IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 1 July 2022, accepted 15 July 2022, date of publication 26 July 2022, date of current version 8 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3193941

RESEARCH ARTICLE

Digital Twin Model: A Real-Time Emotion Recognition System for Personalized Healthcare

BARATHI SUBRAMANIAN[1], JEONGHONG KIM[1], MOHAMMED MARAY[2], (Member, IEEE), AND ANAND PAUL[1], (Senior Member, IEEE)

[1]School of Computer Science and Engineering, Kyungpook National University, Daegu 41566, South Korea
[2]College of Computer Science and Information Systems, King Khalid University, Abha 61421, Saudi Arabia

Corresponding author: Jeonghong Kim (jhk@knu.ac.kr)

This work was supported in part by the National Research Foundation of Korea under Grant 2020R1A2C1012196; and in part by the School of Computer Science and Engineering, Ministry of Education, Kyungpook National University, South Korea, through the BK21 Four Project, AI-Driven Convergence Software Education Research Program, under Grant 4199990214394.

This work involved human subjects or animals in its research. The author(s) confirm(s) that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT Emotion recognition (ER) in healthcare has drawn substantial attention owing to recent advancements in machine-learning (ML) and deep-learning (DL) techniques. The ER system, along with a digital twin of a person in real time, will facilitate the monitoring, understanding, and improvement of the physical entity's capabilities, as well as provide constant input to improve quality of life and well-being for personalized healthcare. However, building such ER systems in real time involves technical challenges, such as limited datasets, occlusion and lighting issues, identifying important features, false classification of emotions, and high implementation costs. To resolve this issue, we built a simple, efficient, and adaptable ER system by acquiring and processing images in real time using a web camera. In addition, we propose an end-to-end framework that combines an ER system with a digital twin setup, in which the predicted result can be analyzed and tested prior to providing the best possible personal healthcare treatment before it leads to any life-threatening disease. Our proposed ER system achieved promising results in less training time without compromising the accuracy. Thus, in real time, it will be helpful in healthcare centers to monitor a patient's health condition, early diagnosis of life-threatening diseases, and to obtain the best and most effective treatment for patients during emergencies.

INDEX TERMS Deep learning, emotion recognition, intelligence system, MediaPipe, smart healthcare system.

I. INTRODUCTION

Over the past few decades, several studies have focused on deducing human emotions from facial expressions, speech/voice expressions, and various body postures. Automatic detection and classification of different expressions and postures is a challenging problem in many fields, such as human-computer interaction, computer vision, and image processing. With recent developments in human-computer interaction (HCI), understanding human emotions using computers has become an essential issue. In addition, the need for real-time analysis of human emotions is increasing in many fields of study, such as psychology, safety, health, games, and robotics. Especially, in medical field, during the time of emergency, it helps doctors to identify the patient's psychological issues and initiate appropriate treatment in effect. For the last 200 years, a good therapy method has been decided based on the positive outcome obtained from testing and trialing drugs on certain people. From the current situation, we can clearly see that common treatments provided

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

for all people are not successful, as we are all unique and susceptible to different types of diseases (e.g., different types of cancers). The manner in which one responds to a certain drug also depends on one's emotional state. We have seen cyber-physical systems for personalized healthcare applications, but until now, no study has focused on personalized healthcare based on emotion-aware intelligence with respect to patients, and this paper takes the initiative to address this issue. Emotional intelligence in healthcare generally applies to physicians' emotional states, prioritizing their patient's urgency and providing necessary services for better treatment. This can also benefit patients [1].

We make judgements about other people's traits based on various forms of nonverbal communication, including their facial expression and tone/stress of voice, leading to the conclusion that he seems to be sad, he sounded really agitated or had no energy, he looked energetic, and so on [2]. This is how humans perceive different emotions, and they are quite good at this [3], [4]. According to a recent report from the World Health Organization, approximately 40% of patients globally are mistreated with medical errors with access to more data, and technology healthcare providers have an opportunity and responsibility to transform patients with personalized healthcare plans. The way we view the patients is constantly evolving from a traditional data analytic perspective, which was historically all about the data we could derive from electronic health records. Nowadays, we can derive more data from personal wearable devices and healthcare devices, which can help us to narrow down our treatment approach to the patient. However, we can obtain a better perspective of the patient when we obtain their emotional state such that treatment demography with patient's history, we can now build a sort of longitudinal record and provide personalized health plans that benefit the patient. This can provide an increasing number of insights about the patient; for example, even interactions the patients had with their healthcare provider could reveal a lot about the patient that we can exploit to care for them.

In this study, we propose a real-time ER system to classify and identify the emotional state of a patient using facial expressions and posture analysis of image/video files for personalized healthcare applications. Various details of posture and face are characterized in terms of feature representation, which can then be deduced to determine emotions, such as whether a person is satisfied, sad, fearful, happy, etc. In addition, we propose an end-to-end emotion-aware decision support framework that concatenates the ER system with a digital twin setup for effective personalized healthcare. The main contributions of this paper are summarized below:

- We created a custom dataset in real time by capturing emotions such as aggressiveness, shock, anxiety/unfriendliness, and openness/willingness, in addition to the six basic universal emotions with various backgrounds, lighting, nationality, and gender using a single web camera.
- We have also incorporated the MediaPipe pipeline framework for pre-processing images and not to detect uncommon features by obtaining landmark keypoints from the face, hands, and body[5]. Thus, the results increased the recognition rate of the proposed ER system model.
- A real-time ER system for personalized healthcare was trained and tested using different ML algorithms that efficiently classify and recognize user emotions in real time without the need for human interpretation in a cost-effective manner.

The rest of this paper is organized as follows: Section 2 introduces existing ER methods and their limitations. Section 3 introduces and describes the proposed methodology. Section 4 presents the experimental settings, results, comparison of our method with other methods and analyzes the model performance. Finally, section 5 presents our contributions and outlines the future work.

II. RELATED WORK

Recent technological advancements in healthcare have attracted attention for building SHC systems [6] during medical emergencies [7]–[9]. In particular, human–computer interaction (HCI) is an active field in emotion recognition as it helps humans to communicate with computers easily. Humans try to interpret others' emotions through emotional behaviors, namely speech and linguistic aspects [10], [11], facial expressions [12]–[14], and physiological signals [15] that originate in the peripheral nervous system and central nervous system dynamics, by observing body postures at different emotional states and intensities of the expressions [16]–[20]. Mehrabian in 1967 established the fact of "3v rule" that says 55% of the communication is visual, 7% verbal and 38% vocal [2]. Although it highlights the importance of nonverbal communication, many challenges remain unsolved and need to be addressed and improved in SHS. This helps doctors provide efficient treatment for patients and vice versa, even during medical emergencies. This section presents a brief review of recent studies on face emotion recognition (FER), speech emotion recognition (SER) and posture emotion recognition (PER).

A. FACE EMOTION RECOGNITION (FER)

The primary means to identify a person is through their face appearance, as it conveys rich information such as a person's age, sex, race, and personality, and is considered to be an important method for visual communication of emotions. Emotions are understood mainly through facial expressions, and FER permits us to understand, react to a large number of stimuli, and in the prognosis of prosocial conduct [21]. Brain damage and neuronal deterioration can cause a deficit in FER, leading to the loss of effective communication. Thus, FER plays a vital role in the early diagnosis of diseases such as Alzheimer's, Parkinson's disease [22], schizophrenia and. In 1971, Ekman and Friesen [23] conducted the first research on FER and stated that humans can universally express six basic emotions: anger, disgust, fear, sadness, happiness, and surprise. They also designed a facial action coding system (FACS) [24], [25], which provides mapping between the emotions and possible movements of the facial muscles. Currently, many datasets are available for research purposes, such as Ekman's FACS facial expression datasets, the Japanese female expression database (JAFFE), the Cohn-Kanade (CK+) expression database [26], and the Facial Expression Recognition 2013 (FER-2013) dataset [27].

In general, emotion recognition involves three major steps: detection, feature extraction, and emotion classification. Several methods have been employed to find the correct position of the face from an image, namely, Haar classifiers [28], Ada-Boost [29], and adaptive skin color models [30] for the discrete emotion model. In addition to this discrete model, the author of [31] proposed a continuous FER model for videos based on an evaluation-activation 2D model. Feature extraction in FER is useful for extracting important features from face-like eyebrows, eyes, nose, and mouth and for accurate recognition of facial expressions. Scholars have used traditional methods, such as geometric features for local binary patterns [32], Gabor wavelet transformation [33], and local directional patterns [34] for facial feature extraction. However, when we compared the overall recognition rate of the traditional methods, it was not high. In recent years, researchers have focused on the DL approach because of its deep and advanced network architectures, such as convolutional neural networks (CNN) for driver emotion classification systems [35], recurrent neural networks (RNN), and time series data. Emotion classification examines the correlation between expression functions and allots them to the subsequent classes. The classification stage generally employs an ML algorithm that uses the extracted feature vector as input and yields the resulting emotions. The most common algorithms used for emotion classification are support vector machines (SVM) [36] owing to their favorable performance, neural networks, and hidden Markov models.

B. POSTURE EMOTION RECOGNITION (PER)

It is estimated that approximately 60–65% of all communication occurs through body language; thus, postures and gestures are effective means of communication, along with speech and facial expression [37]. Postures can effectively convey emotional information and coordinate the speech content. If we can understand someone's body language, it reveals much about their emotional state. It is true that we engage in some form of conversation with another person through eye contact, such as eye gaze and blinking, which can reveal whether a person is distracted, threatened, or concealed feeling or just feel uncomfortable. In this article, we do not focus on eye contact and mouth expression (such as lip biting); rather, we observe gestures and posture to detect emotion and thereby provide personalized healthcare services. Arms, legs, personal space, and manner of communication can also define certain emotional states. While sitting in a closed posture indicates anxiety, unfriendliness and open posture reveal one person's willingness to engage in discussion and openness. Upright posture is often judged as having more positive connotations, while forward leaning posture is defined as negative. Even though the above findings on these posture settings outline a useful framework, they tend to focus on general, frequency dynamic properties of bodies and fail to make clear predictions regarding the specific posture that may be associated with different emotional states [38].

III. PROPOSED METHOD AND FRAMEWORK

Our proposed emotion-aware intelligence decision support system framework for personalized healthcare is based on three major steps in emotion recognition: i) data acquisition, ii) detection and feature extraction, and iii) emotion recognition and classification, as shown in Fig. 1 Once the emotion is detected and classified, healthcare professionals can analyze the data considering the various expressions of the patients and the history of their health record. This helps physicians to diagnose the cause of the disease early. Further, the drug/treatment will be tested on their digital twin for successful physical treatment and effective medical procedures will be provided or recommended according to individual interests and situations. To have better understanding, the overall workflow of our proposed end-to-end emotion-aware decision support framework is shown in Fig. 2, and the following sections explain each of those mentioned above in detail.

A. DATA ACQUISITION

Emotion recognition systems require a large amount of data to train using ML and DL algorithms. Accurate predictions can be achieved when data are diverse. However, the public image/video datasets that are available for emotion recognition are not sufficiently diverse in terms of race and gender and contain limited emotional expressions. To overcome these issues, we created our own dataset for ten different emotions combined with existing universal emotions. RGB images were captured using a 640×480 pixel web camera in various angles, with different lighting backgrounds. Furthermore, to make the model robust, three volunteers were used to perform all the ten emotions including a male from Uzbekistan and two females from different regions of India. To make the model even more effective and unique, we used body postures matching facial expressions when capturing the images. For example, while capturing data for willingness/openness, we captured the entire body by keeping the trunk of the body open.

B. DETECTION AND FEATURE EXTRACTION

Once the data were collected from the web camera, they were passed through the MediaPipe framework to detect and extract features from the frames for FER and PER recognition. MediaPipe is an open-source ML framework that is used for real-time computer vision. The framework combines three individual models for the detection of face, hand, and body pose that extracts 543 landmark keypoints altogether [5], as shown in Fig. 3. The extracted landmarks will have three

FIGURE 1. Overview of our proposed end-to-end framework for personalized healthcare.

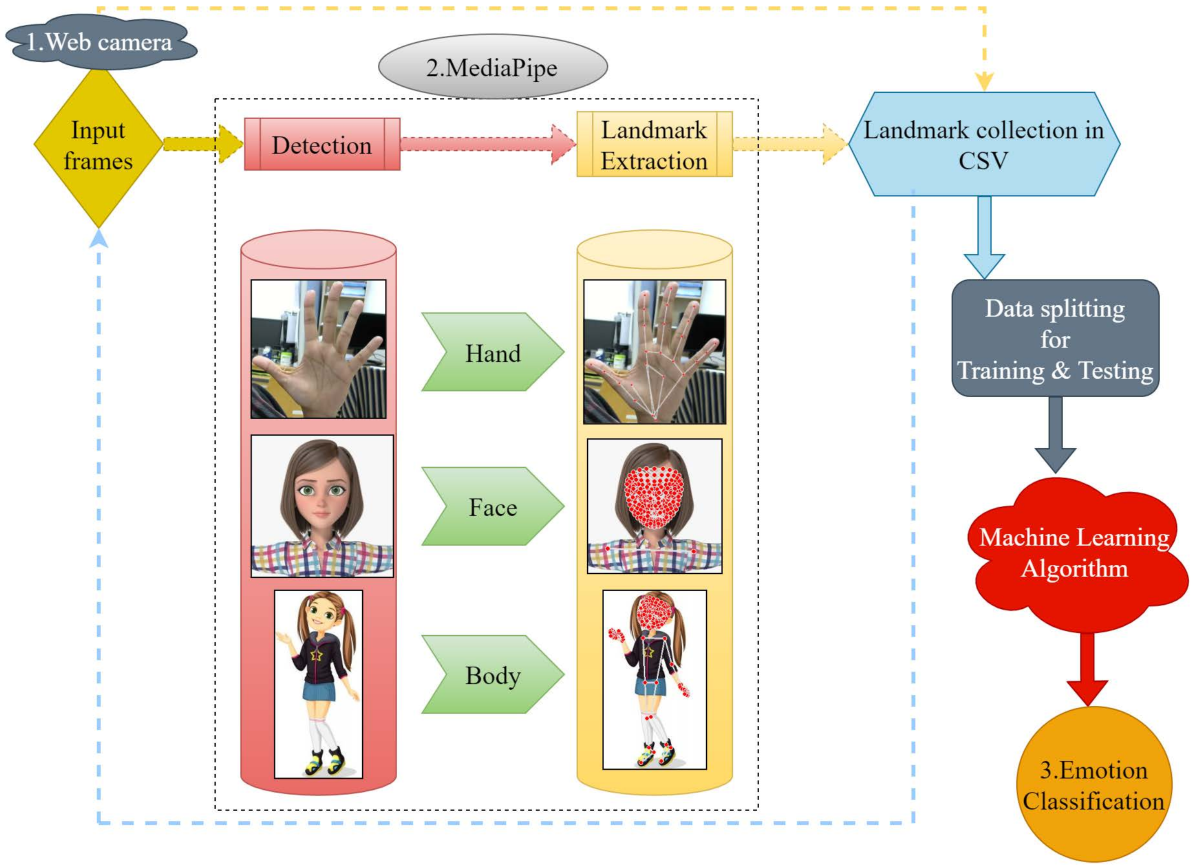


FIGURE 2. Proposed architecture for Realtime ER system with 1. Data acquisition, 2. Detection and feature extraction using MediaPipe and 3.Emotion classification.

FIGURE 3. Result samples taken from Mediapipe framework with extracted 543 keypoints/landmarks from face/hands and body.

coordinates (x, y, z), which makes them work better in any environment by setting up a threshold value for the visibility of other objects. After detecting and obtaining the coordinate values, they are stored in a CSV file format for easy and fast training by sending them to the ML algorithm for emotion classification. We then split the data into training and testing sets. To detect the face, input data are preprocessed because they contain many irrelevant variations, such as background, image illumination, and body poses. We used normalization techniques that used median filters to reduce the illumination and variations in the input images with image quality improvement. In addition, we located the face in the picture and identified the points of interest before passing it for feature extraction. From the experiment, we observed that, MediaPipe performed well for feature detection and extraction, regardless of the background or the environment compared to other feature extraction methods used in the existing works as mentioned in the literature. Also, implementation of this framework is very simple as it does not require any expensive external devices or higher-resolution cameras, which makes it cost-effective and efficient for real time usage.

C. EMOTION RECOGNITION AND CLASSIFICATION

In general, for image processing and classification problems, the SVM algorithm, a linear classification technique, is widely used and has achieved promising results [36], [39].

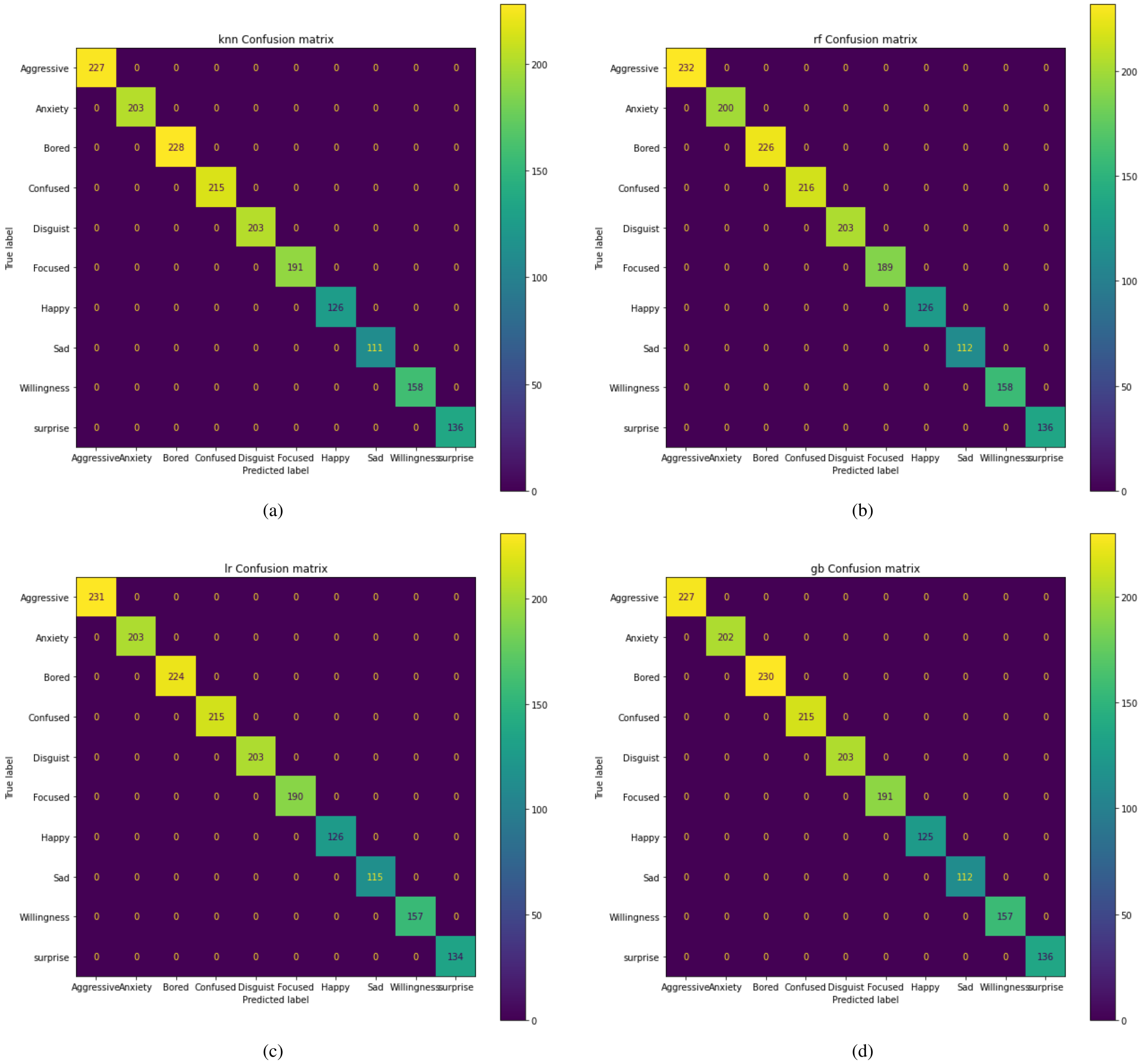


FIGURE 4. Confusion matrix (a) K-Nearest Neighbor; (b) Random Forest; (c) Logistic Regression; and (d) Gradient Boosting.

However, the time taken for classification is low compared with other emotion recognition and classification methods. Similar to SVM, several other classification algorithms are used, such as K-nearest neighbor (KNN), Bayesian classifier (naive Bayes), random forest (RF), logistic regression, and gradient boosting (GB). Because we wanted to build an efficient emotion-aware intelligent support system for health care, we trained the model using eight different machine algorithms and thus selected a gradient boosting algorithm for real-time detection, as it outperformed other algorithms. Finally, our classifier classified the emotion detected by the camera in real time as one out of ten emotions.

IV. RESULTS AND ANALYSIS

A. DATASET

A real-time dataset of total 5,991 labeled images were created with three volunteers performing 10 different expressions, as described in Table 1. Each image was captured using a web camera with a size of 640 × 480 pixels, in different positions and various lighting conditions and stored it in the form of CSV file. We separated the collected dataset in the ratio of 70:30 to form the corresponding training and testing datasets.Thus we used 4,193 images for training and 1,798 images for testing.

B. EXPERIMENTAL SETTING

The simulation was conducted using Python 3.7 version on a desktop computer with 32 GB RAM and an Intel Core i7 processor with a frequency of 3.60 GHz, running on Windows 10 Pro with a 64-bit operating system. The input image was captured using a web camera with a resolution of 720pixels/30 fps of RGB images. This shows that implementing our proposed approach does not require any expensive

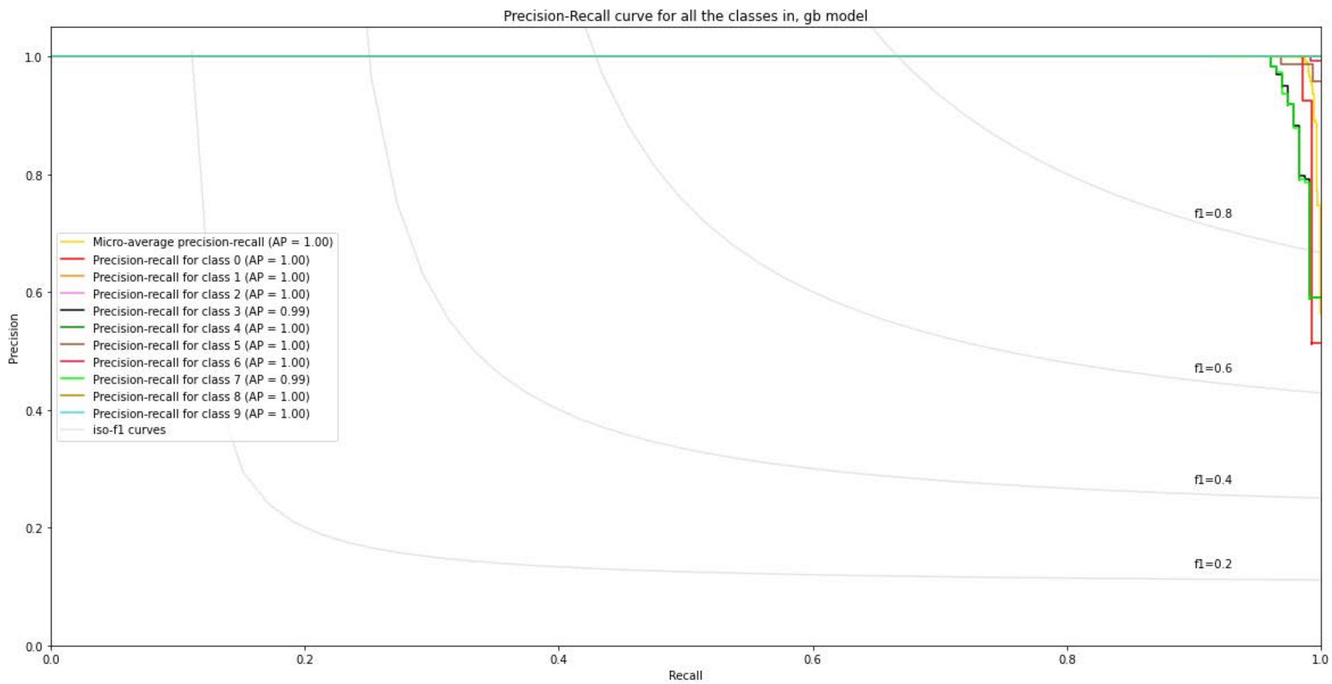


FIGURE 5. Precision, Recall and F1 score curves for all ten classes in GB model.

TABLE 1. Custom dataset for emotion recognition.

CLASSES	EMOTIONS	SAMPLES
Class 0	HAPPY	991
Class 1	SAD	553
Class 2	AGGRESSIVE/ANGRY	421
Class 3	FOCUSED/PAY ATTENTION	556
Class 4	BORED	646
Class 5	SHOCK/SURPRISE	392
Class 6	ANXIETY/UNFRIENDLINESS	655
Class 7	OPENESS/WILLINGNESS	515
Class 8	CONFUSED	512
Class 9	DISGUIST	412

TABLE 2. Average classification accuracy obtained using various ML algorithms.

Algorithm	Accuracy
Logistic Regression	99.1%
Random Forest	99.6%
Gradient Boosting	**99.9%**
KNN	99.7%
SVM	98.1%
Decision Tree Classifier	98.6%
Naive Bayes	76.5%
Ridge Regression	98.1%

external devices or GPU setup for training the data. We ran the simulations to the maximum extent for attaining best accuracy out of all machine learning algorithm pipelines.

C. QUANTITATIVE ANALYSIS

Final outcome of the experiment is presented in Table 2. Further, to extensively analyze the results of the candidate

TABLE 3. Classification report based on various emotions for GB model.

Emotion	Precision	Recall	F1-score
Aggressive	1.00	1.00	1.00
Anxiety	1.00	1.00	1.00
Bored	1.00	1.00	1.00
Confused	1.00	1.00	1.00
Disguist	1.00	1.00	1.00
Focused	1.00	1.00	1.00
Happy	1.00	0.99	1.00
Sad	0.99	1.00	1.00
Willigness	1.00	0.99	1.00
Surprise	1.00	1.00	1.00
Accuracy			**1.00**
Macro-Avg	**1.00**	**1.00**	**1.00**
Weighted-Avg	**1.00**	**1.00**	**1.00**

algorithms, we build confusion matrix, as shown in Fig. 4, to calculate classification performance metrics, such as accuracy, precision, recall, and F1 score, based on the following four values: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The number of accurately predicted data points out of all the data points is called accuracy; ideally, it must be close to one. The following formula (Equation (1)) was used to determine the accuracy of the model:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

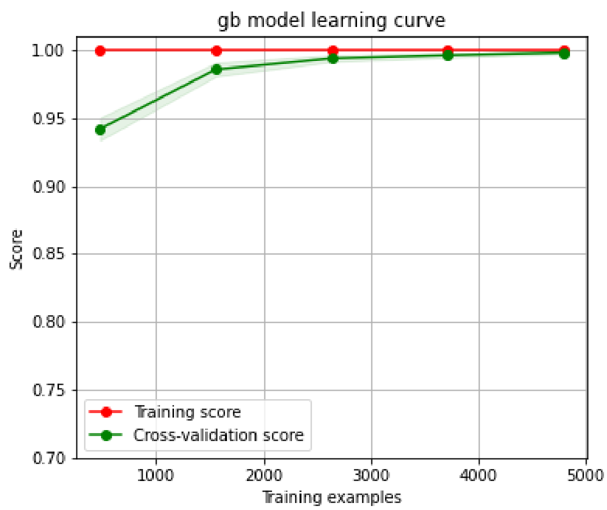


FIGURE 6. GB model learning curve.

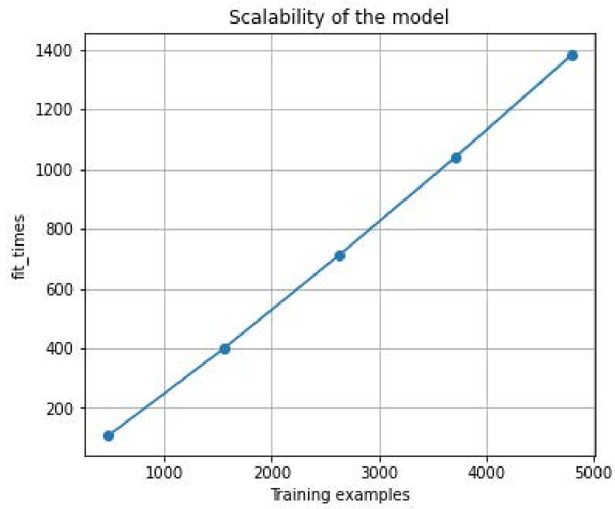


FIGURE 7. GB model scalability.

The positive projected value was equal to the number of positive predictions divided by the total number of predicted positive class values. This, similar to the accuracy score, should be as close to one as possible. The mathematical formula shown in Equation (2) is used to calculate the precision:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

The fraction of positive occurrences anticipated by the model is called recall, and is calculated as shown in Equation (3):

$$Recall = \frac{TP}{FP + FN} \tag{3}$$

The F1-score in Equation (4), also known as the F-measure, is the harmonic mean of precision and recall, implying that it provides a balance between memory and precision. The F1 score reaches its highest value of 1 when precision and recall are flawless, and it can be computed using the following formula:

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{4}$$

Table 3 summarizes the performance of our best algorithm, GB, on the test set, which achieves an average F1-score and standard deviation of 1.00. The algorithm obtains the best performance for almost all the emotions.

We examined the confusion matrix for the top four classifiers, as shown in Fig. 4, to obtain a better understanding of the types of errors made by our classifier. The number of correct and incorrect predictions are summarized for the top four performers, K-Nearest Neighbor (Fig. 4(a)), Random Forest (Fig. 4(b)), Logistic Regression (Fig. 4(c)), and Gradient Boosting (Fig. 4(d)), in the experiment.

Precision-recall (PR) curve showed in Fig. 5 explains that the higher area under the curve (AUC), the higher the precision (low FP rate) and high recall (low FN rate). It is primarily

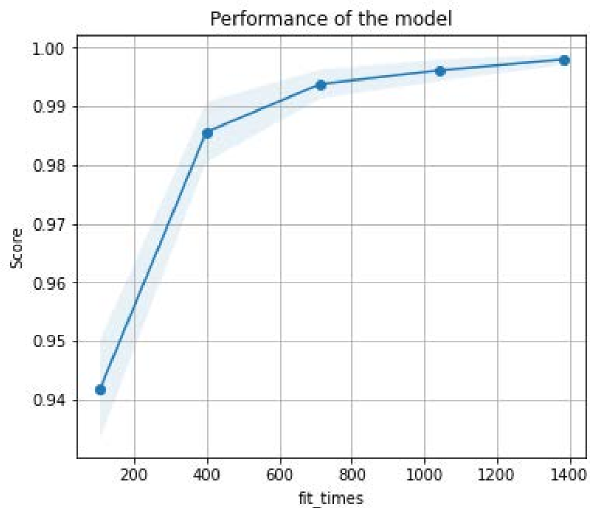


FIGURE 8. GB model performance.

used for the evaluation of binary classification algorithms. Therefore, by using the label_binarize() function, multiclass labels are converted to binary labels. The PR curve slops downward because of the precision loss, as it allows more and more predictions for more recall. This high average precision means that the estimated precision along the curve is at its peak value.

D. PERFORMANCE ANALYSIS

To estimate the model performance and check the prediction result on the new data, cross-validation was performed. This played a vital role in the early identification of over-fitting or under-fitting our data. We also plotted the learning curve to show the efficiency of our selected model out of the eight ML algorithms. The training and cross-validation scores for our GB model were both high. The training score is maintained

FIGURE 9. **Real-time emotion recognition for the emotions: Bored, Surprise, Disgust, Happy, Aggressive and Confused.**

at a score of 1, whereas the cross-validation score starts from approximately 0.95 and increases to a 0.99 score, as shown in Fig. 6. From Fig. 7, we can conclude that as the training data increases, more time is required to fit the data. The model performance plot in Fig. 8 shows that the model can achieve a high score with an increase in time.

E. EMOTION DETECTION IN REAL-TIME

When developing real-time applications, understanding how many frames per second can be handled is crucial. As part of the experiment, we used a single camera with the resolution of 640 * 480-pixel to test the proposed system, and the average number of executed frame rate was 15 frames per second (fps). In terms of human emotion, it is obvious that there is little difference among three successive frames. Because of this, we do not need to extract each individual's emotions frame by frame. Therefore, the proposed system only requires 15 fps (Not 30 frames) to ensure human emotion tracking. In order to test whether our proposed system is accurate in real time, we asked three volunteers of different gender and nationality to show some emotions like surprise, disgust, happy, bored, confused and aggressive in front of the camera. As shown in Fig. 9, our ER system trained on the ML algorithm produces an average accuracy of 99% for all the emotions detected in real time.

Our ER model is suitable for deployment in mobile applications because the trained model is lightweight. The proposed methodology is fast, robust, accurate, and customizable for real-time emotion detection. The state-of-the-art MediaPipe simplifies feature extraction by breaking down and analyzing complex data without the need to create a CNN from scratch. The proposed solution uses the smallest amount of computational resources and devices. Hence, from the accurate results produced by the ER system, physicians/doctors can diagnose the disease early, during emergencies, or even if the person is mute, and can recommend the patient for an effective treatment by testing the drug on their digital twin prior to the physical treatment.

V. CONCLUSION

Recognition of patient emotions has become increasingly important as it helps in many different areas, including the medical sector for personalized healthcare. The ability to identify an individual's emotions can help healthcare centers build smart diagnostic tools such as an ER system that can detect depression and stress among the patients in the early stages so that the medication can be given in prior. However, building such ER systems in real time involves lot of technical challenges. Hence to solve this, we built an efficient, simple, and customized ER system by acquiring and processing images in real-time using a single web camera. In addition, we proposed an end-to-end emotion-aware decision support framework that combines an ER system with a digital twin setup in which the predicted result can be analyzed and tested in advance to provide the best possible personal healthcare treatment before any life-threatening disease develops. Our proposed ER system has shown promising outcomes, with an average accuracy of 99%. The model's efficiency was also demonstrated to be better than that of other works through faster real-time detection. In the future, the work can be extended by adding speech/text emotion information along with facial and body emotions using MediaPipe's state-of-art and best possible classification algorithms.

DECLARATIONS

CONFLICT OF INTEREST

To the best of our knowledge, the named authors have no conflict of interest, financial or otherwise.

CONSENT TO PARTICIPATE

Informed consent was obtained from the subject for publication of identifying information/images in an online open-access publication

REFERENCES

[1] N. Libbrecht, F. Lievens, B. Carette, and S. Côté, "Emotional intelligence predicts success in medical school," *Emotion*, vol. 14, no. 1, pp. 64–73, Feb. 2014, doi: 10.1037/a0034392.

[2] A. Mehrabian, "Communication without words," Tech. Rep., 1968.

[3] F. Noroozi, M. Marjanovic, A. Njegus, S. Escalera, and G. Anbarjafari, "Audio-visual emotion recognition in video clips," *IEEE Trans. Affect. Comput.*, vol. 10, no. 1, pp. 60–75, Jan./Mar. 2019, doi: 10.1109/TAFFC.2017.2713783.

[4] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognit.*, vol. 36, no. 1, pp. 259–275, Jan. 2003, doi: 10.1016/S0031-3203(02)00052-3.

[5] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang, and M. Grundmann, "MediaPipe hands: On-device real-time hand tracking," 2020, *arXiv:2006.10214*.

[6] A. Ganesan, A. Paul, G. Nagabushnam, and M. J. J. Gul, "Human-in-the-loop predictive analytics using statistical learning," *J. Healthcare Eng.*, vol. 2021, pp. 1–14, Jul. 2021, doi: 10.1155/2021/9955635.

[7] A. A. Abdellatif, A. Z. Al-Marridi, A. Mohamed, A. Erbad, C. F. Chiasserini, and A. Refaey, "SsHealth: Toward secure, blockchain-enabled healthcare systems," *IEEE Netw.*, vol. 34, no. 4, pp. 312–319, Jul. 2020, doi: 10.1109/MNET.011.1900553.

[8] A. A. Abdellatif, A. Mohamed, C. F. Chiasserini, M. Tlili, and A. Erbad, "Edge computing for smart health: Context-aware approaches, opportunities, and challenges," *IEEE Netw.*, vol. 33, no. 3, pp. 196–203, May/Jun. 2019, doi: 10.1109/MNET.2019.1800083.

[9] A. Z. Al-Marridi, A. Mohamed, and A. Erbad, "Convolutional autoencoder approach for EEG compression and reconstruction in m-Health systems," in *Proc. 14th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2018, pp. 370–375, doi: 10.1109/IWCMC.2018.8450511.

[10] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 6, pp. 1113–1133, Jun. 2015, doi: 10.1109/TPAMI.2014.2366127.

[11] K. Han, D. Yu, and I. Tashev, "Speech emotion recognition using deep neural network and extreme learning machine," in *Proc. Annu. Conf. Int. Speech Commun. Assoc. (INTERSPEECH)*, Sep. 2014, pp. 223–227, doi: 10.21437/interspeech.2014-57.

[12] T. Abegaz, G. Dozier, K. Bryant, J. Adams, K. Popplewell, J. Shelton, K. Ricanek, and D. L. Woodard, "Hybrid GAs for eigen-based facial recognition," in *Proc. IEEE Workshop Comput. Intell. Biometrics Identity Manage. (CIBIM)*, Apr. 2011, pp. 7–10.

[13] W. Yi Zhao and R. Chellappa, "SFS based view synthesis for robust face recognition," in *Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Mar. 2000, pp. 285–292, doi: 10.1109/AFGR.2000.840648.

[14] B. Gökberk, A. A. Salah, and L. Akarun, "Rank-based decision fusion for 3D shape-based face recognition," in *Proc. Int. Conf. Audio Video-Based Biometric Person Authentication*, in Lecture Notes in Computer Science, vol. 3546, 2005, pp. 1019–1028, doi: 10.1007/11527923_106.

[15] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, and X. Yang, "A review of emotion recognition using physiological signals," *Sensors*, vol. 18, no. 7, p. 2074, Jun. 2018, doi: 10.3390/S18072074.

[16] A. Brogi *et al.*, "Survey on AI-based multimodal methods for emotion detection," *Futur. Gener. Comput. Syst.*, vol. 29, no. 1, pp. 1–18, 2018, doi: 10.1007/978-3-030-16272-6.

[17] S. Singh, V. Sharma, K. Jain, and R. Bhall, "EDBL–algorithm for detection and analysis of emotion using body language," in *Proc. 1st Int. Conf. Next Gener. Comput. Technol. (NGCT)*, Sep. 2015, pp. 820–823, doi: 10.1109/NGCT.2015.7375234.

[18] C. Ma, Q. Liu, and Y. Dang, "Multimodal art pose recognition and interaction with human intelligence enhancement," *Frontiers Psychol.*, vol. 12, pp. 1–13, Nov. 2021, doi: 10.3389/fpsyg.2021.769509.

[19] Z. Yang, A. Kay, Y. Li, W. Cross, and J. Luo, "Pose-based body language recognition for emotion and psychiatric symptom interpretation," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 294–301, doi: 10.1109/ICPR48806.2021.9412591.

[20] K. Schindler, L. Van Gool, and B. de Gelder, "Recognizing emotions expressed by body pose: A biologically inspired neural model," *Neural Netw.*, vol. 21, no. 9, pp. 1238–1246, Nov. 2008, doi: 10.1016/j.neunet.2008.05.003.

[21] N. C. Ebner, M. K. Johnson, and H. Fischer, "Neural mechanisms of reading facial emotions in young and older adults," *Frontiers Psychol.*, vol. 3, pp. 1–19, Jul. 2012, doi: 10.3389/fpsyg.2012.00223.

[22] J. Pietschnig, R. Aigner-Wöber, N. Reischenböck, I. Kryspin-Exner, D. Moser, S. Klug, E. Auff, P. Dal-Bianco, G. Pusswald, and J. Lehrner, "Facial emotion recognition in patients with subjective cognitive decline and mild cognitive impairment," *Int. Psychogeriatrics*, vol. 28, no. 3, pp. 477–485, Mar. 2016.

[23] P. Ekman, "Universal facial expressions of emotion," *California Mental Health*, vol. 8, no. 4, pp. 151–158, 1970.

[24] S. Naik and R. P. K. Jagannath, "GCV-based regularized extreme learning machine for facial expression recognition," in *Advances in Machine Learning and Data Science*, vol. 705. Jan. 2018.

[25] P. Ekman and W. V. Friesen, "Facial action coding system," *Man*, 2002.

[26] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn–Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPRW Workshops)*, Jun. 2010, pp. 94–101, doi: 10.1109/CVPRW.2010.5543262.

[27] I. J. Goodfellow *et al.*, "Challenges in representation learning: A report on three machine learning contests," *Neural Netw.*, vol. 64, pp. 59–63, Apr. 2015, doi: 10.1016/j.neunet.2014.09.005.

[28] S. L. Happy and A. Routray, "Robust facial expression classification using shape and appearance features," in *Proc. 8th Int. Conf. Adv. Pattern Recognit. (ICAPR)*, Jan. 2015, pp. 1–5, doi: 10.1109/ICAPR.2015.7050661.

[29] S. Kang, B. Choi, and D. Jo, "Faces detection method based on skin color modeling," *J. Syst. Archit.*, vol. 64, pp. 100–109, Mar. 2016, doi: 10.1016/j.sysarc.2015.11.009.

[30] C. Wang, "Human emotional facial expression recognition," pp. 1–8, 2018, *arXiv:1803.10864*.

[31] M. Zöllner *et al.*, "Recognizing emotions from video in a continuous 2D," vol. 6949, pp. 584–587, 2011, doi: 10.1007/978-3-642-23768-3.

[32] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image Vis. Comput.*, vol. 27, no. 6, pp. 803–816, 2009, doi: 10.1016/j.imavis.2008.08.005.

[33] S. Zhang, L. Li, and Z. Zhao, "Facial expression recognition based on Gabor wavelets and sparse representation," in *Proc. IEEE 11th Int. Conf. Signal Process.*, vol. 2, Oct. 2012, pp. 816–819.

[34] T. Jabid, M. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern," *ETRI J.*, vol. 32, no. 5, pp. 784–794, 2010, doi: 10.4218/etrij.10.1510.0132.

[35] R. A. Naqvi, M. Arsalan, A. Rehman, A. U. Rehman, W.-K. Loh, and A. Paul, "Deep learning-based drivers emotion classification system in time series data for remote applications," *Remote Sens.*, vol. 12, no. 3, p. 587, Feb. 2020.

[36] M. S. Bilkhu, S. Gupta, and V. K. Srivastava, "Emotion classification from facial expressions using cascaded regression trees and SVM," in *Computational Intelligence: Theories, Applications and Future Directions*. Springer, 2019, pp. 585–594.

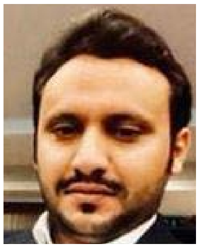
[37] G. N. Foley and J. P. Gentile, "Nonverbal communication in psychotherapy," *Psychiatry, Edgemont*, vol. 7, no. 6, pp. 38–44, 2010, doi: 10.4324/9781315134925-16.

[38] M. Coulson, "Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence," *J. Nonverbal Behav.*, vol. 28, no. 2, pp. 117–139, 2004, doi: 10.1023/B:JONB.0000023655.25550.be.

[39] M. Abdulrahman and A. Eleyan, "Facial expression recognition using support vector machines," in *Proc. 23rd Signal Process. Commun. Appl. Conf. (SIU)*, May 2015, pp. 276–279.

BARATHI SUBRAMANIAN is currently pursuing the Ph.D. degree in computer science and engineering with Kyungpook National University, South Korea. Her research interests include sign language recognition and computer vision using machine learning and deep learning techniques.

JEONGHONG KIM received the B.S. and M.S. degrees from Kyungpook National University, Daegu, South Korea, in 1986, and the Ph.D. degree from Chungnam National University, Daejeon, South Korea, in 2001. He worked as a Senior Researcher with the Electronics and Telecommunications Research Institute, from 1988 to 1996. He worked as a Professor with Sangju National University, from 1996 to 2008. He is currently working as a Professor with the School of Computer Science and Engineering, Kyungpook National University. His current research interests include bio signal processing and pattern recognition using deep learning techniques.



MOHAMMED MARAY (Member, IEEE) received the Ph.D. degree from Warwick University, U.K. He is currently an Assistant Professor with King Khalid University, Saudi Arabia, and working as an Assistant Researcher at Petras Project 2020. His current research interests include cloud, the IoT, fog and edge networks, computational offloading, MEC, and MCC. He is a member of ACM, CISSP, and Saudi Academy of Engineering.

ANAND PAUL (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the National Cheng Kung University, Tainan, Taiwan, in 2010. He is currently working as a full-time Associate Professor with the School of Computer Science and Engineering, Kyungpook National University, Daegu, South Korea. He is a delegate representing Korea for M2M focus group and for MPEG. His research interests include algorithm and architecture re-configurable embedded computing. He has guest edited various international journals.