which have different properties and importance in achieving the results. Hence weighted feature fusion addresses this issue and helps us find optimized features from different CNN layers. After we obtain the weighted CNN features, we fuse them using a global fully connected layer with the size of 1024. Subsequently, we use the softmax layer to classify the input into respective classes. The pretrained model is trained and optimized. After this the all the weights of the pretrained model are frozen. We then train the additional fully connected layers of the fusion model. Reduced learning rate is used for training the fusion model. Greedy optimization is performed with different size of features from all the pooling layers. The set of features giving the best results are retained at the end of the learning process.

TABLE 4. Performance of CNN models with different number of convolution layers (trained from scratch).

No. of CNN layers	1	2	3	4
Accuracy	72.7%	68.2%	69.5%	71.9%

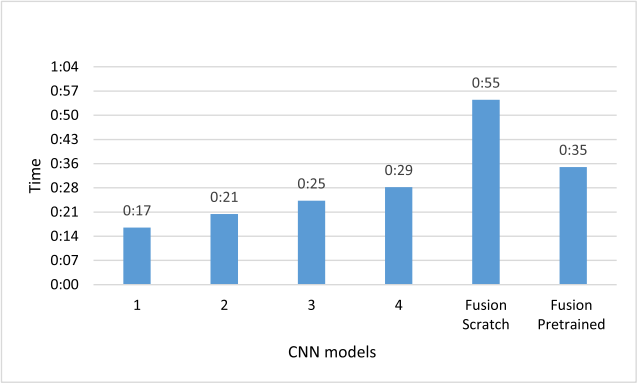

FIGURE 4. Mean training time (h:mm) across subjects for CNN models.

IV. EXPERIMENTS AND RESULTS

For experiments, we use Intel Xeon E5-2650 2.60 Hz CPUs with 17 cores and 64 GB RAM. For deep learning, we use GeForce GTX 1080 GPU with 8 GB memory. CNN was implemented using PyTorch deep learning framework and MNE-Python for EEG data preprocessing.

As previously discussed, we use cropped training strategy to test the performance of the CNN model. For the testing phase, the results for all the input crops are averaged to determine a single output per trial.

The results are better than those of the state of the art for EEG MI classification on the same dataset. Table 3 shows the comparison of our proposed model with the previous state-of-the-art approaches. In [40], CNN was used with cropped training and various selections to achieve 72.0% accuracy on BCI Competition IV dataset 2a MI dataset. Researchers in [30] used SAE over 1D CNN to obtain 69% accuracy on the same MI dataset. In [33], a compact CNN-based model named EEGNet was proposed for multiple EEG datasets, which reached 69% accuracy. FBCSP method and CNN were

TABLE 5. Performance of fusion model at different fusion stages (pretrained).

Fusion Stages	FC-1	FC-2	FC-3	FC-4	FC-3 + FC-4	FC-2 + FC-3 + FC-4	All
Accuracy	67.7%	66.2%	70.5%	73.2%	72.5%	73.7%	**74.5%**

TABLE 6. Confusion matrix for the fusion mode.

	Left hand	Right hand	Feet	Rest	Precision
Left hand	**452**	89	30	29	75.30%
Right hand	89	**440**	45	26	73.30%
Feet	28	31	**436**	105	72.60%
Rest	29	15	96	**460**	76.60%
Sensitivity	75.50%	76.50%	71.80%	74.10%	**74.50%**

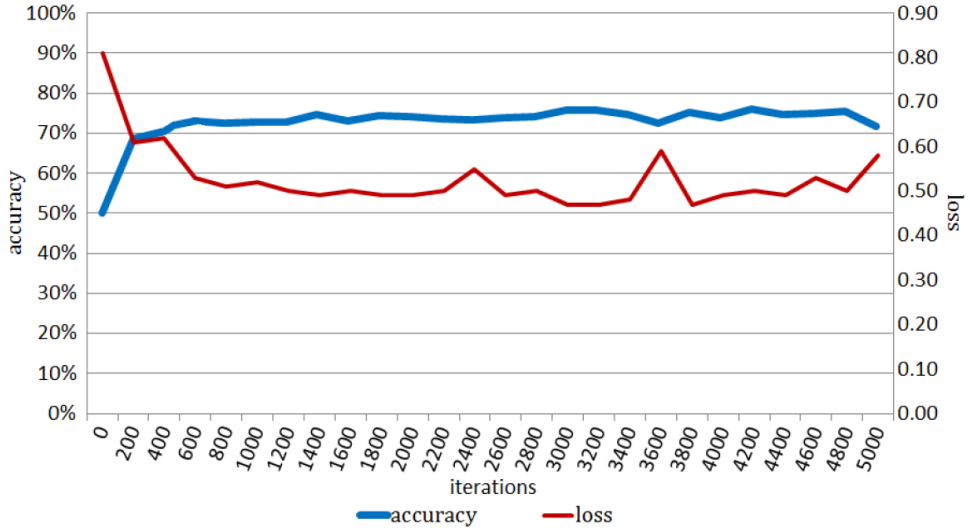

FIGURE 5. Performance of fusion model on BCI dataset.

used in [41] as a novel strategy. FBCSP algorithm was used for data preparation and create envelope representation input EEG signals to preserve the temporal and spatial properties of the signal. This dataset achieved 73.4% accuracy. Our proposed method presents an improvement in the result when pretrained CNN model is used compared with the training from scratch. The accuracy of the CNN model is 69.0%. Applying pretrained model provides us 73.2% accuracy. We achieve improvement in performance when we apply the proposed multilayer fusion model on our pretrained CNN model, obtaining 75.4% accuracy on the BCI dataset in comparison with the state-of-the-art models. The sensitivity of the system is also improved; thus, the multilayer CNN fusion architecture can be used as a subject-specific MI classifier. We also test the system with cross-subject data and achieve good specificity.

Table 4 shows the accuracies obtained when training CNN models from scratch with different number of convolution layers. We used CNN models starting with one convolution pooling block till models having four convolution pooling blocks. Beyond four convolution layers the performance degraded substantially. We also tried the fusion model by considering each FC layer at a time, and with various possible combinations of the FC layers. In table 5 we show the accuracy obtained training each FC layer and some of their combinations which gave good result. However, the best

result was 74.5% which was achieved by fusing all FC layers. Table 6 presents the confusion matrix for the fusion model which shows the overall accuracy, sensitivity and precision. As we can see, the sensitivity of the model for hand imagery movements is better than other classes. The performance enhancement of the fusion model comes with degradation of the training time, as shown in figure 4. It gives the average training time per subject for different CNN models. The fusion model trained from scratch took the most time. The accuracy and loss curve for the fusion model are shown in figure 5.

Although the preliminary results are encouraging, we still need to overcome some issues. Deep learning requires a considerable amount of training data, which is currently lacking. We also need to acquire additional datasets to increase the training set size. Most of the publicly available dataset have limited size because the larger ones are expensive. We also need to test pretrained models because they have yet to be used for EEG classification.

V. CONCLUSION

We propose a multilevel weighted feature fusion architecture based on CNN for EEG MI classification. Our method proves that different CNN layers can extract some abstract representations of the features. When these extracted features are fused, the resulting combined features can improve the overall classification accuracy. The study also shows that deep CNNs when pretrained with similar EEG data can aid CNN to learn small-sized datasets. Our fused framework outperforms state-of-the-art models on subject-specific EEG motor classification. The fused model can learn a general representation of EEG signals; hence our system has remarkable improvement in classification results. We achieve good accuracy for subject-specific data, with better sensitivity when compared to other methods on the BCI dataset.

The architecture proposed is designed to extract spectral, temporal features from EEG motor data while learning general spatially invariant characteristics of MI tasks. The multilayer feature fusion methods based on CNN have yet to be tested on other EEG datasets. Therefore, our method can also be used for other EEG applications to improve the results.

However, we would still want to investigate the system's performance on other EEG MI datasets that have more data. In the future, we also aim to study other ways of feature aggregation to enhance the performance of our proposed method.

REFERENCES

[1] L. J. Greenfield, Jr., J. D. Geyer, and P. R. Carney, *Reading EEGs: A Practical Approach*. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2012.

[2] G. Pfurtscheller and F. H. L. da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, pp. 1842–1857, Nov. 1999.

[3] Y. Zhang, X. Ma, S. Wan, H. Abbas, and M. Guizani, "CrossRec: Cross-domain recommendations based on social big data and cognitive computing," *Mobile Netw. Appl.*, vol. 23, no. 6, pp. 1610–1623, 2018.

[4] M. Grosse-Wentrup and M. Buss, "Multiclass common spatial patterns and information theoretic feature extraction," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1991–2000, Aug. 2008.

[5] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers Neurosci.*, vol. 6, p. 39, Mar. 2012.

[6] L. Tonin, T. Carlson, R. Leeb, and J. del R. Millán, "Brain-controlled telepresence robot by motor-disabled people," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug./Sep. 2011, pp. 4227–4230.

[7] M. S. Hossain, S. U. Amin, G. Muhammad, and M. Alsulaiman, "Applying deep learning for epilepsy seizure detection and brain mapping visualization," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 15, no. 1s, Jan. 2019, Art. no. 10.

[8] K. Das, B. Giesbrecht, and M. P. Eckstein, "Predicting variations of perceptual performance across individuals from neural activity using pattern classifiers," *NeuroImage*, vol. 51, no. 4, pp. 1425–1437, 2012.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates, Inc., vol. 25, 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[10] M. S. Hossain and G. Muhammad, "Emotion recognition using deep learning approach from audio–visual emotional big data," *Inf. Fusion*, vol. 49, pp. 69–78, Sep. 2019.

[11] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016, doi: 10.1109/TMI.2016.2528162.

[12] M. S. Hossain and G. Muhammad, "Environment classification for urban big data using deep learning," *IEEE Commun. Mag.*, vol. 56, no. 11, pp. 44–50, Nov. 2018, doi: 10.1109/MCOM.2018.1700577.

[13] M. Alhussein, G. Muhammad, M. S. Hossain, and S. U. Amin, "Cognitive IoT-cloud integration for smart healthcare: Case study for epileptic seizure detection and monitoring," *Mobile Netw. Appl.*, vol. 23, pp. 1624–1635, Dec. 2018.

[14] G. Muhammad, M. Masud, S. U. Amin, R. Alrobaea, and M. F. Alhamid, "Automatic seizure detection in a mobile multimedia framework," *IEEE Access*, vol. 6, pp. 45372–45383, 2018.

[15] S. M. Plis *et al.*, "Deep learning for neuroimaging: A validation study," *Frontiers Neurosci.*, vol. 8, p. 229, Aug. 2014, doi: 10.3389/fnins.2014.00229.

[16] J. Lee and J. Nam, "Multi-level and multi-scale feature aggregation using pretrained convolutional neural networks for music auto-tagging," *IEEE Signal Process. Lett.*, vol. 24, no. 8, pp. 1208–1212, Aug. 2017, doi: 10.1109/LSP.2017.2713830.

[17] S. Soleymani, A. Dabouei, H. Kazemi, J. Dawson, and N. M. Nasrabadi. (2018) "Multi-level feature abstraction from convolutional neural networks for multimodal biometric identification." [Online]. Available: https://arxiv.org/abs/1807.01332

[18] E. Li, J. Xia, P. Du, C. Lin, and A. Samat, "Integrating multilayer features of convolutional neural networks for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5653–5665, Oct. 2017.

[19] P. Zhang, D. Wang, H. Lu, H. Wang, and X. Ruan, "Amulet: Aggregating multi-level convolutional features for salient object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2017, pp. 202–211.

[20] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik, "Hypercolumns for object segmentation and fine-grained localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2015, pp. 447–456.

[21] P. Bhattacharjee and S. Das, "Two-stream convolutional network with multi-level feature fusion for categorization of human action from videos," in *Pattern Recognition and Machine Intelligence* (Lecture Notes in Computer Science), vol. 10597. Cham, Switzerland: Springer, 2017.

[22] K. Ueki and T. Kobayashi, "Multi-layer feature extractions for image classification—Knowledge from deep CNNs," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, London, U.K., Sep. 2015, pp. 9–12.

[23] Y. Qian, Y. Zhang, X. Ma, H. Yu, and L. Peng, "EARS: Emotion-aware recommender system based on hybrid information fusion," *Inf. Fusion*, vol. 46, pp. 141–146, Mar. 2019.

[24] M. T. F. Talukdar, S. K. Sakib, N. S. Pathan, and S. A. Fattah, "Motor imagery EEG signal classification scheme based on autoregressive reflection coefficients," in *Proc. Int. Conf. Inform., Electron. Vis. (ICIEV)*, Dhaka, Bangladesh, May 2014, pp. 1–4.

[25] N. F. Ince, S. Arica, and A. Tewfik, "Classification of single trial motor imagery EEG recordings with subject adapted non-dyadic arbitrary time-frequency tilings," *J. Neural Eng.*, vol. 3, no. 3, pp. 235–244, 2006.
[26] A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller, "Characterization of four-class motor imagery EEG data for the BCI-competition," *J. Neural Eng.*, vol. 2, pp. L14–L22, Aug. 2005.
[27] 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, doi: 10.1109/TPAMI.2010.125.
[28] N. F. Güler, E. D. Übeyli, and I. Güler, "Recurrent neural networks employing Lyapunov exponents for EEG signals classification," *Expert Syst. Appl.*, vol. 29, no. 3, pp. 506–514, 2005, doi: 10.1016/j.eswa.2005.04.011.
[29] P. Thodoroff, J. Pineau, and A. Lim, "Learning robust features using deep learning for automatic seizure detection," in *Proc. Mach. Learn. Healthcare Conf.*, 2016, pp. 178–190.
[30] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, no. 1, Feb. 2017, Art. no. 016003.
[31] X. An, D. Kuang, X. Guo, Y. Zhao, and L. He, "A deep learning method for classification of EEG data based on motor imagery," in *Intelligent Computing in Bioinformatics*. Berlin, Germany: Springer, 2014, pp. 203–210.
[32] H. Yang, S. Sakhavi, K. K. Ang, and C. Guan, "On the use of convolutional neural networks and augmented CSP features for multi-class motor imagery of EEG signals classification," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 2620–2623.
[33] V. Lawhern, A. Solon, N. Waytowich, S. M. Gordon, C. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Jun. 2018, Art. no. 056013.
[34] Y. Zhang, Y. Qian, D. Wu, M. S. Hossain, A. Ghoneim, and M. Chen, "Emotion-aware multimedia system security," *IEEE Trans. Multimedia*, to be published, doi: 10.1109/TMM.2018.2882744.
[35] 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.
[36] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 153–160.
[37] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, Jan. 2012.
[38] R. C. Leeb, G. Brunner, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI competition 2008-Graz data set A and B," Inst. Knowl. Discovery, Lab. Brain-Comput. Interfaces, Graz Univ. Technol., Graz, Austria, Tech. Rep. 1–6, 2008, pp. 136–142.
[39] R. T. Canolty *et al.*, "High gamma power is phase-locked to theta oscillations in human neocortex," *Science*, vol. 313, no. 5793, pp. 1626–1628, Sep. 2006.
[40] R. T. Schirrmeister *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapp.*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017, doi: 10.1002/hbm.23730.2017.
[41] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5619–5629, Nov. 2018.

SYED UMAR AMIN received the master's degree in computer engineering from Integral University, India, in 2013. He is currently pursuing the Ph.D. degree with the Department of Computer Engineering, College of Computer and Information Sciences, King Saud University. His research interests include deep learning, biologically inspired artificial intelligence, and data mining in healthcare.

MANSOUR ALSULAIMAN received the Ph.D. degree from Iowa State University, Ames, IA, USA, in 1987. Since 1988, he has been with the Computer Engineering Department, King Saud University, Riyadh, Saudi Arabia, where he is currently a Professor with the Department of Computer Engineering. He is also the Director of the Center of Smart Robotics Research, King Saud University. His research areas include automatic speech/speaker recognition, automatic voice pathology assessment systems, computer-aided pronunciation training system, and robotics. He was the Editor-in-Chief of the *King Saud University Journal Computer and Information Systems*.

GHULAM MUHAMMAD received the B.S. degree in computer science and engineering from the Bangladesh University of Engineering and Technology, in 1997, and the M.S. degree and the Ph.D. degree in electrical and computer engineering from the Toyohashi University and Technology, Japan, in 2003 and 2006, respectively. He is currently a Professor with the Department of Computer Engineering, College of Computer and Information Sciences, King Saud University (KSU), Riyadh, Saudi Arabia. His research interests include image and speech processing, cloud and multimedia for healthcare, serious games, resource provisioning for big data processing on media clouds, and biologically inspired approach for multimedia and software system. He has authored and co-authored more than 200 publications, including the IEEE/ACM/Springer/Elsevier journals, and flagship conference papers. He holds a U.S. patent on audio processing. He has supervised over 10 Ph.D. and master's theses. He is involved in many research projects as a Principal Investigator and a Co-Principal Investigator. He was a recipient of the Japan Society for Promotion and Science Fellowship from the Ministry of Education, Culture, Sports, Science and Technology, Japan. He received the Best Faculty Award from the Computer Engineering Department, KSU, from 2014 to 2015.

MOHAMED A. BENCHERIF received the Ph.D. degree in electronics from the University of Blida, Algeria, in 2015. He is currently a Researcher with the Center for Smart Robotics Research, King Saud University (KSU), Riyadh, Saudi Arabia. He is also teaching courses with the Department of Computer Engineering, KSU. His research interests include robot design and implementation, FPGA and DSP algorithm implementation, speech processing, and cryptography.

M. SHAMIM HOSSAIN (SM'09) received the Ph.D. degree in electrical and computer engineering from the University of Ottawa, Canada. He is currently a Professor with the Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. He is also an Adjunct Professor with the School of Electrical Engineering and Computer Science, University of Ottawa. He has authored and co-authored approximately 200 publications in refereed journals and conferences and, books, and book chapters. His research interests include cloud networking, smart environment (smart city, smart health), social media, the IoT, edge computing and multimedia for health care, deep learning approach to multimedia processing, and multimedia big data. He is a Senior Member of the ACM. He has served as a member of the organizing and technical committees of several international conferences and workshops. He was a recipient of a number of awards, including the Best Conference Paper Award, the 2016 *ACM Transactions on Multimedia Computing, Communications and Applications* Nicolas D. Georganas Best Paper Award, and the Research in Excellence Award from the College of Computer and Information Sciences, King Saud University (three times in a row). He has served as the Co-Chair, General Chair, Workshop Chair, Publication Chair, and TPC for over 12 IEEE and ACM conferences and workshops. He is currently the Co-Chair of the second IEEE ICME Workshop on Multimedia Services and Tools for Smart-Health (MUST-SH 2019). He is on the Editorial Boards of the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE NETWORK, the IEEE MULTIMEDIA, the IEEE WIRELESS COMMUNICATIONS, the IEEE ACCESS, the *Journal of Network and Computer Applications* (Elsevier), the *Computers and Electrical Engineering* (Elsevier), the *Human-centric Computing and Information Sciences* (Springer), the *Games for Health Journal*, and the *International Journal of Multimedia Tools and Applications* (Springer). He also serves as a Lead Guest Editor for the IEEE NETWORK, the *Future Generation Computer Systems* (Elsevier), and the IEEE ACCESS. Previously, he served as a Guest Editor for the *IEEE Communications Magazine*, the IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE (currently JBHI), the IEEE TRANSACTIONS ON CLOUD COMPUTING, the *International Journal of Multimedia Tools and Applications* (Springer), the *Cluster Computing* (Springer), the *Future Generation Computer Systems* (Elsevier), the *Computers and Electrical Engineering* (Elsevier), the *Sensors* (MDPI), and the *International Journal of Distributed Sensor Networks*.