LSTM-Based Emotion Detection Using Physiological Signals: IoT Framework for Healthcare and Distance Learning in COVID-19

Muhammad Awais, *Member, IEEE*, Mohsin Raza, *Member, IEEE*, Nishant Singh, Kiran Bashir, Umar Manzoor, Saif ul Islam, and Joel J. P. C. Rodrigues, *Fellow, IEEE*

Abstract—Human emotions are strongly coupled with physical and mental health of any individual. While emotions exibit complex physiological and biological phenomenon, yet studies reveal that physiological signals can be used as an indirect measure of emotions. In unprecedented circumstances alike the coronavirus (Covid-19) outbreak, a remote Internet of Things (IoT) enabled solution, coupled with AI can interpret and communicate emotions to serve substantially in healthcare and related fields. This work proposes an integrated IoT framework that enables wireless communication of physiological signals to data processing hub where long short-term memory (LSTM)-based emotion recognition is performed. The proposed framework offers real-time communication and recognition of emotions that enables health monitoring and distance learning support amidst pandemics. In this study, the achieved results are very promising. In the proposed IoT protocols (TS-MAC and R-MAC), ultralow latency of 1 ms is achieved. R-MAC also offers improved reliability in comparison to state of the art. In addition, the proposed deep learning scheme offers high performance (f-score) of 95%. The achieved results in communications and AI match the interdependency requirements of deep learning and IoT frameworks, thus ensuring the suitability of proposed work in distance learning, student engagement, healthcare, emotion support, and general wellbeing.

Index Terms—Artificial intelligence (AI), coronavirus (Covid-19), human emotion analysis, Internet of Things (IoT), long short-term memory (LSTM), wearable physiological signals.

Muhammad Awais and Mohsin Raza are with the Department of Computer Science, Edge Hill University, Ormskirk L39 4QP, U.K. (e-mail: mawais@ieee.org; mohsinraza119@gmail.com).

Nishant Singh is with the School of Psychology, University of Birmingham, Birmingham B15 2TT, U.K. (e-mail: nishant22jan@gmail.com).

Kiran Bashir is with the Department of Computer Science, Virtual University, Lahore 55150, Pakistan (e-mail: mc180400249@vu.edu.pk).

Umar Manzoor is with the Faculty of Science and Engineering, University of Hull, Hull HU6 7RX, U.K. (e-mail: u.manzoor@hull.ac.uk).

Saif ul Islam is with the Department of Computer Science, Institute of Space Technology, Islamabad 44000, Pakistan (e-mail: saiflu2004@gmail.com).

Joel J. P. C. Rodrigues is with the Post Graduate Program in Electrical Engineering, Federal University of Piauí, Teresina 64049, Brazil, and also with the Instituto de Telecomunicações, 1049-001 Lisboa, Portugal (e-mail: joeljr@ieee.org).

Digital Object Identifier 10.1109/JIOT.2020.3044031

I. INTRODUCTION

CORONAVIRUS outbreak (Covid-19) has hit the world hard with a dramatic effect on the way humans live and thrive. Human race is seeing unprecedented restrictions in performing the day-to-day activities. A large portion of the world population is forced to stay indoors and work from home. Healthcare and education are among the largely affected sectors. Remote teaching and healthcare facilities have become the need of the hour. This has brought forth with it the need to deliver online lectures and provide remote healthcare.

Among others, keeping a sound and healthy mental state is most important with people confined in their households. Another challenge is being faced by the tutors to maintain student's engagement throughout the lecture. This has also increased the need of state-of-the-art facility for pastoral support to the students. Since the students come from diverse social and ethnic backgrounds, their physical, mental, and emotional wellbeing need to be very well supported. This demands for the online teaching programs to be directed toward imparting teaching and knowledge along with a health and emotions monitoring system. A probable solution could be the development of a real-time emotion monitoring and analysis system.

Human emotions are the results of a conscious experience that takes place in an event and is characterized by the brain activity of individuals. Cognition is another branch of emotion and is highly applicable in artificial intelligence. Emotions are very complex entities that are difficult to characterize. Recently, with the advancement of human–computer interaction techniques and its demand, the importance of emotion recognition has suddenly increased.

The nature of human emotions is psychologically characterized by two major aspects: 1) valence and 2) arousal [1]. Fig. 1 represents the valence and arousal aspects of human emotions. Both these concepts are not mutually exclusive but are conceptually different frameworks. Valence refers to positive or negative, and arousal provides information about the silence level [1].

These emotions are reflections of an individual's mental state and recognizing such emotions has been very critical in understanding human thoughts. This has been applied in driving, mental, and social healthcare monitoring and so on. The use of physiological signals is thus a more reliable method
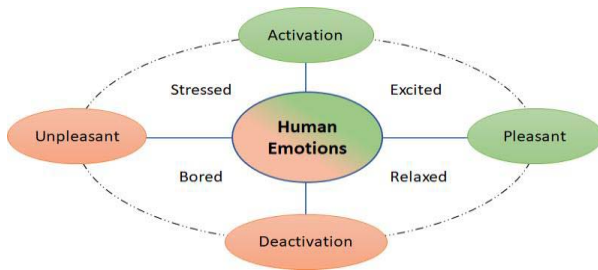
Fig. 1. Schematic representation of a 2-D valence and arousal.

to track emotions of individuals and their internal cognitive processes. Human physiology within the context of emotion recognition is imperative as these emotions are the results of the central nervous system (CNS) and the autonomic nervous system (ANS) activities, which cannot be imitated easily.

A. Human Emotion Analysis

Human emotion analysis has become a significant area in recent years in effective computing due to its strong relevance with physical and mental health, quality of life, and wellbeing. Among several other application domains, human emotion analysis has been applied significantly in human–computer interaction (HCI). Ideally, there are two ways by which human recognition is done: 1) speech signal analysis and facial recognition and 2) analysis of physiological signals. Physiological signals are often preferred over voice signals due to the strong association and these signals exhibit with physical and mental state of human beings [2].

The frequently used multimodal physiological sensing modalities are electroencephalogram (EEG), electrocardiogram (ECG), electromyogram (EMG), galvanic skin response (GSR), respiration (RSP), skin temperature (SKT), blood volume pressure (BVP), and Photoplethysmography (PPG). Researchers have used these modalities in the past, such as respiration, EEG, GSR, SKT, EMG, heart rate, and heart rate variability (HRV) derived from ECG, to recognize and track human emotion for different applications [3], [4]. Human emotions were elicited by movie clips, camera images, music, and through difficult mathematical tasks. For instance, Ayata et al. [3] developed a wearable computing device with GSR and PPG sensors. The device was used to detect valence and arousal emotions that were triggered by music tracks. Similarly, emotions induced through video recordings were classified by coupling facial expression with ECG respiration signals for emotion classification in [4].

Although there has been significant work on analysis of emotions with EEG, it does not compete with the other physiological signals in terms of wearability and viability. EEG systems often seem to struggle when it comes to portability and ease of use. This is not the case with the new wearables in the market that can be easily used for acquisition of physiological data (GSR, EMG, BVP, and SKT) and vital body stats. The multimodal fusion mechanism of

biosignal analysis other than EEG provides interesting outcomes upon analysis and the study in this article does this very well.

B. Artificial Intelligence for Emotion Analysis

The recent advancements in the domain of artificial intelligence (AI) and computational power have become a key part of any system recognizing human emotions. This is also since physiological signals are very complex by nature and interpreting the underlying behaviors from simple visualization or signal processing is not trivial. Researchers in the past have used a hybrid particle swarm optimization support vector machine (PSO-SVM) and the use of a hierarchical classifier-based emotion recognition system was proposed [5]. In recent years, the developments within the subbranch of AI called "deep learning" has been widely used to analyze physiological signals. This is because deep learning is a branch that is based on feature-free algorithms, which automatically computes complex features from signals, easily ignored by hand-crafted feature extraction techniques. Yin et al. [6] developed an ensemble deep learning method using EEG to characterize human emotions. Tripathi et al. [7] also developed a convolutional neural network (CNN)-based emotion recognition system using speech features. Moreover, facial recognition for analysis of human emotions had been used with the application of deep convolutional neural networks in [8]. The use of an autoassociative neural network-based face emotion recognition from the video [9] and deep transfer learning and score fusion was also carried out [10].

These advancements in emotions analysis have brought with it a new hope for better analysis and judgement of mental illness. Neurodegeneration can bring with it the deficits in emotion perception. Some of the neurodegenerative conditions that may be very well be affected with decreased emotion perception along with the advancement of the disease are Alzheimer's disease (AD), behavioral variant frontotemporal dementia (bvFTD), semantic variant primary progressive aphasia (svPPA), nonfluent variant primary progressive aphasia (nfvPPA), progressive supranuclear palsy (PSP), and corticobasal syndrome (CBS). A very recent finding suggests that emotion perception changes with the progressing of neurodegeneration disorders, although differently mainly due to the unique anatomical correlation of different diseases [11].

C. IoT for Emotion Analysis

Prior to the advancement of research in the Internet of Things (IoT) and wearable technologies, there were several difficulties in using the physiological data for emotion recognition. As physiological patterns vary, it was hard to map them largely across the participants and subjects. Earlier recording of the biosignals required the participants to be stationary as the biosensors were much prone to artifacts. However, with the advancement of technology and wearable sensors came the multifunctionality of biosensors. Moreover, the use of wearables and an intelligent IoT framework for the acquisition and analysis of these physiological signals will have a way forward for a real-time analysis [12]. This will allow a simultaneous collection and analysis of an individual's emotion as
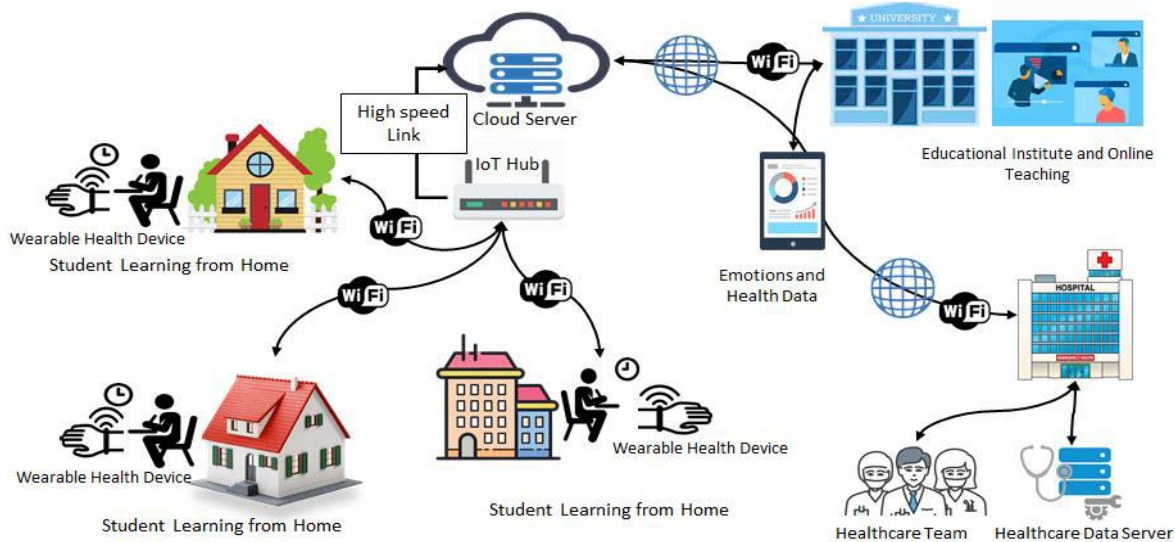
Fig. 2.   Cloud-based data processing of physiological sensors for emotion recognition through IoT-based framework.

this has made data collection much simpler and wireless, with minimum artifact that can be potentially translated to IoT [13].

In recent year, AI-enabled IoT-based emotion recognition has attained much attention. An IoT knowledge-based wellbeing recommendation system named "IAMHAPPY" is developed that implements a rule-based engine on wearable sensors' data suggesting people on how to achieve everyday happiness [14]. Mano *et al.* [15] developed a camera-based IoT system to recognize patient's emotion through camera images.

The relation between the emotions and the physiological signals have been widely studied for many applications. However, an important application of the emotions analysis is overlooked where schools and educational institutions can provide significant support to students when delivering distance learning. Moreover, this context has become much more relevant amid the Covid-19 outbreak, where a major paradigm shift is observed from face-to-face learning to distance learning. An online IoT-enabled real-time emotions analysis system will also make the required pastoral support very efficient. This will allow the tutors provide personalized attention to every individual in the class. Students with special requirements will benefit from it, by allowing tutors to keep an eye on their essential vitals remotely, throughout the teaching session.

Despite the notable work done in the domain of emotion recognition, to the best of our knowledge, none of the existing systems integrates multimodal methodologies at large and study the effect of combining a variety of sensors to system performance. This is vital since human beings are intelligent creatures and can camouflage their emotional expressions. This poses a big limitation to the developed facial recognition methods as well as methods using individual physiological sensors for human emotions. This work develops a long short-term memory (LSTM)-based emotion recognition through multimodal sensing modalities and integrates its outcome with the IoT framework to provide feedback for customized student education, their health, and wellbeing. The proposed IoT framework provides connectivity between sensing-based

output and the healthcare infrastructure for bespoke feedback. The contributions of the proposed work are as follows.

1) The development of LSTM-based emotion recognition system through multimodal physiological sensors with high performance.
2) An IoT-based framework to enable ultrareliable low-latency communications (URLLC) for student learning and healthcare amid pandemics alike Covid-19.
3) An integrated IoT-enabled machine learning-based emotion recognition framework to support distance learning and healthcare applications.

The remainder of this article is organized as follows. System model is presented in Section II. Section III covers results and discussion, whereas the conclusion and future directives are presented in Section IV.

II. SYSTEM MODEL

The emergence of Covid-19 has resulted in drastic changes throughout the world. One such change introduced in the traditional educational practices is the shift of conventional lectures conducted in physical environments to virtual lecturing environments. While the proposed machine learning techniques use physiological sensors to depict behavioral attributes of the students, it is still needed to collect and covey the sensory feedback to IoT hub in timely fashion. The overall schematic of the proposed cloud server-based data processing of physiological sensors for emotion recognition through IoT-based framework is presented in Fig. 2. The data recorded through physiological sensors will be transferred to the cloud server through the proposed IoT framework, then the cloud server processes the data analytics pipeline to identify the underlying patterns from the data and to profile the students' emotional state. The emotion's profiling will be available to educational institute as well as to healthcare infrastructure to make informed decision about educational as well as healthcare interventions, respectively, for affective learning and wellbeing. The following sections provide more details about the

proposed AI-enabled data processing system and the IoT-based data communication framework.

A. LSTM-Based Emotion Detection

The interpretation of human emotions is relatively a challenging task. It involves the analysis of several sensing modalities to interpret inner and inaccessible human system of emotions. These sensing modalities can be used to develop objective measure of classifying human emotion. Therefore, one of the objectives of this study was to develop a deep learning-based emotion detection platform that exploits multimodal physiological sensing data to interpret students' mental health via emotional and mental recognition during distance learning for potential applications.

1) Data Set: This study utilizes a novel physiological sensors-based data set [16], which is first of its kind to detect the human emotions. The multimodal physiological sensors used in this system are RSP, GSR, ECG, EMG, SKT, and BVP sensors. The placement and quantity of these sensors are as follows: RSP (1): high on torso, above chest; GSR(1): nondominant hands index finger; BVP(1): nondominant hand's middle finger, and EMG(3): two on the face and one on the back.

Thirty participants participated in data acquisition experiment [16]. Participants were shown a total of eight videos related of four different categories of emotions to trigger diverse emotional experiences. The different categories of the video were relaxing, boring, amusing, and scary. These categories were used as the four classes of internal emotions to be classified. The individual time duration of the videos was between 140 and 200 s and all the participant watched the same videos during the experiment at their turn. Continuous recording of the physiological signals at 1000 Hz/s was performed to record the objective measure. Further details of the experimental conditions and physiological sensors' setup can be obtained in [16]. In this work, the recorded objective measures are used to train the deep learning model and to classify the emotions.

2) Long Short-Term Memory for Emotion Recognition: The LSTM network is a special class of deep neural networks with the ability to memorize long term dependences of a time-series data set. Such capabilities are complimented by incorporating a variety of memory cells and gating operations inside an LSTM network. The memory cells are updated after performing a variety of gating operations, which in turn updates which values to remember and which to forget in the temporal sequence. Therefore, it is highly suited to model temporal dynamics in a robust and effective way.

There are three type of gating operation in LSTM: 1) input gate; 2) output gate; and 3) forget gate. The expressions that lay the foundation of LSTM are presented in

$$i_t = \sigma_t(W_i[h_{t-1}, x_t] + b_i) \tag{1}$$

$$f_t = \sigma_f(W_f[h_{t-1}, x_t] + b_f) \tag{2}$$

$$c_t = f_t c_{t-1} + i_t \sigma_c(W_c[h_{t-1}, x_t]) + b_c) \tag{3}$$

$$o_t = \sigma_o(W_o[h_{t-1}, x_t] + b_o) \tag{4}$$

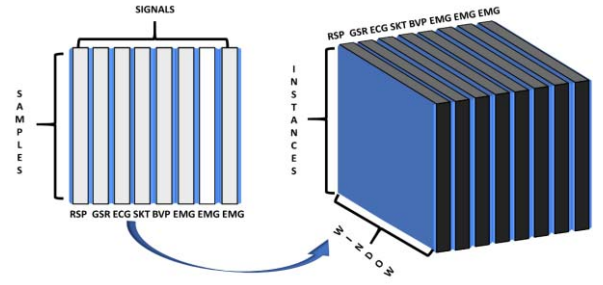$$h_t = o_t \sigma_h(c_t) \tag{5}$$

Fig. 3. Data preprocessing reshaping for to run LSTM model.

where x_i is the input data sequence; i, f, and o are input, output, and forget gate, respectively; c and h correspond to cell and hidden state; b_i, b_f, b_c, and b_o are biases related to input, forget gate, cell state, and output; W_i, W_f, W_c, W_o being the weight matrix of input gate, forget gate, cell state, and output gate respectively, and σ_t, σ_f, σ_c, σ_o, and σ_h are the activation functions of input, forget, cell state, output gate, and hidden state.

3) Data Preparation and LSTM Model Implementation: This work provides a feasibility study of real-time emotion detection on IoT hub/edge server. The physiological sensor data are resampled at 200 Hz to reduce the computational power, data transmission, and storage requirement over the cloud.

The overlapping windowing is very important in signal processing, especially when handling temporal data sequences, as it significantly affects the system performance [17]. Therefore, in this study, 4-s window with 50% overlap [18] is used to convert the data samples into data instances and windows. All the sensing modalities are first combined in a way that each column represents a single sensor, and rows represent the total number of samples per sensor (Fig. 3). Then, the data is transformed to make it compatible with the input shape of LSTM (instances*windows*sensors, see Fig. 3) used in TensorFlow. Each data instance contains 1000 samples of data, which corresponds to 5 s of data considering resampled frequency of 200 Hz.

The LSTM model parameters and information regarding data processing are presented in Table I. The data set is divided into 80/20 instances, where 80% of the data instances were used to train the LSTM model and remaining 20% were used for performance validation and testing. The LSTM model will be implemented on cloud server in Fig. 2.

F-score is used as a performance measure to compute and compare the performance of the proposed LSTM model. *F*-score is preferred over accuracy as it combines precision and recall, thus providing better insight in system performance across unbalanced data sets. The expression to find *F*-score is presented in (6), where TP, FP, and FN are true positives, false positives, and false negatives derived from the confusion matrix

$$F\text{-score} = \frac{2 * \text{TP}_C}{2 * \text{TP}_C + \text{FP}_C + \text{FN}_C} * 100. \tag{6}$$

4) Impact of Sensor Combinations on System Performance: The feasibility and computational power of any wearable
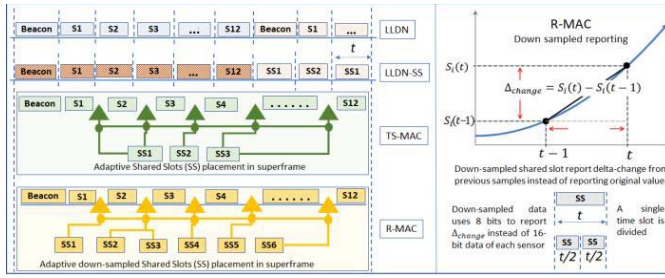
Fig. 4.   Superframe structure of legacy and proposed schemes.

TABLE I
LSTM MODEL PARAMETERS

Parameter	Value/dimension
window	4 sec (800 samples)
Overlapping windows	2 sec (400 samples)
Total data instances	[19050 × 800 × 8] (Train: 80%) (Test: 20%)
Batch size	[512 × 1000 × 8]
W_i, W_f, W_c, W_o	[128 × 128]
b_i, b_f, b_c, b_o, c, h	128
Lambda loss	0.0015
Learning rate	0.001
Optimization	Adam optimizer
Output classes	4

TABLE II
ANALYSING IMPACT OF SENSOR COMBINATION
ON SYSTEM PERFORMANCE

Combinations	Sensing modalities
C1	EMG 1, EMG 2, EMG 3
C2	ECG, BVP, GSR, RSP, SKT
C3	ECG, BVP, GSR, SKT
C4	ECG, BVP, GSR, RSP, SKT, EMG 1, EMG 2, EMG 3

TABLE III
DESCRIPTION OF FREQUENTLY USED VARIABLES IN THIS ARTICLE

Parameters	Variables	Value
Superframe duration	T	10ms
Total scheduled slots	s_{sch}	12
Total shared slots	s_{sh}	3
Timeslot duration	t	~410μs
Total sensors	S_T	12
Retransmission slot duration	t_r	205 μs, 410 μs
Probability of a failed communication	p	0 - 0.25

system are essential considerations in practical real-time implementations [19]. Therefore, the performance of different combinations of sensing modalities is also analyzed by generating the respective LSTM models. This is crucial as some of the sensing modalities in data set, such as respiratory sensors or EMG, are placed on body that might not be feasible in real-life implementation in the context of wearability, portability, and battery life. The different combinations analyzed in this study are presented in Table II, whereas further discussion on performance of all combinations is presented in Section IV. The sensing combinations are formed by considering the wearability and intrinsic nature of the sensing modalities as these can significantly influence the wearability and performance. For instance, C1 only contains EMG sensors to classify emotions while C3 excludes EMG and RSP.s

B. IoT Framework

The proposed machine learning techniques benefit from the sensory data, which can easily be sampled by a purpose-built smart watch. The IoT framework allows hassle-free data communication from the wearable sensors embedded in watch to IoT hub, and then to the cloud server as presented in Fig. 2.

The data collected from each sensor are 16 b with sampling frequency of 1 kHz. Thus, each sensor requires a 16-kb/s data rate. As the proposed work currently requires up to eight sensors, each of size 16 b and sampling frequency of 1 kHz, therefore, a minimum of 128 kb/s is required for effective communications. Additional, 64 kb/s is reserved for future changes, updates, and new sensors. However, the delay is relatively crucial in this case as the sampling time is 1 ms, thus, communication of each data sample should take less than 1 ms to deliver to the IoT hub. To enable high reliability and collision avoidance in the communications, time-division multiple

access (TDMA) is used. The communication from the sensors to IoT hub takes place in a superframe of duration 10 ms. Each superframe is further divided into 15 timeslots, where duration of each timeslot is 410 μs. The communication overhead time is taken to be 3.84 ms in a superframe. Since TDMA is used, therefore, the IoT hub emits a beacon at the start of every superframe to eliminate inaccurate local clock timing and if needed, manage communications from multiple sources without time overlap in communications. The frequently used parameters are presented in Table III.

In the proposed work, two MAC protocols, reliability-enabled MAC (R-MAC) and time-sensitive MAC (TS-MAC), are proposed. The proposed work offers suitable changes in existing IEEE802.15.4e low latency deterministic networks (LLDN) to meet the desired outcomes. Later in Section III, the performance of proposed work is compared with the LLDN to offer comparative analysis [20], [21].

C. Delay Analysis

In the traditional LLDN, the communication of all the sensor readings is scheduled periodically. In addition to the traditional LLDN, the use of shared slots (SSs) is also introduced to improve the reliability of the LLDN by enabling retransmission of failed slots. While the use of SSs improves reliability, yet the overall communications delay in failed communications also increases. Therefore, a suitable adaptive time sensitive MAC protocol is proposed, which takes into consideration the retransmission delay of failed communications. The superframe structure of LLDN, LLDN with SSs (LLDN-SS), and proposed TS-MAC and R-MAC is presented in Fig. 4.

The delay is a highly critical attribute in the current application scenario; therefore, it is necessary to investigate the average delay in communications of each of the protocol. The

delay in LLDN can be modeled as

$$d_{\text{LLDN}} = (1 - p) \times t + p \times T \tag{7}$$

where d_{LLDN} is the average delay of LLDN communications, p is the probability of failed communications, t is the time slot duration, and T is the superframe duration. In the case of LLDN-SS, the average delay can be expressed as

$$
\begin{aligned}
d_{\text{L-SS}} = &\{(1 - p) \times t\} + \{p \times (S_T \times t)\} \\
&+ \Bigg[u((p \times S_T) - s_{\text{sh}}) \\
&\quad \times \left\{ \frac{((p \times S_T) - s_{\text{sh}})}{S_T} \times (T - (S_T \times t)) \right\} \Bigg]
\end{aligned} \tag{8}
$$

where $d_{\text{L-ss}}$ is the average delay of LLDN-SS, S_T is the number of sensors, and s_{sh} is the number of SSs. $u((p \times S_T) - s_{\text{sh}})$ is the unit step function.

The delay in TS-MAC/R-MAC can be expressed as

$$
\begin{aligned}
d_R = d_{\text{TS}} = &\{(1 - p) \times t\} + \{p \times (t_r + t)\} \\
&+ \Bigg[u((p \times S_T) - s_{\text{sh}}) \\
&\quad \times \left\{ \frac{((p \times S_T) - s_{\text{sh}})}{S_T} \times (T - (t_r + t)) \right\} \Bigg]
\end{aligned} \tag{9}
$$

where d_R and d_{TS} are the average delays of R-MAC and TS-MAC respectively, and t_r is the retransmission slot/SS duration. In TS-MAC, $t_r = t$ whereas in R-MAC, $t_r = t/2$.

D. Error Rate

The overall communication reliability strongly depends on the channel conditions. The added SSs allow certain improvements in the error rate by introducing necessary retransmission ability for failed communications. In any case, the frame error rate can depict the reliability of the communication schemes. Here, the mathematical model for probability of failure in superframe is presented for LLDN, LLDN-SS, TS-MAC, and R-MAC.

The probability of failure in super frame communication in LLDN ($P(f_{\text{LLDN}})$) can be expressed as

$$P(f_{\text{LLDN}}) = \sum_{x=1}^{S_T} \binom{S_T}{x} p^x (1 - p)^{S_T - x}. \tag{10}$$

Since no SSs are used in traditional LLDN, so the overall superframe failure rate will only depend on the probability of failure in individual communications. In LLDN-SS, the probability of failure in super frame communication ($P(f_{\text{LLDN-SS}})$) can be expressed as

$$P(f_{\text{LLDN-SS}}) = \sum_{x=s_{sh}+1}^{S_T} \binom{S_T}{x} p^x (1 - p)^{S_T - x} \tag{11}$$

where s_{sh} is the total number of SSs. Since the same number of SSs is allocated in TS-MAC as in LLDN-SS, therefore, the overall reliability of the two do not differ. Thus, the probability of failure in super frame communication in TS-MAC is $(P(f_{\text{TS-MAC}})) = (P(f_{\text{LLDN-SS}}))$. Whereas in R-MAC, downsampling is introduced, which allows retransmission of the failed communications of 16-b sensors into 8-b delta indicator
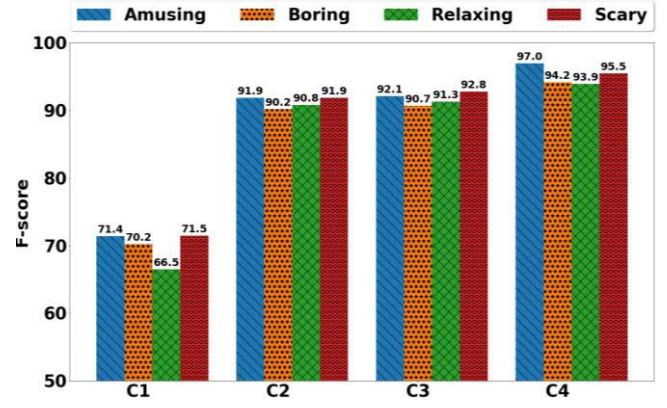

Fig. 5.   Performance analysis of different sensor combinations to classify emotions.

evaluated using last reading. In R-MAC, the retransmission communicates the change (Δ_{change}) from last successful communication instead of the sensor readings. The graphical representation of R-MAC downsampled reporting is presented in Fig. 4, whereas the reported change can be expressed as

$$
\begin{aligned}
\Delta_{\text{change}} &= S_i(t) - S_i(t - 1) \mid \Delta_{\max} \\
&= \{(|(\max(S_i) - S_i(t)|), (|\min(S_i) - S_i(t)|)\}
\end{aligned} \tag{12}
$$

where $S_i(t)$ is the current sensor reading and $S_i(t - 1)$ is the previous sensor reading. The sampling resolution is evaluated using $\Delta_{\max}$ and downsampled bits. Note that the signed numbers are used. In a more ideal scenario, instead of downsampling, the same resolution is maintained if $\sigma < \Delta_{\text{th}}$, where σ is the standard deviation in past consecutive readings and Δ_{th} is the change threshold. In any case, the retransmission size is reduced, thus leading to more possibilities of retransmissions. Therefore, probability of failure in super frame communication in R-MAC ($P(f_{\text{R-MAC}})$) is given by

$$P(f_{\text{R-MAC}}) = \sum_{x=(2 \times s_{sh})+1}^{S_T} \binom{S_T}{x} p^x (1 - p)^{S_T - x}. \tag{13}$$

The performance of the proposed schemes in comparison to legacy systems is presented in the following section.

III. RESULTS AND DISCUSSION

A. Emotion Recognition Through Physiological Sensors

The proposed LSTM model that classifies various emotions is validated on testing data set using f-score. The graphical presentation of performances obtained through the several models developed for each of sensing combinations is presented in Fig. 5, whereas the respective confusion matrices are presented in Table IV. In Fig. 5, the performance obtained in recognizing the four emotions, i.e., amusing, boring, relaxing, and scary, is presented. Each bar (column) represents the f-score obtained in classifying a particular emotion, where the f-score value is also presented on top of each bar for better readability. The overall f-score, which is the mean of f-scores of four emotions, is presented in Table IV. Among all sensor combinations, the best overall performance

TABLE IV
CONFUSION MATRIX OF DIFFERENT CLASSIFIER COMBINATIONS. (a) C1. (b) C2. (c) C3. (d) C4

f-score 69.9%	(a)				
Actual Class	classified as ➔	Amuse	Bore	Relax	Scary
	Amuse	746	103	124	104
	Bore	76	617	93	49
	Relax	93	99	605	79
	Scary	99	104	122	697
f-score 91.2%	(b)				
Actual Class	classified as ➔	Amuse	Bore	Relax	Scary
	Amuse	992	26	20	39
	Bore	48	733	37	17
	Relax	21	12	796	47
	Scary	20	20	25	957
f-score 91.7%	(c)				
Actual Class	classified as ➔	Amuse	Bore	Relax	Scary
	Amuse	999	35	11	32
	Bore	32	773	13	17
	Relax	34	39	769	34
	Scary	27	22	16	957
f-score 95.1%	(d)				
Actual Class	classified as ➔	Amuse	Bore	Relax	Scary
	Amuse	1050	16	3	8
	Bore	14	787	23	11
	Relax	10	22	816	28
	Scary	14	11	20	977

of 95.1% is achieved by sensor combination C4 that utilities all the recorded physiological sensors and is followed by combination C2 and C4 performing equally well by obtaining f-score of over 91%. Whereas the performance sensor combination C1 is much below the rest of combinations. The high f-score achieved in three sensor combinations C2–C4 confirms the potential of using the proposed system in recognizing human emotions is real-life settings with high performance of above 95%.

B. Which Sensor Combination to Choose?

The selection of individual sensors or a combination of sensors in crucial in real life where sensors placements, feasibility, portability, and computational requirements have significance. Therefore, this study analyzes the effect of reducing the number and type of sensors and their impact on the overall performance and performance by class of each recognized emotion.

The presented performance of each emotion class in Fig. 5 highlights the fact that the LSTM model, which considers all sensors, outperforms the others. As shown in the figure, this model achieves a minimum f-score of above 93% in recognizing all four cases. However, the LSTM models developed using all sensor combinations except EMG (C2, C3) also performed comparatively well by achieving performance of above 90% in classifying each of the four emotions (see Fig. 5). The worst performer for each class was combination of only EMG

sensors (C1) whose performance was around 70% for most of the classes.

While the best performance is obtained using the LSTM-based model that incorporates all eight sensing modalities (C4), yet, using a large number of sensors are often impractical. Since the proposed work targets healthcare and distance learning applications, therefore, the sensors should be easily wearable. Intended application of using the proposed system is to monitor students' mental health and effectiveness of the lecture contents in virtual environments currently enforced in amid—and probably in post Covid-19 situations. On the contrary, using the least number of sensors in the form of EMG (C1) is the worst performer is our case. Besides, the placement of all of three EMG sensors is quite impractical, considering the real-life conditions. This is because two of the EMG sensors were placed on *zygomaticus* and *supercilli* muscle group on face, and the third was mounted on upper back at *trapezius* muscle. Therefore, as per the findings of this study, the insignificant performance and the impractical sensors' placement, makes EMG sensors the least relevant in the context of emotion recognition. Another interesting finding is that there is no major difference in performance ($<0.05\%$) observed for C2 and C3. The combination C3 utilizes ECG, BVP, GSR, and SKT while the combination C2 includes RSP sensors in addition to the sensors used by C3. This suggests that RSP sensor do not contribute much to the system performance but only makes system less practical as the RSP is placed above chest and under the armpit.

Therefore, these findings suggest that the best tradeoff between the system performance and the wearability is C2 utilizing ECG, BVP, GSR, and SKT sensors. Such sensor combinations are readily available in the offshelf wrist wearable devices such as shimmer [22] wrist bands, which support the findings of our study and transformation of the proposed system into real life.

C. IoT Framework Results

The performance of the proposed protocols is evaluated based on two primary performance metrics: 1) delay and 2) reliability. Since the desired application has a very strict delay restriction, therefore, the average communication delay in proposed protocols, TS-MAC and R-MAC, along with state of the art is evaluated. The average delay in individual communications from sensors to IoT hub is presented as a function of probability of failure in communication (i.e., channel conditions), as represented in Fig. 6. The average delay is an important parameter and based on results, the suitability of outcomes can be argued. The figure shows that the delay in both proposed schemes remains below 1-ms threshold. As represented in Fig. 6, the delay of TS-MAC and R-MAC is well within the limits even for relatively high communication failure probability. In comparison to this, the delay in both LLDN and LLDN-SS crosses the 1-ms threshold once the communication failure probability reaches above 0.06 and 0.12, respectively. While both state-of-the-art protocols also offer relatively low delay, however, in the given scenario, delay restrictions were posed at 1 ms, which is even below most of the robotics and process control applications [23].
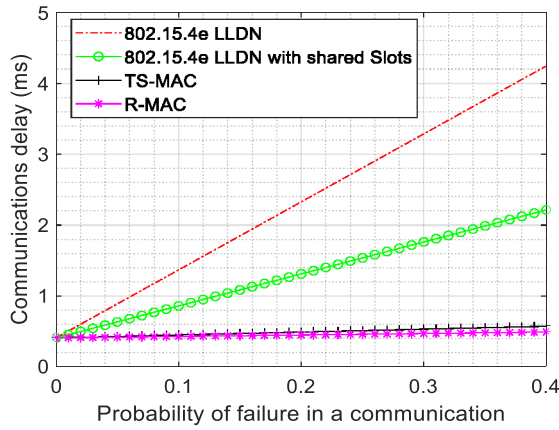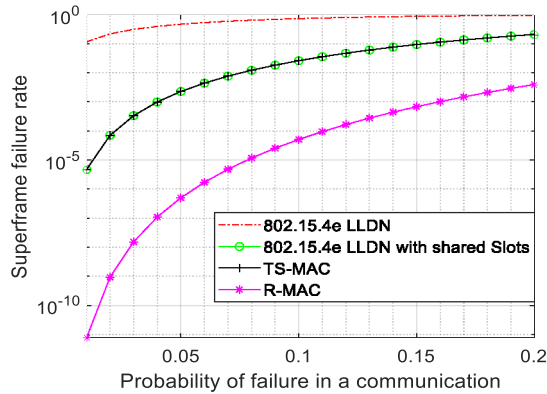
Fig. 6. Delay analysis.

Fig. 7. Frame error rate.

In addition to the delay reduction, the proposed schemes offer relatively high communication reliability. The results presented in Fig. 7 show that R-MAC offers a notable improvement in reliability as frame error rate is much lower compared to both TS-MAC and LLDN-SS. The reliability of the communication system is very important and higher packet reception rate can justify the suitability of work for ultrareliable lowlatency communications such as healthcare systems. It is worth mentioning that a frame is considered in error if there is at least one failed communication in a superframe. Given the strict criteria for failure rate evaluation, the results of R-MAC are very positive and are suitable for the proposed healthcare and distance learning enhancement framework. On the downside, the R-MAC sacrifices the data precision to achieve high reliability and low latency, which is acceptable in the presented scenario but can cause certain limitations in other applications.

IV. CONCLUSION

In this article, a comprehensive IoT-enabled human emotion analysis framework is proposed. Physiological signals are analyzed through LSTM-based deep learning model to recognize various emotions, i.e., amusing, boring, relaxation, and scary. The developed physiological-based LSTM model recognizes human emotions with very high performance of above 95%. Moreover, the proposed intelligent IoT framework provide seamless data communication with low latency and high reliability between the sensing devices and the

IoT hub. To meet the stringent sampling and communication needs of physiological sensors, low latency of less than 1 ms is achieved. The proposed IoT-enabled emotion recognition paradigm is expected to support and assist students, educational institutions, and healthcare infrastructure toward effective distance-learning paradigm while keeping the optimal health and wellbeing standards in the Covid-19 outbreak and future pandemic threats. As future research directions, the article can be further extended with a focus on end-to-end communications and visual aids to support distance learning. In addition, the incorporation of edge services can enhance the feasibility of the proposed work.

REFERENCES

[1] C. E. Izard, *Human Emotions*. New York, NY, USA: Springer, 2013.
[2] M. Awais, N. Badruddin, and M. Drieberg, "Driver drowsiness detection using EEG power spectrum analysis," in *Proc. IEEE Reg. 10 Symp.*, 2014, pp. 244–247.
[3] D. Ayata, Y. Yaslan, and M. E. Kamasak, "Emotion based music recommendation system using wearable physiological sensors," *IEEE Trans. Consum. Electron.*, vol. 64, no. 2, pp. 196–203, May 2018.
[4] D. S. Wickramasuriya, M. K. Tessmer, and R. T. Faghih, "Facial expression-based emotion classification using electrocardiogram and respiration signals," in *Proc. IEEE Healthcare Innovat. Point Care Technol.*, 2019, pp. 9–12.
[5] E. M. Albornoz, D. H. Milone, and H. L. Rufiner, "Spoken emotion recognition using hierarchical classifiers," *Comput. Speech Lang.*, vol. 25, no. 3, pp. 556–570, 2011.
[6] Z. Yin, M. Zhao, Y. Wang, J. Yang, and J. Zhang, "Recognition of emotions using multimodal physiological signals and an ensemble deep learning model," *Comput. Methods Programs Biomed.*, vol. 140, pp. 93–110, Mar. 2017.
[7] S. Tripathi, A. Kumar, A. Ramesh, C. Singh, and P. Yenigalla, "Deep learning based emotion recognition system using speech features and transcriptions," 2019. [Online]. Available: arXiv:1906.05681.
[8] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order," *Pattern Recognit.*, vol. 61, pp. 610–628, Jan. 2017.
[9] K. S. Rao, V. K. Saroj, S. Maity, and S. G. Koolagudi, "Recognition of emotions from video using neural network models," *Expert Syst. App.*, vol. 38, no. 10, pp. 13181–13185, 2011.
[10] H. Kaya, F. Gürpinar, and A. A. Salah, "Video-based emotion recognition in the wild using deep transfer learning and score fusion," *Image Vis. Comput.*, vol. 65, pp. 66–75, Sep. 2017.
[11] C. L. Brown et al., "Comparing two facets of emotion perception across multiple neurodegenerative diseases," *Social Cogn. Affect. Neurosci.*, vol. 15, no. 5, pp. 511–522, 2020.
[12] T. Qiu et al., "Intelligent systems for the Internet of Things," *IEEE Access*, vol. 7, pp. 146342–146347, 2019.
[13] J. J. P. C. Rodrigues, S. Jabbar, M. Abdallah, C. Verikoukis, and M. Guizani, "Future communication trends toward Internet of Things services and applications," *IEEE Wireless Commun.*, vol. 26, no. 6, pp. 6–8, Dec. 2019.
[14] A. Gyrard and A. Sheth, "IAMHAPPY: Towards an IoT knowledge-based cross-domain well-being recommendation system for everyday happiness," *Smart Health*, vol. 15, Mar. 2020, Art. no. 100083.
[15] L. Y. Mano et al., "Exploiting IoT technologies for enhancing health smart homes through patient identification and emotion recognition," *Comput. Commun.*, vols. 89–90, pp. 178–190, Sep. 2016.
[16] K. Sharma, C. Castellini, E. L. van den Broek, A. Albu-Schaeffer, and F. Schwenker, "A dataset of continuous affect annotations and physiological signals for emotion analysis," *Nat. Sci. Data*, vol. 6, no. 1, pp. 1–13, 2019.
[17] M. Awais, N. Badruddin, and M. Drieberg, "A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability," *Sensors*, vol. 17, no. 9, p. 1991, 2017.
[18] M. Awais, L. Palmerini, A. K. Bourke, E. A. F. Ihlen, J. L. Helbostad, and L. Chiari, "Performance evaluation of state of the art systems for physical activity classification of older subjects using inertial sensors in a real life scenario: A benchmark study," *Sensors*, vol. 16, no. 12, p. 2105, 2016.

[19] M. Awais, L. Chiari, E. A. F. Ihlen, J. L. Helbostad, and L. Palmerini, "Physical activity classification for elderly people in free-living conditions," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 1, pp. 197–207, Jan. 2019.

[20] M. Raza, H. Le-minh, N. Aslam, and S. Hussain, "A novel MAC proposal for critical and emergency communications in industrial wireless sensor networks," *Comput. Electr. Eng.*, vol. 72, pp. 976–989, Nov. 2018.

[21] *IEEE Standard for Local and Metropolitan Area Networks—Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs) Amendment 1: MAC Sublayer*, IEEE Standard 802.15.4e-2012 (Amendment to IEEE Std 802.15.4-2011), Apr. 2012, doi: 10.1109/IEEESTD.2012.6185525.

[22] A. Burns *et al.*, "SHIMMER^{TM}—A wireless sensor platform for non-invasive biomedical research," *IEEE Sensors J.*, vol. 10, no. 9, pp. 1527–1534, Sep. 2010.

[23] M. Raza, N. Aslam, H. Le-Minh, S. Hussain, Y. Cao, and N. M. Khan, "A critical analyssis of research potential, challenges, and future directives in industrial wireless sensor networks," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 39–95, 1st Quart., 2018.

Kiran Bashir is currently pursuing the M.C.S. degree with the Department of Computer Science, Virtual University, Lahore, Pakistan.

Her research interests include image processing, machine learning, and wireless networks.

Umar Manzoor received the Ph.D. degree in computer science from the University of Salford, Salford, U.K.

He is a Lecturer with the University of Hull, Hull, U.K. Previously, he worked as a Research Fellow with the Heterogenous Machine Learning Group, Tulane University, New Orleans, LA, USA. He has strong interest in machine learning, natural language processing, multiagent systems, and artificial intelligence.

Muhammad Awais (Member, IEEE) received the Ph.D. degree in biomedical, electrical and system engineering from the University of Bologna, Bologna, Italy, in 2018.

He is a Senior Lecturer with Edge Hill University, Ormskirk, U.K. He worked as a Research Fellow in data analytics and AI with the University of Hull, Hull, U.K., and in signal processing and machine learning with the University of Leeds, Leeds, U.K. His research interests are applied machine learning to develop ICT based systems for Internet of things, Industry 4.0, biomedical and health care domain.

Saif ul Islam received the Ph.D. degree in computer science from the University Toulouse III Paul Sabatier, Toulouse, France, in 2015.

He is an Assistant Professor with the Department of Computer Science, KICSIT, Institute of Space Technology, Islamabad, Pakistan. His research interests include resource and energy management in large-scale distributed systems and Internet of Things.

Mohsin Raza (Member, IEEE) received the Ph.D. degree from the Math, Physics and Electrical Engineering Department, Northumbria University (NU), Newcastle upon Tyne, U.K., in 2018.

He is a Senior Lecturer with Edge Hill University, Ormskirk, U.K. Prior to this, he worked as a Lecturer with Northumbria University, Newcastle upon Tyne, from 2018 to 2019, a Postdoctoral Fellow with Middlesex University, London, U.K., from 2018 to 2019, a Junior Lecturer from 2010 to 2012, later as a Lecturer with Mohammad Ali Jinnah University, Karachi, Pakistan, from 2012 to 2015, and a Hardware Engineer with USS, Karachi, from 2009 to 2010. His research interests include IoT, 5G, ITS, machine learning, industry 4.0, and digital twins.

Joel J. P. C. Rodrigues (Fellow, IEEE) received the Ph.D. degree in informatics engineering and the M.Sc. degree from the University of Beira Interior, Covilhã, Portugal.

He is a Professor with the Federal University of Piauí, Teresina, Brazil; a Senior Researcher with the Instituto de Telecomunicações, Lisboa, Portugal; and a Collaborator of the Post-Graduation Program on Teleinformatics Engineering with the Federal University of Ceará (UFC), Fortaleza, Brazil. He is the Leader of the Next Generation Networks and Applications (NetGNA) Research Group (CNPq), an IEEE Distinguished Lecturer, a Member Representative of the IEEE Communications Society on the IEEE Biometrics Council. He has authored or coauthored over 950 papers in refereed international journals and conferences, three books, two patents, and one ITU-T Recommendation.

Nishant Singh received the Ph.D. degree in biomedical technology from the Birla Institute of Technology, Ranchi, India, in 2015.

He is currently a Postdoctoral Research Fellow with the School of Psychology, University of Birmingham, Birmingham, U.K. He worked as a Postdoctoral Fellow from with the Computer Science Department, Middlesex University, London, U.K., 2018 to 2019.