Received May 3, 2019, accepted May 10, 2019, date of publication May 30, 2019, date of current version June 21, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2919995

Healthcare IoT-Based Affective State Mining Using a Deep Convolutional Neural Network

MD. GOLAM RABIUL ALAM[1,2], (Member, IEEE), SARDER FAKHRUL ABEDIN[1], (Student Member, IEEE), SEUNG IL MOON[1], ASHIS TALUKDER[1], (Member, IEEE), AND CHOONG SEON HONG[1], (Senior Member, IEEE)

[1]Department of Computer Science and Engineering, Kyung Hee University, Seoul 17104, South Korea
[2]Department of Computer Science and Engineering, BRAC University, Dhaka 1212, Bangladesh

Corresponding author: Choong Seon Hong (cshong@khu.ac.kr)

This work was supported in part by the Ministry of Science and ICT (MSIT), South Korea, under the ICT Consilience Creative Program supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP) under Grant IITP-2019-2015-0-00742, and in part by the IITP Grant funded by the MSIT (Resilient/Fault-Tolerant Autonomic Networking Based on Physicality, Relationship, and Service Semantic of IoT Devices) under Grant 2015-0-00557.

ABSTRACT Human effects are complex phenomena, which are studied for pervasive healthcare and well-being. The legacy pen and paper-based affective state determination methods are limited in their scientific explanation of causes and effects. Therefore, due to advances in intelligence technology, researchers are trying to apply some advanced artificial intelligence (AI) methods to realize individuals' affective states. To recognize, realize, and predict a human's affective state, domain experts have studied facial expressions, speeches, social posts, neuroimages, and physiological signals. However, with the advancement of the Internet of Medical Things (IoMT) and wearable computing technology, on-body non-invasive medical sensor observations are an effective source for studying users' effects or emotions. Therefore, this paper proposes an IoMT-based emotion recognition system for affective state mining. Human psychophysiological observations are collected through electromyography (EMG), electro-dermal activity (EDA), and electrocardiogram (ECG) medical sensors and analyzed through a deep convolutional neural network (CNN) to determine the covert affective state. According to Russell's circumplex model of effects, the five basic emotional states, i.e., happy, relaxed, disgust, sad, and neutral, are considered for affective state mining. An experimental study is performed, and a benchmark dataset is used to analyze the performance of the proposed method. The higher classification accuracy of the primary affective states has justified the performance of the proposed method.

INDEX TERMS Affective computing, convolutional neural network, emotion recognition, healthcare IoT, Internet of Medical Things.

I. INTRODUCTION

Human affects are the key factors of motivation, influence, strength, and empathy. Therefore, recognizing human affect is necessary for mental health care and well-being. Moreover, the emotionomics is a promising research domain in marketing and business, acknowledging that the emotional brain processes human perception five times faster than the cognitive brain [1]. Thus, a consumer's feelings drive him/her further than cognitive thinking to buy a given product. Therefore, product designs, shopping mall decorations, model selection for advertisements, and black-Friday sales offer are all the outcomes of consumer affect or emotion studies. After all, science applied to emotion is needed for mental healthcare, business studies, recommender systems, affective computing, social networks, emotion communication [2] and emotion-aware robot design.

Affective computing deals with the constructs of psychophysiology and generalizes human emotions as affective states [3]. In psychology, emotion is a conscious experience that can be characterized by functional mental activity and by the degree of contentment and discontentment.

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

Therefore, affective states are inherently thought of as non-scientific and affective disorders are assessed through a psychological questionnaire, e.g., the Korean Mood Disorder Questionnaire (K-MDQ) [4]. According to neuroimaging research, the brain of a human consists approximately 100 billion neurons and glial cells, forming a well-structured communication network among themselves [5]. As emotion is characterized by functional mental activity, determining the true emotional state is very challenging. However, with advances in technology, scientific methodologies and tools are successfully being applied in emotion recognition [6], [7], activity recognition and monitoring [8], [9], stress measurement [10], depression assessment [11], [12], and mental healthcare [13]–[15]. This research applies signal processing [5], big-data management [16], Internet of Medical Things (IoMT) [17] and advanced machine learning [18] methodologies for mining the affective states of individuals.

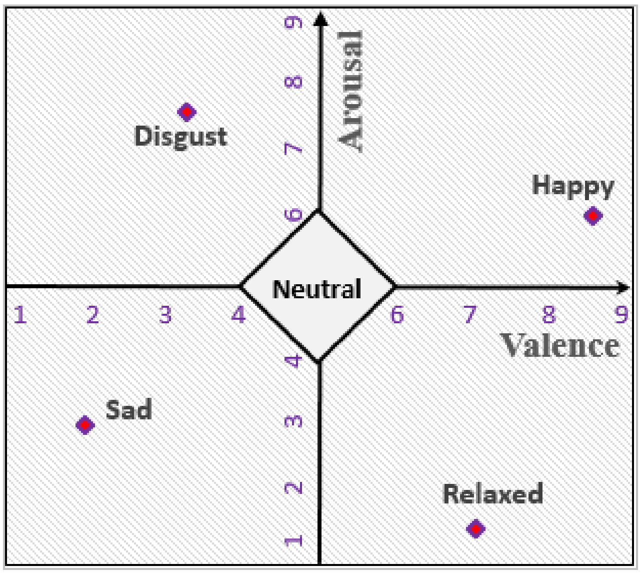

FIGURE 1. The five basic affective states in Russell's emotion circumplex.

The Geneva Emotion Wheel [19] represents 40 emotions, where valence and arousal levels are considered as the axes of 2D representation. However, it is difficult to classify each of these 40 emotions in the natural sciences, medical sciences and engineering perspectives and therefore only four to six basic emotions are studied in the facial expression recognition and image processing domains. Similarly, according to Russell's circumplex [20] of emotion, this paper studied five basic emotions such as Happy, Sad, Relaxed, Disgusted, and Neutral, as shown in Fig. 1.

Most of the existing studies detect affective states by employing facial expression [21]–[23], mood recognition by cognitive analysis [24], speech pattern mining based emotion recognition [25], sentiment analysis from social network posts [26], and psychophysiological observation-based affective mood classification [27]. However, facial expressions can't express internal affective states, speech analysis is not feasible for continuous emotion analysis, individuals' are not posting on the social network at every moment, and people may hide information in cases of the questionnaire-based affective state assessment. In contrast, an internal affective state can be mined through medical sensor-based physiological observations. Moreover, psychophysiological observation based affective mood classification methods are also feasible for continuous and real-time monitoring of affective states. Therefore, this research focused on the analysis of arousal, and valence [28] levels in physiological signals to mine the hidden affective states or emotions. Here, IoMT-based [29] physiological sensors such as Electro-Dermal Activity (EDA), Electromyography (EMG), and Electrocardiogram (ECG) are applied to extract the levels of valence, and arousal of a user. However, the key challenge is mapping such medical sensor observations to user emotions, i.e., determining the biosensor signal patterns for each of the affective states. Therefore, the deep Convolutional Neural Network (CNN) model is used for mining the affective state from large volumes of medical sensor observations.

The major contributions of this research are 1) an IoMT-based Affective State Mining (IASM) framework, 2) a testbed of real-time emotion recognition, 3) a performance study on DEAP dataset:

1) Affective state mining framework: Convolutional neural network (CNN)-based affective state mining framework is presented for real-time emotion recognition. The IoMT-based framework is composed of a Hadoop Distributed File System (HDFS)-based storage module, biosignal processing and feature extraction module, and a distributed CNN-based affective state miming module. Unlike legacy CNNs, in the proposed novel CNN architecture, the extracted features of the signal processing unit and high-level features from the last pooling layer are fused and fed into the fully connected layer to extract global discriminative features. Therefore, this affective state mining framework enables a platform of distributed processing and mining of massive data streams generated by the on-body medical sensors of a user.
2) Performance study on the DEAP dataset: The proposed CNN based affective mood mining approach is applied to the benchmark DEAP (Database of Emotion Analysis through Physiological Signals) dataset to study the performance of the proposed approach. The higher accuracy in the classification of valence and arousal justifies the performance gain of the proposed affective state mining approach.
3) Real-time emotion recognition test-bed: A test bed of real-time emotion recognition is developed via LuaJIT programming and the Torch scientific computing library. BioSignalPlux sensors and API are used for collecting medical-grade sensor data from users. The trained model of the proposed CNN is applied for online emotion classification.

The remaining sections of this paper are organized as follows: Section II reviews the related works. Section III discusses the details of the IoMT-based affective state mining framework. A comprehensive discussion of the proposed emotion classification methodology and performance study using DEAP dataset are stated in Section IV. The prototype test-bed implementation of real-time emotion recognition and performance evaluation are presented in section V. Section VI concludes the paper along with future directions.

II. RELATED WORKS

According to the survey presented in [30], human affects has been studied by ancient Greek philosophers. In this ancient theory of affects, emotions are described as the causes of a change of mind. Beyond this, the theory of human emotions has been vastly studied in social psychology. Psychologists introduced self-assessment-based emotion measurement scales and defined different emotions as affective states. The Self-Assessment Manikin (SAM) [31] is one of the pioneering picture oriented affective state measurement scales for measuring arousal, pleasure, and dominance induced by certain stimuli. However, there are some impediments in self-report and questionnaire-based affective state measurement models, e.g., off-line and abstract methods, where subjects can hide necessary information, provide misleading information, or not remember information. In a self-report, subjects may feel cautious in disclosing personal traits, or the questions may be misapprehended or misinterpreted. Moreover, it is also difficult to judge the significance of a solicited qualitative answer to a question in self-report-based methods of affective state determination. On the other hand, the biomedical sensor-based emotion recognition can overcome the limitations mentioned above.

In addition to psychological emotion assessment scales, several scientific studies have been carried out to detect human emotions automatically. Facial expression-based emotion recognition is studied in [22], where the authors' used a maximum entropy Markov model (MEMM) [15] to recognize basic emotions. A decision tree-based typical facial expression recognition (FER) system is proposed in [32]. Support vector machines (SVMs) were used in [33] for facial expression-based emotion recognition. A hidden Markov model (HMM)-based FER is studied in [34], where the authors' extract the emotions from time-sequential facial expression images of a video. However, such facial image analysis-based emotion recognition system analyzes facial expressions [35], not emotions. In other words, FER can measure the valence level but not the arousal level of human emotions.

To overcome the limitations of the traditional self-assessment-based psychological emotion recognition and facial image analysis-based emotional expression recognition, the body sensor network [36], [37] or physiological observation-based affective state recognition system has been studied. AlZoubi *et al.* [38] used brain electroencephalogram (EEG) signals to determine the affective state. An SVM classifier is applied to discriminate ten emotion classes with a maximum of 55% accuracy. In [39], the higher order crossings of EEG are analyzed for emotion recognition. Quadratic discriminant analysis (QDA) and an SVM classifier are used to discriminate six basic emotion classes. The single and multi-channel EEG signal models achieve 62.3% and 83.33% classification accuracy, respectively. However, the EEG signal has a lower spatial resolution [40] and implements sophisticated methods to analyze the EEG signal observations, which requires a laboratory or clinical environment for data collection and experimental study. Therefore, EEG signal-based affective state mining approaches are generally not feasible for an ambient assisted-living environment. Furthermore, unlike a fully functional 16 or 32 electrode EEG sensor, the EDA, ECG and EMG sensors are wearable as a wrist or chest band, which are suitable for ubiquitous monitoring.

Liu *et al.* [41] proposed a physiology-based affect recognition system for patients with autism spectrum disorder (ASD). They used skin conductance (SC), ECG, EMG, and skin temperature (ST) sensors to analyze the three different affective states. An SVM classifier is used for successful classification of 82.9% of affective states. However, the study was solely on children with ASD and defined affective states such as anxiety, engagement, and liking. In a recent study [42], radio frequencies are used for human heart-beat recognition, and measured heart-beat signals can be used for emotion recognition. That study considered four basic emotion classes and applied a 1-norm support vector machine ($l_1 - SVM$) classifier to achieve 72% classification accuracy in a person independent experimental setup. However, heart-rate variability determination through radio frequency is a passive measurement approach, which is unable to recognize individual emotions in a crowd, e.g., in a hospital, apartment or outdoor environment.

Koelstra *et al.* [43] proposed physiological signal-based emotion analysis, where the authors used 40 channels of biomedical and peripheral sensors, including EEG, ECG and EMG signals [43]. That work is a pioneer in physiological signal-based emotion analysis and utilized a naive Bayes classifier to classify the valence and arousal levels. They gained 61.8% and 65.1% accuracy in valence and arousal level classification, respectively. The authors used an exceedingly large amount of sensors, including the complex 32-channel EEG signals, which requires an extensive environmental setup and is thus still not feasible for real-life usage. Therefore, this paper proposed a convolutional neural network (CNN) [44] based emotion recognition method, where wearable and light-weight electromyography (EMG), electro-dermal activity (EDA) and electrocardiogram (ECG) sensors are used to observe the physiological changes in different affective states. This research determines the valence and arousal [45] levels from physiological observations induced while watching the video stimuli of different affective states.The affective valence is described through a scale of unpleasantness to pleasantness, whereas the affective arousal is defined through a range of calming to exciting [45]. Based on the valence and

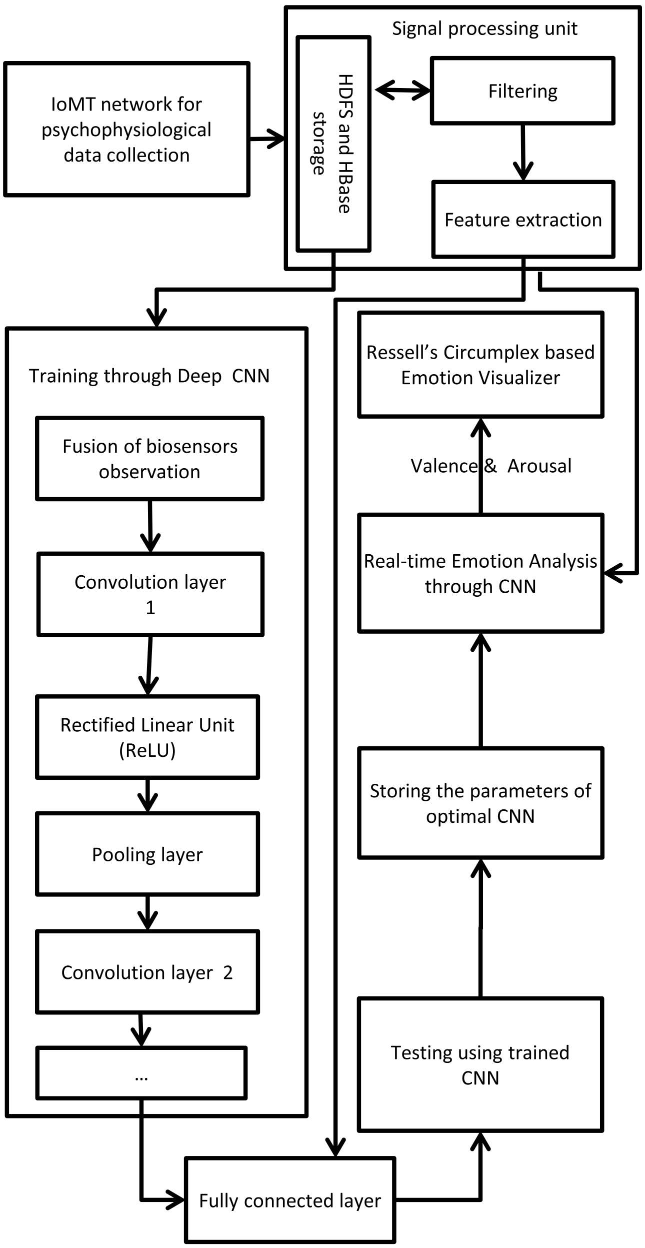

FIGURE 2. Internet-of-medical-things (IoMT)-based affective state mining framework via a deep convolutional neural network (CNN).

arousal levels, the affective states are determined according to Russell's emotion circumplex.

III. IOMT-BASED AFFECTIVE STATE MINING (IASM) FRAMEWORK

The internet-of-medical-things (IoMT)-based affective state mining framework is presented in Fig. 2. Processing biomedical signals and mining the affective states from the biosensor observations require complex processing environment as in [46]. Lightweight IoMT wearables [47] and hand-held devices with limited processing and memory capability are not feasible for analyzing such massive data. The proposed affective state mining framework leverages the hidden iceberg computation and enables real-time emotion recognition via hand-held or wearable devices.

A. IOMT NETWORK FOR PHYSIOLOGICAL DATA COLLECTION

In the proposed IASM framework, an IoT network of biomedical sensors is formed for collecting subjects' physiological markers for mining their affective states. There are three types of biomedical sensors such as electro-dermal activity (EDA), electrocardiogram (ECG), and electromyogram (EMG), as well as a sink node in the proposed IoMT network. The EDA sensor is used to determine the skin conductance level, which is based on the secretion of sweat and reflects the arousal level [48]. The ECG sensor is used to measure the arousal level by comparing the frequency bands of parasympathetic and sympathetic signals [49]. The electrical impulses of the zygomaticus muscle, which are observed via facial electromyogram (EMG) sensor, are used to discriminate the valence levels [50]. However, none of these biosensors' channels are individually adequate for extracting the valence and arousal levels of induced emotion. Therefore, the observation level fusion [9], [37], [51] of different biomedical sensor channels is used to classify human emotions. The gateway or sink node of the IoMT network collects the signals from the biosensors and sends those to the signal processing module for filtering and feature extraction.

B. SIGNAL PROCESSING UNIT (SPU)

The signals are stored in a Hadoop Distributed File System (HDFS) and HBase storage for distributed processing and feature extraction. The HDFS enables a cluster of computing nodes for parallel processing. The different channels of EMG, ECG, and EDA sensors are processed in parallel fashion for signal preprocessing and feature extraction. Therefore, the sensor data are split into time-sequential windows to speed-up processing through parallelism, and eradicating the need of ETL (extract, transform, and load) operations. The collected and stored physiological signals are down-sampled to 128Hz. Subsequently, the bandpass filtering is applied, and the filtered biosignals are used in sensible features extraction for pattern classification. Three different HDFS cluster nodes are used for feature extraction of the three different biomedical sensors.

The time and frequency domain features are extracted from ECG, EDA and EMG sensors. The peak amplitude, i.e., the signal peaks and index locations, are identified through a *state-machine* design approach. The extracted ECG features are: $\{mean(HRV), sd(HRV), sd(d'(HRV)), p(NN50), psd(HF), psd(LF), r(psd(LF), psd(HF)), poincare(SD_1), poincare(SD_2), r(poincare(SD_1), poincare(SD_2))\}$; where, HRV, sd, d', d'', $p(NN50)$, $psd(HF)$, $psd(LF)$, $r()$, $poincare(SD_1)$, and $poincare(SD_2)$ represent the heart rate variability [52], standard deviation, first derivation, second derivation, number of successive normal-to-normal (NN) interval pairs differing by greater than 50 ms, power spectral

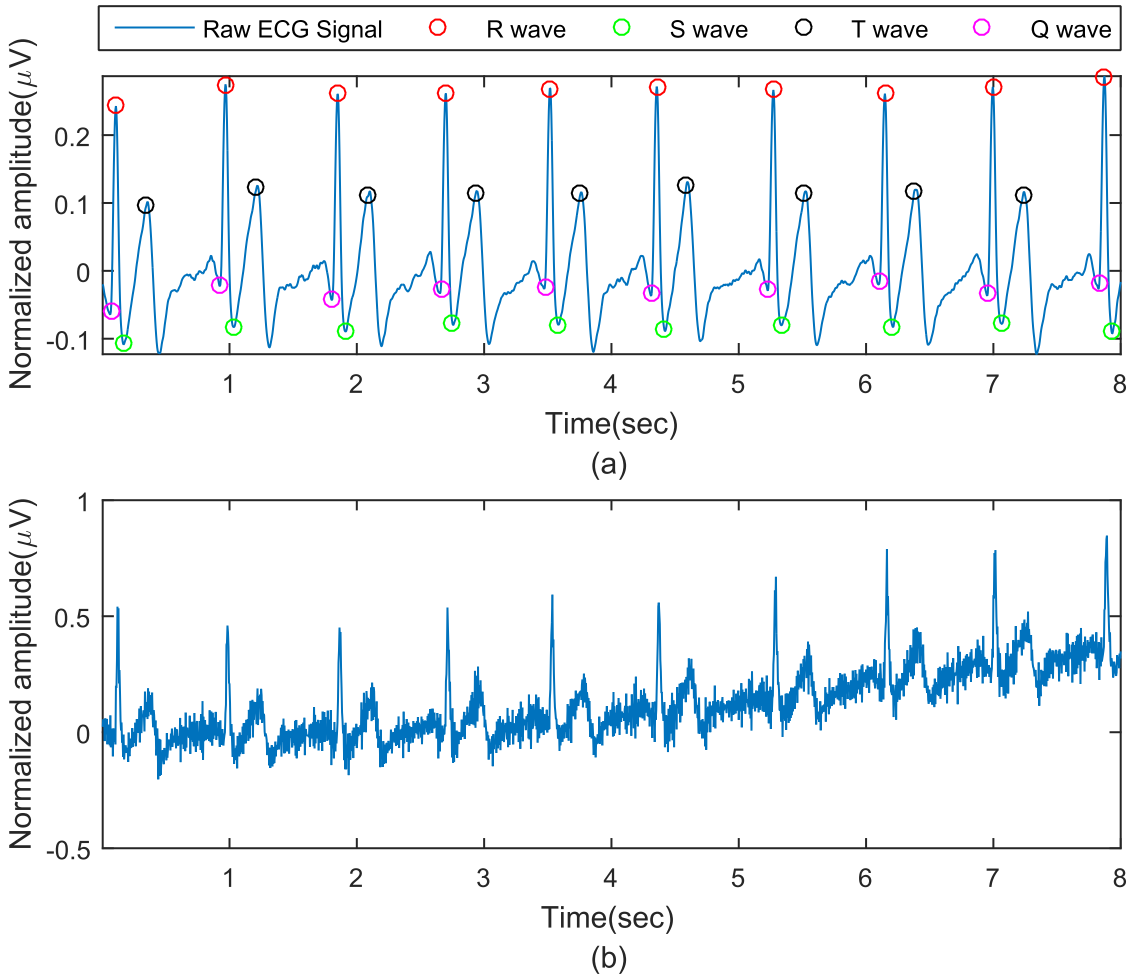

FIGURE 3. Bio-signal processing and feature extraction: (a) ECG peak detection for HRV determination and QRS complex identification (b) raw ECG signal.

density of the high-frequency band, power spectral density of the low-frequency band, ratio, Poincare standard deviation one (i.e., fast beat-to-beat variability), and Poincare standard deviation two (i.e., longer term variability of interbeat interval), respectively. The raw and filtered ECG signals with detected peaks are shown in Fig. 3(a) and Fig. 3(b), respectively. The power spectral density (PSD) of the ECG signal is shown in Fig. 4. For analyzing the distribution of heart rate variability patterns, Poincare plots are utilized. The standard deviation of the ECG voltage dispersion is measured in the diagonal line (SD_1). Afterwards, the voltage dispersion of (SD_1) the normal or perpendicular line is determined as (SD_2). The Poincare graph is shown in Fig. 5. However, the sharpness of the scatter pattern is evaluated via the SD_1/SD_2 ratio.

The skin conductance level (SCL) associated with the various affective state is measured through EDA sensors. The electric field potential is measured between a specific point and reference point of skin to determine the SCL. The electrical conductivity and it's phasic change are determined to measure the skin conductance response (SCR). The extracted

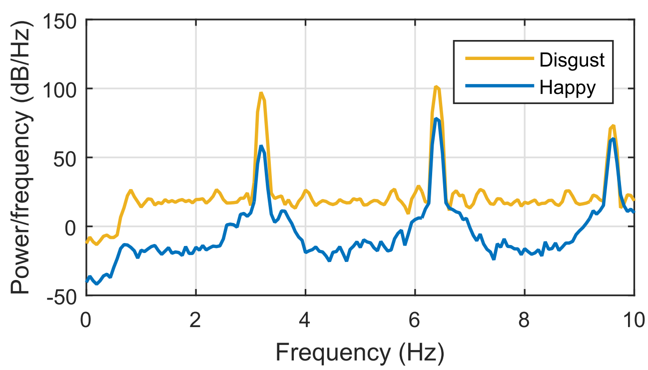

FIGURE 4. Power spectral density (PSD) comparison of the disgusted and happy emotions.

EDA features are: $\{mean(SC_n), mean(SC_l), sd(SC_n), sd(SC_l), mean(d'(SC_n)), mean(d'(SC_l)), mean(d''(SC_l)), occ(SCR), (mean(SCR_{amp}))\}$; where, SC_n, SC_l, occ, and SCR_{amp} represent the normalized skin conductivity, low-passed (cutoff frequency of 0.2 Hz) skin conductivity, occurrences or frequency of SCR, and the SCR amplitude, respectively. The measured

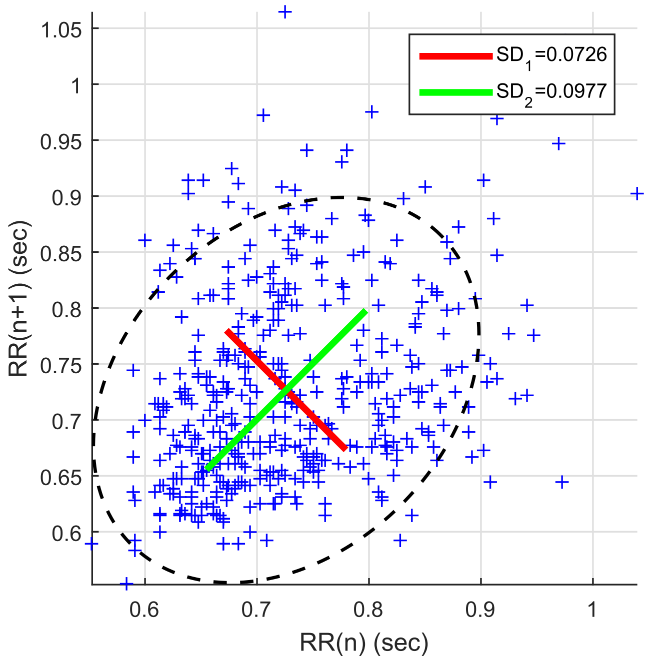

FIGURE 5. The poincare plots to analyze the heart rate variability patterns.

SCL of a human subject is shown in Fig. 6. Moreover, the first derivative of the measured SC signal is shown in Fig. 7.

The electrical impulses of zygomaticus muscles were measured through zEMG, which holds the influential features of joy or happiness and therefore the valence level as well. The raw signal of the zygomaticus major EMG is shown in Fig. 8. The extracted EMG features are: $\{mean(ZMA), sd(ZMA), integral(ZMA), d'(ZMA)), mean(fft(ZMA)), median(fft(ZMA))\}$; where, *ZMA*, *integral* and *fft* represent the zygomatics major activity (ZMA), integrated ZMA and fast Fourier transformation, respectively. The integral of the Zygomaticus EMG (zEMG) determines the area under the curve (AUC) of the EMG signal, as shown in Fig. 9, which is an estimation of the muscular efficiency.

C. TRAINING THROUGH A DEEP CONVOLUTIONAL NEURAL NETWORK (CNN)

The convolutional neural network (CNN) is a class of feed-forward neural networks which were developed for handling deep network structures. There are three hierarchical layers in CNN design: 1) Convolution layer: A convolutional layer provides translation invariance. Since the convolution kernel works on every part of the sensor observation tensor, basically those kernels are searching for the same feature everywhere in the tensor. The shallower convolution layer extracts the edge features. However, the deeper convolution layers extract the potential features. 2) Pooling layer: A convolution layer is followed by a pooling layer for down-sampling the feature maps generated by the convolution layer. The down-sampling in pooling layer reduces the number of learn-able hyper-parameters of CNN without loss of generality. 3) Fully connected neural network layer: The fully connected layer follows the traditional neural network architecture. Here, every input unit (i.e., the outputs of the last pooling layer) are connected to every hidden or output layer neuron. This layer produces the classification results. The convolution and pooling operations are depicted in Fig. 10.

There are also three major operations in CNN: 1) forward propagation, 2) backward propagation, and 3) parameter up-gradation. In deep CNNs, the forward propagation $P_{forward}$ is the composite function (1) of the convolution, sigmoid and pooling operations, where $O = \{o_{11}, o_{12}, \ldots, o_{1M}; o_{21}, o_{22}, \ldots, o_{2M}; o_{31}, o_{32}, \ldots, o_{3M}\}$ is the fused sensor observations.

$$P_{forward} = f^{(pool)} \circ f^{(ReLU)} \circ f^{(conv)}(O) \tag{1}$$

If O is considered as the input of the convolution layer and P is considered as the output of the convolution layer, then (2) holds. Here, $W = \{w_{11}, w_{12}, \ldots, w_{1N}; \ldots\ldots\ldots; w_{N1}, w_{N2}, \ldots, w_{NN}\}$ is the $N \times N$ convolution kernel. For simplicity, the linear dimension is considered as $|W| = n$, where, n must be greater than or equal to 1 but less than or equal to both $|P|$ and $|Q|$.

$$P = O * W = W^T O_{n:n+|W|-1} \tag{2}$$

Now, the partial derivative of the output P of (2) with respect to the kernel parameter W, will result in (3), which will be used in backward propagation and $1 \le i \le |W|$.

$$\frac{\partial P_n}{\partial W_i} = O_{n+i-1} \tag{3}$$

Again, the partial derivative of the output P of (2) with respect to the input observations O will result in (4) for backward propagation.

$$\frac{\partial P_{n-i+1}}{\partial O_n} = W_i \tag{4}$$

The activation function is considered as the rectified linear unit (5).

$$ReLU(P) = max\,(0, P) \tag{5}$$

In the pooling layer, considering the $m \times m$ max pooling, the single output element of sub-sampling will be (6). Here, the function g is the aggregate function of max pooling as defined in (7).

$$P_n = g\left(ReLU(P)_{(n-1)m+1:nm}\right) \tag{6}$$

$$g\,(ReLU(P)) = max(ReLU(P)) \tag{7}$$

The cost function is considered as the squared error function (8), which measures the difference or error among the ground truth labels L and forward pass estimations.

$$E(W, b, O, L) = \frac{1}{2} \parallel L - P_{forward} \parallel^2 \tag{8}$$

The gradient of the error is estimated through (9), where $\frac{\partial P}{\partial W_i}$ is determined by (3).

$$\frac{\partial E}{\partial W_i} = \frac{\partial E}{\partial P}\frac{\partial P}{\partial W_i} = \sum_{n=1}^{|O|-|W|+1} \eta_n^P O_{n+i-1} \tag{9}$$

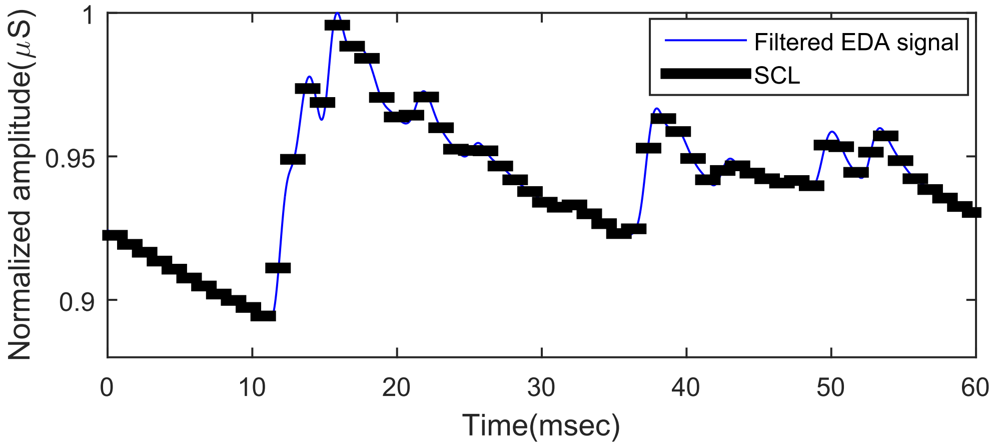

FIGURE 6. The measured skin conductance levels (SCLs) of the electro-dermal activity (EDA) sensor.

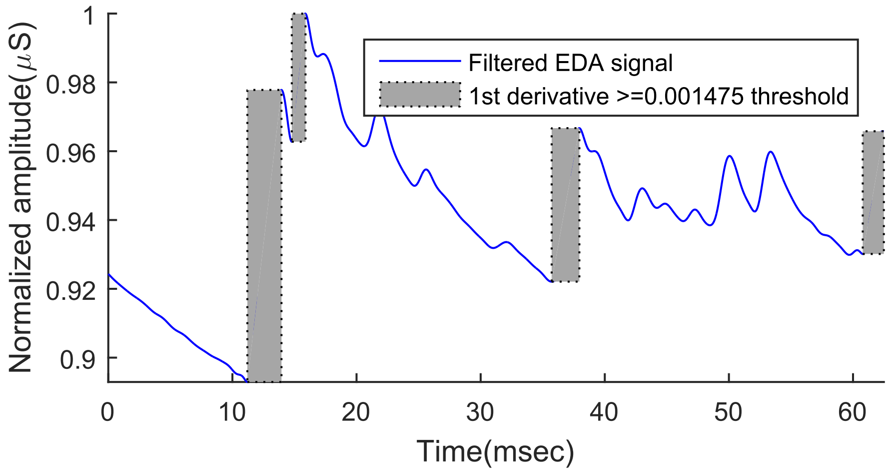

FIGURE 7. The EDA signal feature determined through differentiating the measured skin conductance levels (SCLs).

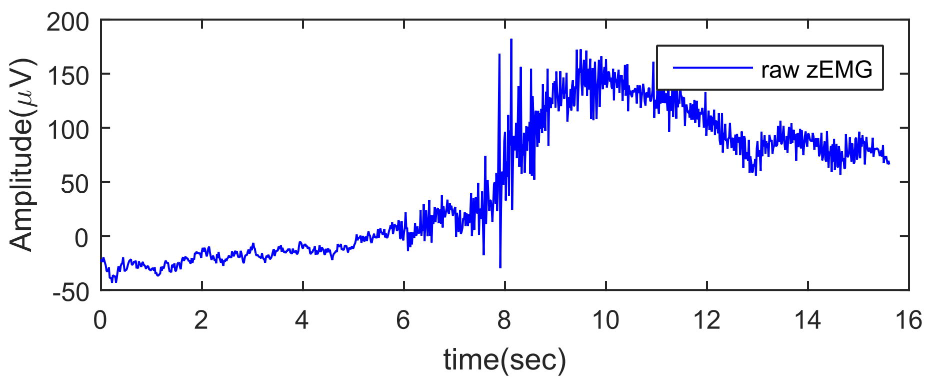

FIGURE 8. The raw zygomaticus electromyogram (zEMG) signal.

Moreover, η_n^O is the gradient of the error with respect to the input O and can be determined through (10). In backward propagation, η_n^O is determined in the convolution layer.

$$\eta_n^O = \frac{\partial E}{\partial O_n} = \frac{\partial E}{\partial P}\frac{\partial P}{\partial O_n} = \sum_{i=1}^{|W|} \eta_{n-i+1}^P W_i \tag{10}$$

The error signal is passed along the backward direction as well as the pooling layer and is determined by (11), where g' is the derivative of the aggregate function g of (7).

$$\eta_{(n-1)m+1:nm}^O = \eta_n^P g_n' \tag{11}$$

In backward propagation, the determined error signal is passed through the activation layer (ReLU), followed by the

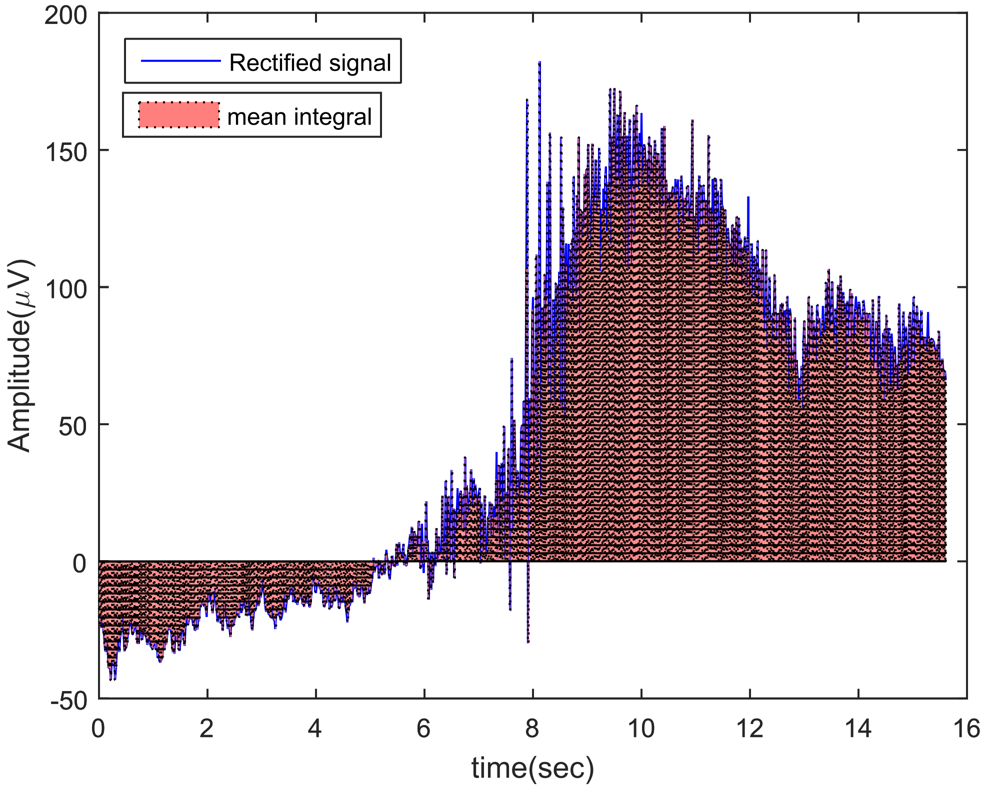

FIGURE 9. The mean integral of the EMG signal to estimate muscular efficiency.

convolution layer and the input layer. Therefore, the input layer receives $\eta^{O}_{(n-1)m+1:nm}$ as an input using the following equation (12), where $\bullet$ is used to denote the Hadamard product.

$$\eta^{O} = \left(\eta^{P} g'\right) \bullet ReLU'(P) * W \tag{12}$$

Using (9), the gradient of the error $\frac{\partial E}{\partial W} = \nabla_W E$ of the full network is determined, and the network parameters are upgraded through (13).

$$W = W - \zeta \nabla_W E \tag{13}$$

In the proposed IoMT-based affective state mining framework, the biosignals sample data are partitioned into training, and testing sets. The described convolutional neural network (CNN) is used to train the proposed IASM model. To extract the high-level features, different kernels are used in the convolution layer to produce various feature maps. Additionally, to adopt non-linear properties into the decision function, the rectified linearization (ReLU) method is used on the output of the convolution layer and then, the feature maps are down-sampled in pooling layer. To achieve the convergence of the deep CNN model, several blocks of convolution, ReLU, and pooling operations are executed iteratively. The fully connected layer is embedded at the end of the CNN model. In the legacy CNN architecture, the outputs of the last pooling layer are fed into the fully connected layer. However, in the proposed CNN architecture, the extracted features of the signal processing unit and high-level features from the last pooling layer are fused and fed into the fully connected layer to extract the global discriminative features. This feature adds novelty to the proposed CNN-based IASM model. The output of the fully connected layer produces the predicted class labels of emotions. In the training phase, the hyper-parameters of the proposed CNN are updated through back-propagation and the stochastic gradient methods. The trained CNN model is stored and applied for real-time emotion recognition. To analyze real-time emotions, the collected biosensor observations are fed into the trained CNN, which predicts the real-time arousal and valence levels of the induced affective state of the user.

IV. PERFORMANCE EVALUATION

The performance of the proposed emotion recognition approach, i.e., biomedical sensor observation-based affective state mining using a convolutional neural network, is studied using a benchmark DEAP dataset.

A. DESCRIPTION OF DATASET

The standard DEAP dataset [43] is used to evaluate the performance of the proposed IASM framework for affective state mining. The DEAP is a limited access dataset, which is developed for physiological signal-based human emotion recognition. Data were collected from 32 subjects. The total 40 video clips each of one-minute length are used as stimuli.

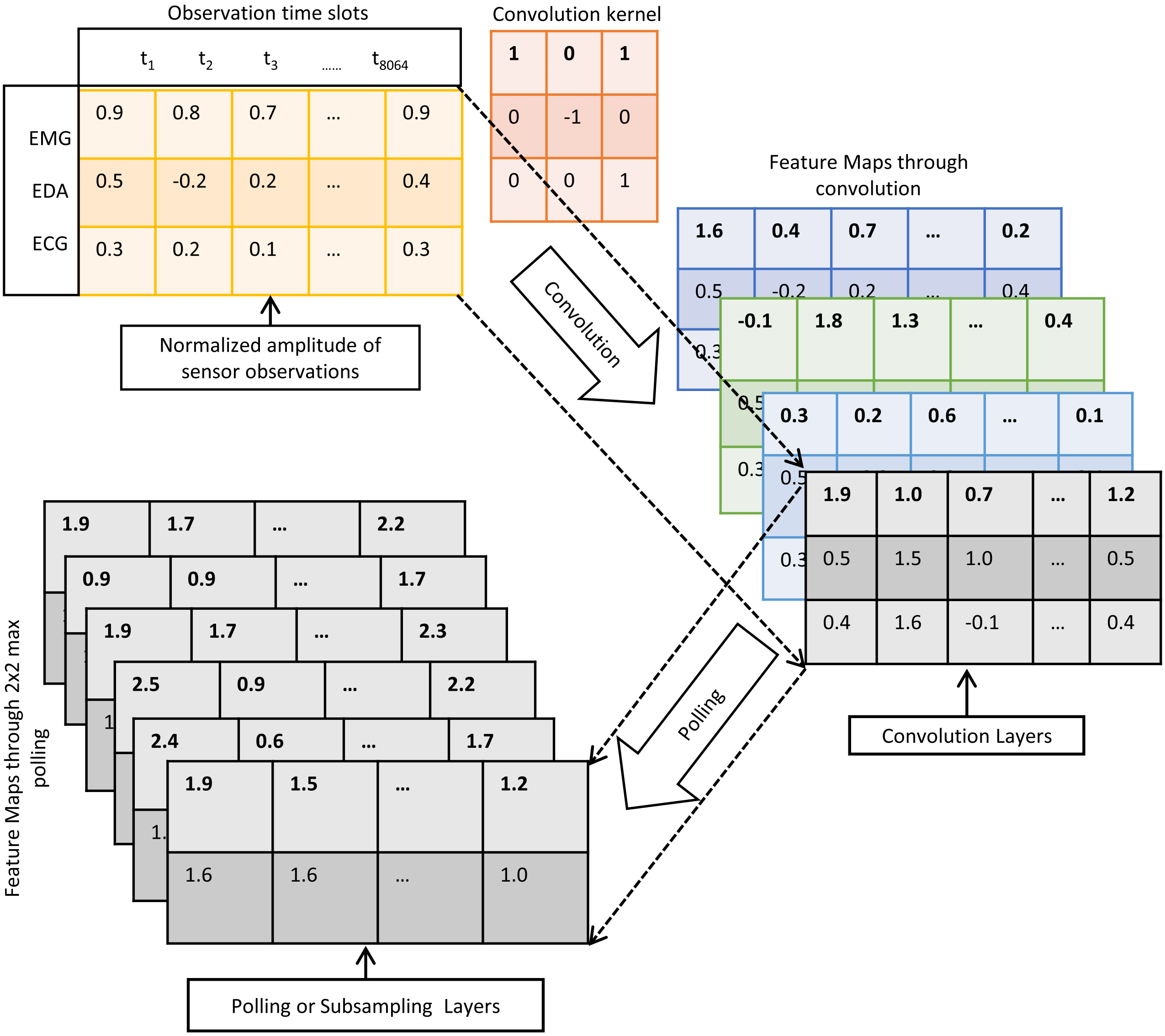

FIGURE 10. The convolution and pooling operations of the CNN model for IoMT-based affective state mining.

The physiological observations of 32 channels EEG, 12 channels peripheral sensors, and one channel status are recorded for affective states recognition. However, this research used only 3 peripheral sensors i.e., electro-dermal activity (EDA), electromyography (EMG) and photoplethysmogram (PPG) observations for evaluating the performance of the proposed model of affective state mining. A 9-point Likert scale is used to collect the ratings of arousal, valence, liking, and dominance levels of the video stimuli. Therefore, the dimension of the used dataset is $32 \times 40 \times 3 \times 8064$; where 32 is the number of subjects, 40 is the number of video stimuli, 3 is the number of sensor channels, and 8064 is the number of sample data points.

The performance of the proposed IoMT-based affective state mining approach IASM is shown in Fig. 11. The receiver operating characteristic (ROC) curves of the five affective states in DEAP dataset is presented in Fig. 11. Five-fold cross-validation is applied in this study. The area under the ROC curve (AUC) in the overall ROC plot is 0.9474, 0.9433, 0.9389, 0.9155, and 0.9112 respectively for Sad, Happy, Disgusted, Neutral and Relaxed affective states, respectively.

B. TESTBED IMPLEMENTATION

A testbed is implemented to evaluate the performance of the proposed IASM system via a convolutional neural network (CNN). The Torch 7.0 deep learning framework is used for the implementation. The Lua programming language is used for development of the CNN model. After installing Torch, LuaJIT is installed, which is a just-in-time interpreter for compilation and interpretation of the Lua language. Moreover, for processing vast amounts of data, GPU programming is designed using the CUDA and CuDNN packages on top

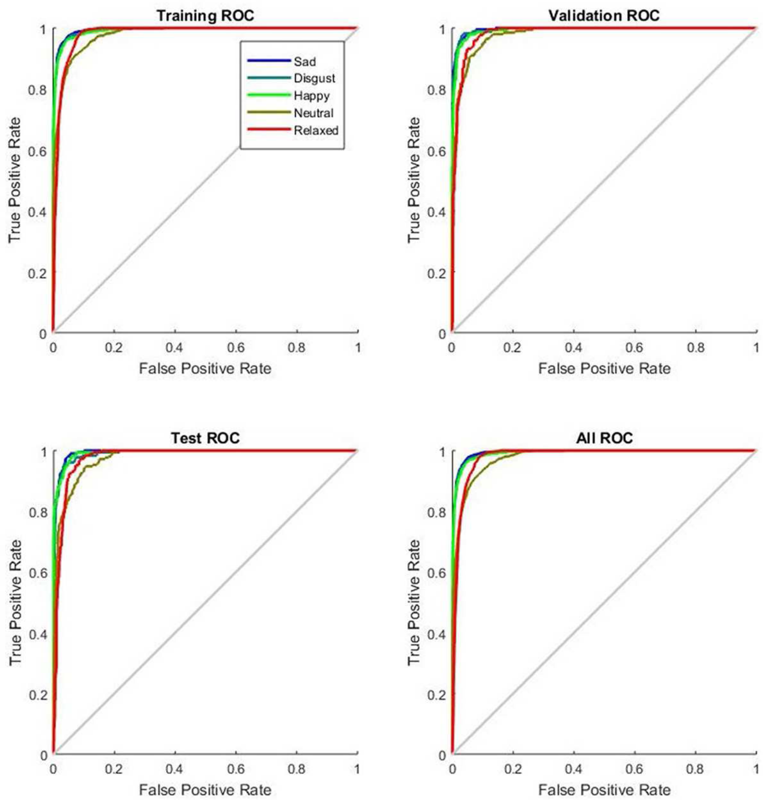

FIGURE 11. The receiver operating characteristic (ROC) curves of five affective states.

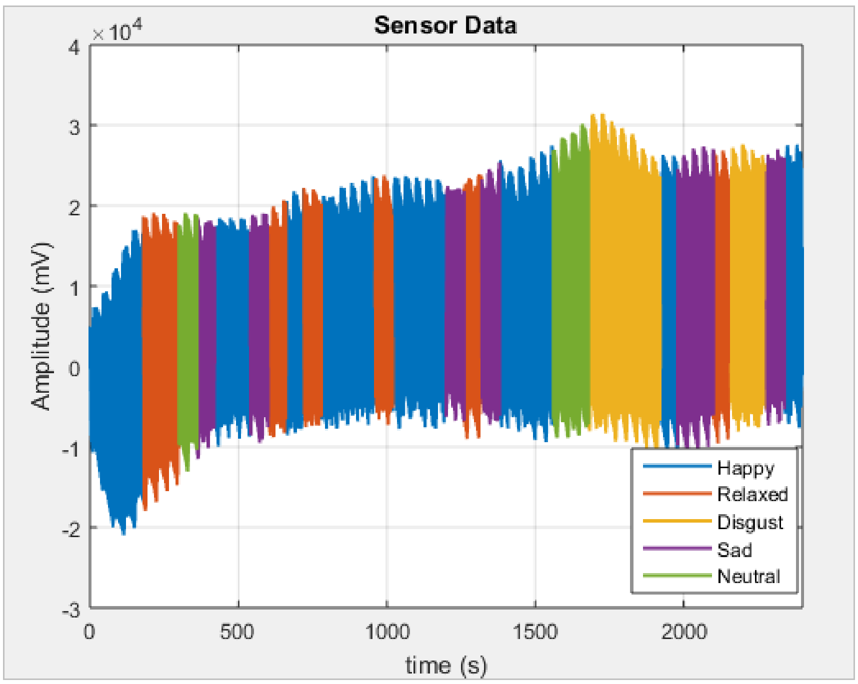

FIGURE 12. The PPG signal pattern of the five basic affective states.

of the Torch interface. For the big-data loading interface, PyTorch API is used, which enables a Python 3.0 interface for data loading and storing.

To expose emotions of an individual, the video stimuli of DEAP dataset are used. The EDA, ECG, PPG, and zEMG IoMT-sensors are placed on the human body to

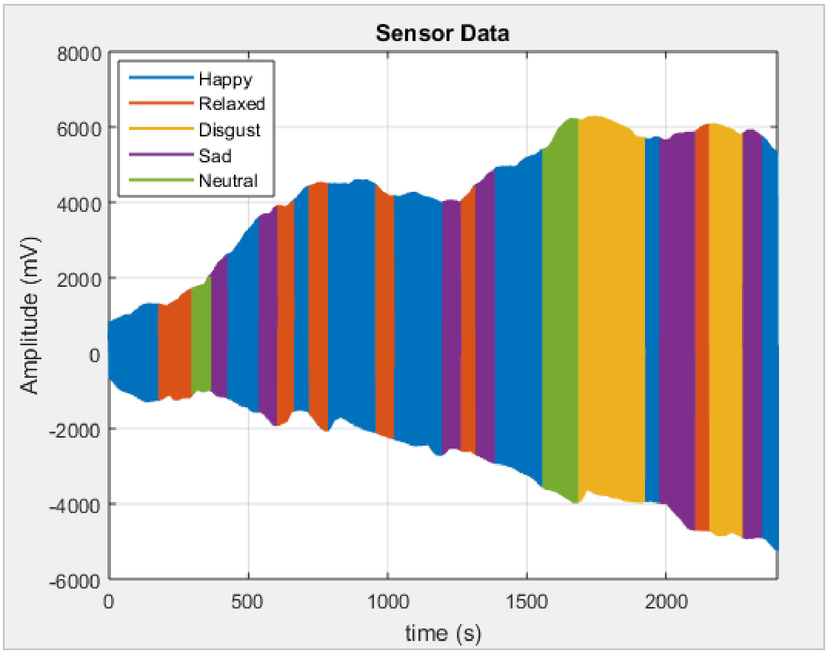

FIGURE 13. The EDA signal pattern of the five basic affective states.

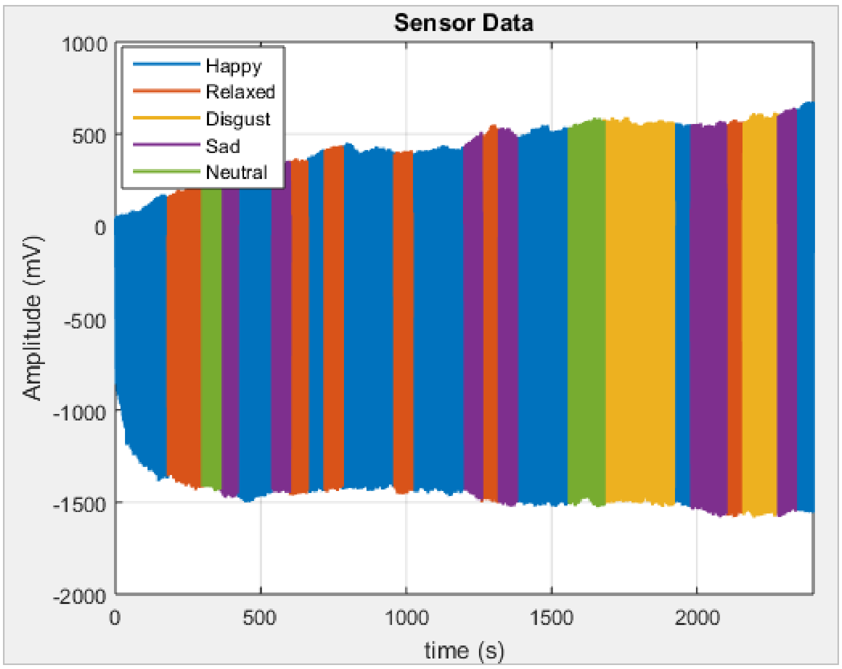

FIGURE 14. The EMG signal pattern of the five basic affective states.

collect bio-signal observations while the subject is watching the video stimuli. As the electroencephalogram (ECG) and photoplethysmogram (PPG) sensors are complementary, the ECG sensor observations are used instead of PPG, which is used in the DEAP dataset. The significance of using ECG signal over PPG is that the ECG facilitates to extract potential discriminative features, e.g., $p(NN50)$, $poincare(SD)$ (see section III(B)), other than the normal heart rate variability. To prepare the training and testing dataset, the subjects are requested to rate their arousal and valence levels using a 9-point Likert scale, while watching the video stimuli.

The collected PPG, EDA and EMG signal patterns are shown in Fig. 12, Fig. 13 and Fig. 14, respectively for a subject. Refereeing those figures, it is shown that there is no pattern to individual biosensor observations which can distinguish the Happy, Relaxed, Disgusted, Sad and Neutral moods. The figures demonstrate the difficulty of finding psychophysiological markers for affective state mining. Therefore, we need to combine the sensor observations to uncover patterns for emotion classification. Hence, the proposed non-linear deep convolutional neural network model of IASM framework is trained and applied for identifying the real-time affective states of users.

TABLE 1. **Confusion matrix of the proposed affective state mining approach PAC using DEAP dataset [43] (Unit:%).**

	Happy	Relaxed	Disgust	Sad	Neutral
Happy	**87.88**	4.24	2,42	1.6	3.84
Relaxed	4.40	**85.53**	1.26	3.77	5.03
Disgust	3.7	1.39	**86.57**	4.6	3.7
Sad	1.54	1.92	3.08	**90.77**	2.69
Neutral	2.67	5.33	4.00	4.00	**84.00**
Average		**87.5**			

The performance of the developed IASM prototype is shown in Table 1. The EDA observations induced very clear discriminative factors in the case of negative arousal, which is the key for sad emotions. Therefore, the proposed IASM framework has shown highest 90% classification accuracy in the case of mining sad emotions. The standard emotion circumplex of Russell didn't define the neutral emotion through arousal and valence axes. For simplicity, the valence and arousal levels between 4 to 6 of the 9-point Likert scale are considered as neutral emotion in the proposed IASM model. However, because of such linear definition, the CNN classifier suffers to discriminate neutral emotion from other affective states and produces a classification accuracy of only 84%. In summary, the proposed IASM model has achieved 87.5% classification accuracy in general, compare to the 72% of the radio frequency-based emotion analyzer [42], 82.9% of the SVM-based emotion classifier [41], and 61.8% valence and 65.1% arousal level classification accuracy of DEAP [43]. The CNN-based depth features of the proposed IASM play key roles in achieving the overall classification accuracy of 87.5%.

V. CONCLUSION

The wearable internet-of-medical-things market is rapidly growing, especially for health status monitoring and athletic training. Affective state recognition from wearable biosensors can complement context-aware recommendation, mood stabilization, and stress and depression management, especially for mental well-being. This research used the observations of biomedical sensors, which are lightweight and readily available within consumer products to infer human emotions. The proposed affective state mining method ensures a higher accuracy of 87.5% compared to the state-of-the-art physiology-based emotion recognition methods. The depth level features, which are generated through a convolutional neural network, play key roles in discriminating the affective states. The prototype implementation justifies the suitability of using the IASM framework for real-time emotion recognition. Even though the accuracy of the proposed deep CNN-based affective state mining method is relatively improved, there are still different approaches that can be taken to enhance the performance gain further. The biomedical sensor data are continuous and sequential, and therefore, the use of a time-sequential deep learning architecture e.g., deep recurrent neural network may enhance the classification accuracy.

REFERENCES

[1] D. Hill, *Emotionomics: Leveraging Emotions for Business Success*. London, U.K.: Kogan Page, 2010.

[2] M. Chen, P. Zhou, and G. Fortino, "Emotion communication system," *IEEE Access*, vol. 5, pp. 326–337, 2017.

[3] R. W. Picard, "Affective computing. M.I.T Media Laboratory Perceptual Computing Section," Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. 321, 1995. [Online]. Available: https://affect.media.mit.edu/pdfs/95.picard.pdf

[4] S. O. Bae, M. D. Kim, J. G. Lee, J.-S. Seo, S.-H. Won, Y. S. Woo, J.-H. Seok, W. Kim, S. J. Kim, K. J. Min, D.-I. Jon, Y. C. Shin, W. M. Bahk, and B.-H. Yoon, "Is it useful to use the korean version of the mood disorder questionnaire for assessing bipolar spectrum disorder among Korean college students?" *Asia–Pacific Psychiatry*, vol. 6, no. 2, pp. 170–178, Jun. 2014.

[5] M. G. R., E. J. Cho, E.-N. Huh, and C. S. Hong, "Cloud based mental state monitoring system for suicide risk reconnaissance using wearable bio-sensors," in *Proc. 8th Int. Conf. Ubiquitous Inf. Manage. Commun.*, Jan. 2014, Art. no. 56.

[6] M. M. Hassan, M. G. R. Alam, M. Z. Uddin, S. Huda, A. Almogren, and G. Fortino, "Human emotion recognition using deep belief network architecture," *Inf. Fusion*, vol. 51, pp. 10–18, Nov. 2019.

[7] M. Z. Uddin, M. M. Hassan, A. Almogren, M. Zuair, G. Fortino, and J. Torresen, "A facial expression recognition system using robust face features from depth videos and deep learning," *Comput. Electr. Eng.*, vol. 63, pp. 114–125, Oct. 2017.

[8] H. Ghasemzadeh, P. Panuccio, S. Trovato, G. Fortino, and R. Jafari, "Power-aware activity monitoring using distributed wearable sensors," *IEEE Trans. Human-Mach. Syst.*, vol. 44, no. 4, pp. 537–544, Aug. 2014.

[9] Z. Wang, D. Wi, R. Gravina, G. Fortino, Y. Jiang, and K. Tang, "Kernel fusion based extreme learning machine for cross-location activity recognition," *Inf. Fusion*, vol. 37, pp. 1–9, Sep. 2017.

[10] C. Kappeler-Setz, F. Gravenhorst, J. Schumm, B. Arnrich, and G. Tröster, "Towards long term monitoring of electrodermal activity in daily life," *Pers. Ubiquitous Comput.*, vol. 17, no. 2, pp. 261–271, Feb. 2013.

[11] O. Mayora, "The MONARCA project for bipolar disorder treatment," *J. Cyber Therapy Rehabil.*, vol. 1, no. 4, pp. 14–15, 2011.

[12] H. Imaoka, K. Inoue, Y. Inoue, H. Hazama, T. Tanaka, and N. Yamane, "R-R intervals of ECG in depression," *Psychiatry Clinical Neurosciences*, vol. 39, no. 4, pp. 485–487, Dec. 1985.

[13] R. Gravina and G. Fortino, "Automatic methods for the detection of accelerative cardiac defense response," *IEEE Trans. Affect. Comput.*, vol. 7, no. 3, pp. 286–298, Jul./Sep. 2016.

[14] M. G. R. Alam, S. F. Abedin, M. A. Ameen, and C. S. Hong, "Web of objects based ambient assisted living framework for emergency psychiatric state prediction," *Sensors*, vol. 16, no. 9, P. 1431, 2016.

[15] M. G. R. Alam, R. Haw, S. S. Kim, M. A. K. Azad, S. F. Abedin, and C. S. Hong, "EM-Psychiatry: An ambient intelligent system for psychiatric emergency," *IEEE Trans. Ind. Inform.*, vol. 12, no. 6, pp. 2321–2330, Dec. 2016.

[16] G. Harerimana, B. Jang, J. W. Kim, and H. K. Park, "Health big data analytics: A technology survey," *IEEE Access*, vol. 6, pp. 65661–65678, 2018.

[17] C. Xie, P. Yang, and Y. Yang, "Open knowledge accessing method in IoT-based hospital information system for medical record enrichment," *IEEE Access*, vol. 6, pp. 15202–15211, 2018.

[18] M. Z. Uddin and M. M. Hassan, "Activity recognition for cognitive assistance using body sensors data and deep convolutional neural network," *IEEE Sensors J.*, to be published.

[19] V. Sacharin, K. Schlegel, and K. R. Scherer, "Geneva Emotion Wheel rating study (Report)," Swiss Center Affect. Sci., Univ. Geneva, Geneva, Switzerland, Tech. Rep., 2012. [Online]. Available: https://archive-ouverte.unige.ch/unige:97849

[20] R. JA, "A circumplex model of affect," *J. Personality Social Psychol.*, vol. 39, no. 6, pp. 1161–1178, 1980.

[21] M. Z. Uddin, M. M. Hassan, A. Almogren, A. Alamri, M. Alrubaian, and G. Fortino, "Facial expression recognition utilizing local direction-based robust features and deep belief network," *IEEE Access*, vol. 5, pp. 4525–4536, 2017.

[22] M. H. Siddiqi, M. G. R. Alam, C. S. Hong, A. M. Khan, and H. Choo, "A novel maximum entropy Markov model for human facial expression recognition," *PLoS ONE*, vol. 11, no. 9, Sep. 2016, Art. no. e0162702.

[23] M. Z. Uddin, W. Khaksar, and J. Torresen, "Facial expression recognition using salient features and convolutional neural network," *IEEE Access*, vol. 5, pp. 26146–26161, 2017.

[24] D. C. Ong, J. Zaki, and N. D. Goodman, "Affective cognition: Exploring lay theories of emotion," *Cognition*, vol. 143, pp. 141–162, Oct. 2015.

[25] P. Gangamohan, S. R. Kadiri, and B. Yegnanarayana, "Analysis of emotional speech—A review," in *Toward Robotic Socially Believable Behaving Systems*, vol. 1. Berlin, Germany: Springer, 2016, pp. 205–238.

[26] A. A. A. Esmin, R. L. De Oliveira, and S. Matwin, "Hierarchical classification approach to emotion recognition in Twitter," in *Proc. 11th Int. Conf. Mach. Learn. Appl.*, vol. 2, Dec. 2012, pp. 381–385.

[27] J.-J. Cabibihan and S. S. Chauhan, "Physiological responses to affective tele-touch during induced emotional stimuli," *IEEE Trans. Affective Comput.*, vol. 8, no. 1, pp. 108–118, Jan./Mar. 2015.

[28] S. P. Robbins, T. Judge, and T. T. Campbell, *Organizational Behaviour*. Upper Saddle River, NJ, USA: Prentice-Hall, 2010.

[29] S. F. Abedin, M. G. R. Alam, R. Haw, and C. S. Hong, "A system model for energy efficient green-IoT network," in *Proc. Int. Conf. Inf. Netw. (ICOIN)*, Jan. 2015, pp. 177–182.

[30] D. Konstan, *The Emotions of the Ancient Greeks: Studies in Aristotle and Classical Literature*, vol. 5. Toronto, ON, Canada: Toronto Univ. Press, 2006.

[31] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *J. Behav. Therapy Experim. Psychiatry*, vol. 25, no. 1, pp. 49–59, Mar. 1994.

[32] S. Mohseni, H. M. Kordy, and R. Ahmadi, "Facial expression recognition using DCT features and neural network based decision tree," in *Proc. ELMAR*, Sep. 2013, pp. 361–364.

[33] T. Ahsan, T. Jabid, and U.-P. Chong, "Facial expression recognition using local transitional pattern on Gabor filtered facial images," *IETE Tech. Rev.*, vol. 30, no. 1, pp. 47–52, Sep. 2013.

[34] M. Z. Uddin, J. J. Lee, and T.-S. Kim, "An enhanced independent component-based human facial expression recognition from video," *IEEE Trans. Consum. Electron.*, vol. 55, no. 4, pp. 2216–2224, Nov. 2009.

[35] M. Shoyaib, M. Abdullah-Al-Wadud, S. M. Z. Ishraque, and O. Chae, "Facial expression classification based on dempster-shafer theory of evidence," in *Proc. Belief Functions: Theory Appl.* Berlin, Germany: Springer, 2012, pp. 213–220.

[36] G. Fortino, R. Giannantonio, R. Gravina, P. Kuryloski, and R. Jafari, "Enabling effective programming and flexible management of efficient body sensor network applications," *IEEE Trans. Human-Mach. Syst.*, vol. 43, no. 1, pp. 115–133, Jan. 2013.

[37] R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino, "Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges," *Inf. Fusion*, vol. 35, pp. 68–80, May 2017.

[38] O. AlZoubi, R. A. Calvo, and R. H. Stevens, "Classification of EEG for affect recognition: An adaptive approach," in *Proc. Australas. Conf. Artif. Intell.*, vol. 5866. Berlin, Germany: Springer, 2009, pp. 52–61.

[39] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 186–197, Mar. 2010.

[40] E. D. Kondylis, T. A. Wozny, W. J. Lipski, A. Popescu, V. J. DeStefino, B. Esmaeili, V. K. Raghu, A. Bagic, and R. M. Richardson, "Detection of high-frequency oscillations by hybrid depth electrodes in standard clinical intracranial EEG recordings," *Front. Neurol.*, vol. 5, p. 149, Aug. 2014.

[41] C. Liu, K. Conn, N. Sarkar, and W. Stone, "Physiology-based affect recognition for computer-assisted intervention of children with autism spectrum disorder," *Int. J. Human-Comput. Stud.*, vol. 66, no. 9, pp. 662–677, Sep. 2008.

[42] M. Zhao, F. Adib, and D. Katabi, "Emotion recognition using wireless signals," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, Oct. 2016, pp. 95–108.

[43] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis ;using physiological signals," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 18–31, Jan./Mar. 2012.

[44] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.

[45] M. M. Bradley and P. J. Lang, "Affective reactions to acoustic stimuli," *Psychophysiology*, vol. 37, no. 2, pp. 204–215, Mar. 2000.

[46] P. Pace, G. Aloi, R. Gravina, G. Caliciuri, G. Fortino, and A. Liotta, "An edge-based architecture to support efficient applications for healthcare industry 4.0," *IEEE Trans. Ind. Inform.*, vol. 15, no. 1, pp. 481–489, Jan. 2019.

[47] K. K. Wong, *Methods in Research and Development of Biomedical Devices*. Singapore: World Scientific, 2013.

[48] (2017). *Galvanic Skin Response: The Complete Pocket Guide—Imotions. GSR Pocket Guide, Imotions*. [Online]. Available: https://imotions.com/blog/galvanic-skin-response/

[49] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, "ECG pattern analysis for emotion detection," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 102–115, Jan./Mar. 2012.

[50] C. Vera-Munoz, L. Pastor-Sanz, G. Fico, M. T. Arredondo, F. Benuzzi, and A. Blanco, "A wearable EMG monitoring system for emotions assessment," in *Proc. Probing Exper.*, 2008, pp. 139–148.

[51] G. Fortino, S. Galzarano, R. Gravina, and W. Li, "A framework for collaborative computing and multi-sensor data fusion in body sensor networks," *Inf. Fusion*, vol. 22, pp. 50–70, Mar. 2015.

[52] A. Andreoli, R. Gravina, R. Giannantonio, P. Pierleoni, and G. Fortino, "SPINE-HRV: A BSN-based toolkit for heart rate variability analysis in the time-domain," in *Proc. Wearable Auton. Biomed. Devices Syst. Smart Environ.* Berlin, Germany: Springer, 2010, pp. 369–389.

MD. GOLAM RABIUL ALAM (S'15–M'17) received the B.S. and M.S. degrees in computer science and engineering and information technology, respectively, and the Ph.D. degree in computer engineering from Kyung Hee University, South Korea, in 2017. He served as a Postdoctoral Researcher with the Computer Science and Engineering Department, Kyung Hee University, from 2017 to 2018. He is currently an Associate Professor with the Computer Science and Engineering Department, BRAC University, Bangladesh. His research interests include healthcare informatics, mobile cloud and edge computing, ambient intelligence, and persuasive technology. He is a member of the IEEE IES, CES, CS, SPS, CIS, and ComSoc. He is also a member of the Korean Institute of Information Scientists and Engineers (KIISE) and received several best paper awards from prestigious conferences.

SARDER FAKHRUL ABEDIN (S'18) received the B.S. degree in computer science from Kristianstad University, Kristianstad, Sweden, in 2013. He is currently pursuing the Ph.D. degree in computer science and engineering with Kyung Hee University, South Korea. His research interests include the Internet of Things network management, cloud computing, fog computing, and wireless sensor networks. He is a member of the Korean Institute of Information Scientists and Engineers.

SEUNG IL MOON received the B.S., M.S., and Ph.D. degrees from the Department of Computer Engineering, Kyung Hee University, South Korea, in 2011, 2013, and 2018, respectively, where he is currently serving as a Postdoctoral Researcher with the Computer Science and Engineering Department. He has authored several national and international journals and conference proceedings. His research interests include ambient intelligent living, advanced wireless network protocols, and SDN networks. He is a member of KIISE.

ASHIS TALUKDER (S'17–M'18) received the B.S. and M.S. degrees in computer science and engineering from the University of Dhaka, Bangladesh. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Kyung Hee University, South Korea. He has been an Assistant Professor with the Department of Management Information Systems (MIS), University of Dhaka, since 2009. His research interests include social networks, influence maximization, network optimization, and data mining. He is a member of the IEEE Communication Society (IEEE ComSoc), SIGAPP, the Association for Information Systems (AIS), Bangladesh Chapter, and the Internet Society (ISOC), Bangladesh Chapter. He is also a member of the Korean Institute of Information Scientists and Engineers (KIISE).

CHOONG SEON HONG (S'95–M'97–SM'11) received the B.S. and M.S. degrees in electronic engineering from Kyung Hee University, Seoul, South Korea, in 1983 and 1985, respectively, and the Ph.D. degree from Keio University, Minato, Japan, in 1997. In 1988, he joined KT, where he was involved in broadband networks as a Technical Staff Member. In 1993, he joined Keio University. He was with the Telecommunications Network Laboratory, KT, as a Senior Member of Technical Staff and the Director of the Networking Research Team, until 1999. Since 1999, he has been a Professor with the Department of Computer Engineering, Kyung Hee University. His research interests include the future Internet, ad hoc networks, network management, and network security. He is a member of ACM, IEICE, IPSJ, KIISE, KICS, KIPS, and OSIA. He has served as the General Chair, the TPC Chair/Member, or an Organizing Committee Member of international conferences, such as NOMS, IM, APNOMS, E2EMON, CCNC, ADSN, ICPP, DIM, WISA, BcN, TINA, SAINT, and ICOIN. He is currently an Associate Editor of the IEEE Transactions on Network and Service Management, the *International Journal of Network Management*, and the *Journal of Communications and Networks* and an Associate Technical Editor of the *IEEE Communications Magazine*.