A Facial and Vocal Expression Based Comprehensive Framework for Real-Time Student Stress Monitoring in an IoT-Fog-Cloud Environment

MADANJIT SINGH[1], SARVESHWAR BHARTI[1], HARJIT KAUR[1], VAIBHAV ARORA[2], MUNISH SAINI[2], MANEVPREET KAUR[1], AND JASWINDER SINGH[1]

[1]Department of Computer Science, Guru Nanak Dev University, Amritsar 143005, India
[2]Department of Computer Engineering and Technology, Guru Nanak Dev University, Amritsar 143005, India

Corresponding author: Madanjit Singh (madanjit.csc@gndu.ac.in)

ABSTRACT In this era of digital and modern education, the existence of psychological stress on students cannot be denied. The surplus aggregation of the stress may lead to different problems like a decline in student grade (performance), an increase of violence in behavior, and even more extreme cases. The advent of Information Communication and Technology (ICT) and its tools opened the doors to innovations that facilitate interactions among things and humans. In this utilization, the paper proposes a novel, IoT-aware student-centric stress monitoring and real-time alert generating framework to predict student stress index in a particular context. In elaboration, we respectively used extended VGG16, Bidirectional Long Short Term Memory network (Bi -LSTM), and Multinomial Naïve Bayes techniques to generate the scores of emotions from student facial expressions, speech pitch, and content of student speech at the cloud layer. Specifically, the model aims to classify the stress events as normal or abnormal on basis of the overall emotion of the students' physiological data readings. The activation of the abnormal event in case of higher values for negative emotions like stress, fear, sadness, disgust, etc.; a stern alert is sent to the student, coordinators, and caretakers. This proposed framework will ultimately be a great tool that will support the education institutions, students, their parents, and guardians to get a real-time alert on students' overall emotions. The prior knowledge of stress accumulated on the mind of the student will help in overcoming major problems of student dropout, decrease student academic performance, and tackle the stress situation that may lead to the student attempting suicide.

INDEX TERMS IoT, fog, real-time alert, emotion analysis, facial emotions, speech emotions, dropout, student performance, stress.

I. INTRODUCTION

The education teaching and learning pedagogy has evolved throughout the timespan [1], [2]. The technological advancements in the field of Information Communication and Technology (ICT) and other related areas across the globe have a direct repercussion on education [3], [4]. This infusion of technology with education learning and teaching pedagogy, or education theory has brought forth the new education implication referred to as education technology [5], [6]. It is

observed that in the modern era of technology, both teachers and students relied excessively on technology for achieving the goals of successful learning [7]. Moreover, the introduction of contemporary teaching pedagogies like blended learning [8], [9] and flipped learning (classrooms) [10], [11] in education has revealed persuasive evidence of progressive developments in the area of education [12].

With the testimonies of education teaching and learning progression and development, the existence of adverse consequences and their generation due to various extrinsic and intrinsic factors cannot be denied [13]. One of the major issues that prevailed with the regular upgrading of
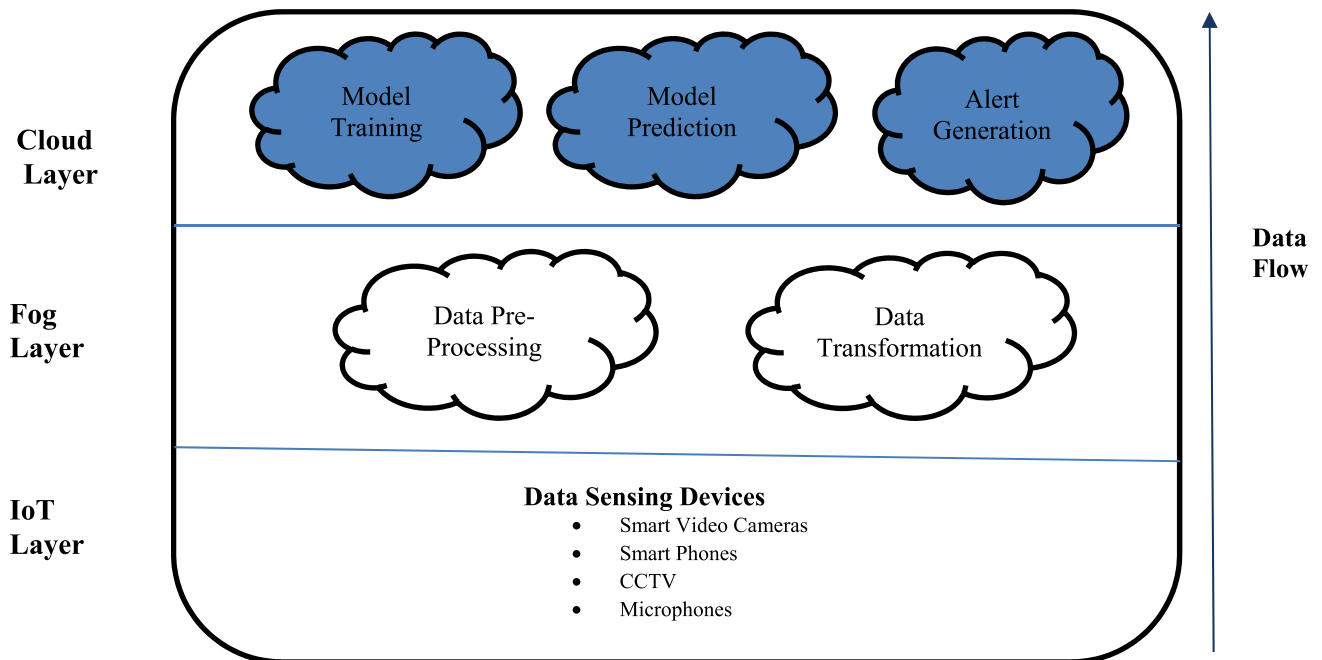
The associate editor coordinating the review of this manuscript and approving it for publication was Rongbo Zhu[ID].

education and its pedagogy is the over-accumulation of stress (and other negative emotions) on stakeholders of education practices [14]. Stress (or other negative emotions) has diverse deleterious repercussions on the physical and mental health of the learners [15], [16]. It escorts to depression, anxiety, eating disorders, insomnia, poor academic performance, dropout situations, or even suicidal conditions, especially among the students [17]. The stress can be measured using computer-based non-invasive methods by understanding physiological and physical changes in a person like changes in a heartbeat, voice attenuation, body posture, and facial expressions. In this paper, we focused to analyze the emotions of the students based on these parameters and evaluate against the purposed IoT-Fog-based model to identify the stress-endangered students.

The recent advancement in IoT-based applications is providing an optimized scenario for different areas such as health care, automation systems, transportation services, etc. [18]. IoT with the services of fog computing becomes a powerful measure for effective and quick response paradigms, especially in the case of emergency services. Fog computing offers several advantages such as low latency, location awareness, quality of service assurance, and real-time analysis and alert generation [19].

The framework of the IoT-Fog computing is a three-layered architecture as depicted in Fig. 1. The IoT layer enables the end-users to fetch data from real-world entities. In our case cameras, smartphones and microphones would act as a primary source of fetching data. The Fog layer act as a bridge between raw data and processed data which would be used for ML model training and prediction. The cloud layer provides data storage, processing, prediction, and alert generation services.

The framework aids in achieving the following set of research objectives.

1) Utilize IoT-Fog capabilities in building a comprehensive framework for monitoring the student emotions in a smart environment.
2) Analyze the level of stress (or other negatives) emotions based on physical measures, voice samples, and content evaluation.
3) To propose a measure indicating the generalized emotion for student behavior during analysis.
4) Generating the stern alert in case of excessive accumulation of stress (or other negative emotions) indicated by the proposed measure.

The paper is organized into the following sections: the former section specifies a basic introduction to the proposed work. Section 2 presents an overview of the previous work done by the different researchers to perform student healthcare monitoring and to propose an IoT-based alert generation and monitoring system. The detailed layout of the proposed framework for the analysis and monitoring process is discussed in Section 3. The outcomes of the proposed framework implementation and its performance analysis are depicted in Section 4. Further, the succeeding Section 5 provides a detailed discussion of the obtained results. The various threat to the validity of the conducted study is represented in Section 6, discussed in the subsequent section followed by some conclusive remarks on the paper along with the possible future recommendations.

II. RELATED WORK

This section reviews some of the recent and important contributions in the field of emotion monitoring and analysis. Mozafari *et al.* [20] in their findings deploy comprehensive generalizable single and multimodal IoT-based techniques for stress level detection based on three attributes; using wearable IoT devices for continuous data collection, multimodal and heterogeneous sensor data collection, and hierarchal data collection consisting of multiple sources. The method monitors stress using four physiological signals; Photoplethysmogram (PPG), Galvanic Skin Response (GSR), Abdominal Respiratory (AR), and Thoracic Respiratory (TR). Further, the LOSO (Leave-OneSubject-Out) method was incorporated to capture the impact variations in subjects. Uday *et al.* [21] developed an IoT system for detection and feedback generation for stressed patients. The proposed system collects data from the smart band and chest strap modules to monitor stress in real-time by sending and processing data on a cloud-based ThingSpeak server.

Shaikh & Ali [22] developed an IoT–aware stress management system for evaluating student stress indexing. The system features the temporal dynamic Bayesian Network (TDBN) to depict stress events by encompassing readings of conventional medical devices at the fog layer. Further, experimental works reveal the outperformance of the Bayesian Classifier over other classification methods. The study also offers a gentle touch in the literature review on current techniques and methods, including deep learning for complex multi-dimensional healthcare sensor data in support of fog computing.

Kocielnik *et al.* [23] proposed a framework for continuous measures of stress-related data from sensor devices in real-world activities. The focus was to address the challenges along with evaluation of approach on on-field studies, long-term measurement of stressed people, and the analysis of their behavioral patterns to perceive meaningful information to yield stress balance. Furthermore, Oti *et al.* [24] in their study focus to monitor maternal stress during their pregnancy period. The proposed k-means algorithm deployed on an IoT-based remote health monitoring system provides stress monitoring in the hospital's environment as well as in daily activities by continuously monitoring patients' health-related parameters.

Verma & Sood [25] present an IoT-based student-centric framework for stress monitoring by incorporating a two-stage Temporal Dynamic Bayesian Network (TDBN) model over the fog layer. The stress index is evaluated based on four parameters; (a) leaf node evidence, (b) workload factor, (c) context (d) student health trait, and the results of stress endangered students are provided by the alert generation mechanism. Pace *et al.* [26] initiate the BodyEdge model for human-centric applications in the healthcare industry. The proposed architecture was implemented to detect high-stress levels for workers and athletes from Heart Rate Variability (HRV) features. This model consists of two modules; one

is a mobile client module for data collection and the second is performing an edge gateway for data processing with the real-time facilities of private and public cloud platforms. Rachakonda *et al.* [27] used a deep neural network method to process stress data collected from the accelerometer, humidity, and temperature measuring sensors in an edge computing framework.

Sheng *et al.* [28] provided an overview of the IETF protocol bundle in support of IoT technology. These protocols were evaluated to provide better guidelines, resulting in the effective and efficient design of the communication system in healthcare scenarios. Further, Alberdi *et al.* [29] concentrated on recent works in automatic stress detection over the measurements executed along with the psychological, physiological, and behavior modalities. Contextual measurements along with these parameters are used to adopt the prominent suitable methods to facilitate the development of stress measuring systems.

Cui *et al.* [30] in their findings propose a non-invasive stress monitoring system with having sensor units; capable to measure temperature using Infrared Temperature Measurement (ITM), PhotoPlethysmoGram (PPG), and Inertial Measurement Units. The focus is to continuously monitor stress signs in sheep during transportation. The designed Wearable Stress Monitoring System (WSMS) showed adequate power in recording and sending detecting information on physiology and climate during transport. This study, based on non-contact and non-destructive monitoring techniques helped in minimizing the effects of stress loads on sheep.

Deploying the AI model at fog nodes may suffer from several adversarial threats. Li *et al.* [31] proposed a framework named as DeSVig (Decentralized swift vigilance) to recognize adversarial attacks and correct the mistakes in a few seconds in artificial intelligence-inspired industrial systems. The framework overall improves the effectiveness of the system by recognizing abnormal inputs.

III. THE PROPOSED FRAMEWORK

The proposed framework (as shown in Fig. 2) demonstrates the internal layout of the student stress monitoring and alert generation system. It consists of three layers: the IoT layer, Fog layer, and Cloud layer.

A. IoT-FOG BASED FRAMEWORK ARCHITECTURE

The raw data is collected from IoT-enabled input devices like smart video cameras, smartphones, and microphones installed around the student under observation. The data fetched at the IoT layer is then passed to the fog layer over the internet for further operations. There is a need of sensing the data and sending it to the cloud layer for processing purposes in real-time. This is only possible with the integration of internet-enabled smart devices. For this purpose, we installed multiple.

IoT devices at the data acquisition layer. The fog layer acts as an interface between the IoT layer and the cloud layer as it
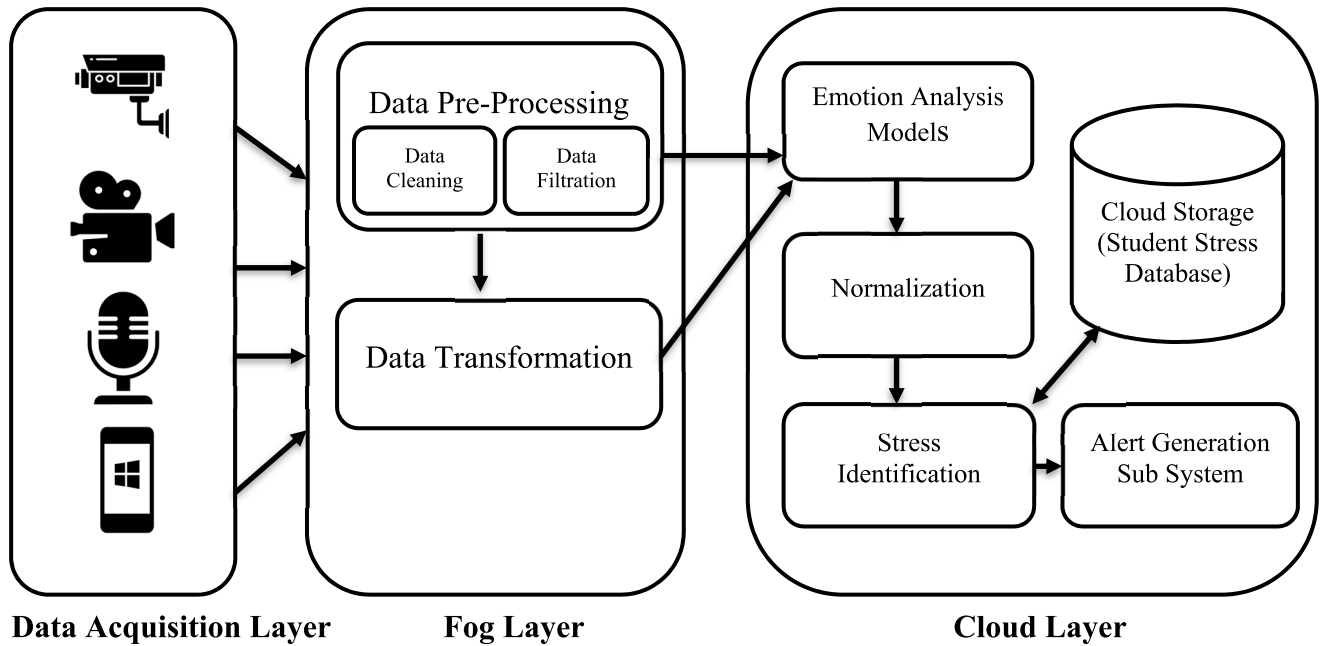
FIGURE 2. The layered architecture of the proposed system.

takes input from the IoT layer and provides output to the cloud layer. The cloud layer is responsible for all major activities (model training, testing, alert generation) in the proposed model. Further, the predictions are made by using trained weights, and alerts are generated for the cases where the stress readings are above the threshold, enabling the stress event.

B. DATA ACQUISITION LAYER

In the process of data acquisition, three different types of scenarios have been taken into consideration. The CCTV and live recording cameras are used for getting the video streams and photos of the persons (students) and their expressions, for analyzing their facial emotions. The dataset was generated and collected based on 6 categories of emotions namely: Surprise, Fear, Neutral, Anger, Sad and Happy. The microphone has been used to record and store audio streams to perform Speech emotion analysis. Further, on audio stream google speech to text API is used to gather textual data for content-based analysis. Both audio and textual data are categorized in three different categories namely: Positive, Neutral, and Negative. For efficient and effective analysis, we have extended the collected dataset by using the Kaggle dataset resources♦. This extension will help us in increasing the efficiency of the trained model. Further, the performance of the proposed model can be tested on the Kaggle test dataset℘. Table 1 gives an overview of different parameters that are considered in the data set along with IoT devices used for the collection of the data.

♦https://www.kaggle.com/search?q=emotion+analysis
℘https://www.kaggle.com/shivamburnwal/speech-emotion-recognition

TABLE 1. Dataset categorization.

Dataset	IoT Technology	Description	Parameters
Visual data	Smart Video Camera	Visual data is collected by recording the facial expressions	Facial expressions, blinking of eyes, eye movement.
Audio data	Microphones	Audio data is collected from the student's answers to the counselor	Speech clarity, texture, speed, pronunciation, fluency.
Textual data	Google Library	Audio data is converted to the textual form	Keywords, no of words, sentence formation, the semantics of the text.

C. FOG LAYER

Data gathered by the data acquisition layer is further fed to the fog layer for analysis. The purpose of embedding the fog layer in between the IoT layer and cloud layer is to reduce the workload of the cloud layer. The fog layer performed all data cleaning, data filtration, and data transformation tasks. It will restrict not loading unnecessary data over the cloud layer which might result in more overhead and an increase in the complexity in the absence of the fog layer. Moreover, the data is processed locally at the fog layer, hence able to save network bandwidth and increase the overall performance of the proposed model. We utilized three different pre-processing approaches for data filtration and cleaning purposes. For the facial emotion dataset, images were cropped and centralized to focus on the facial expression.

The target image size is set to 224 × 224 × 3 for preprocessing as while training we would be using the VGG16 model [31]. Further, the images were normalized and augmented with the use of ImageDataGenerator API by using a sheer range of 20% and horizontal flip as the arguments [32]. For the audio dataset, firstly we loaded the audio files using the Librosa library with Kaiser_fast res_type (refer to Fig. 3) [33]. A further mel-scaled spectrogram is generated (see Fig. 4), which is then converted to a dB-scaled spectrogram (refer to Fig. 5) [34]. With these, we are having features in the image format.
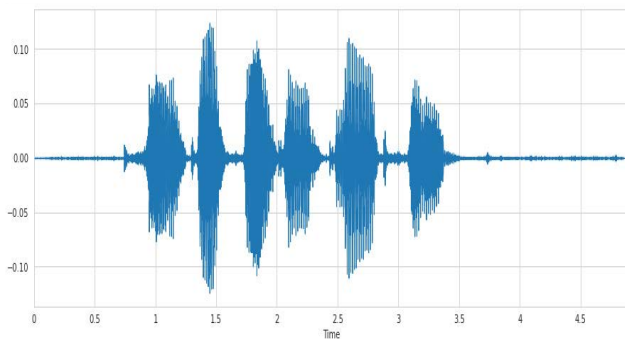

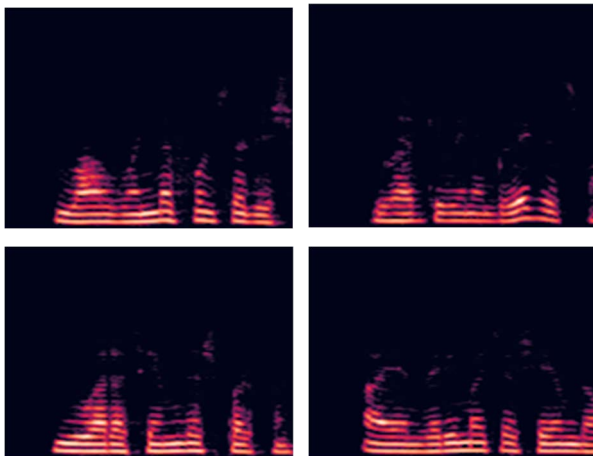**FIGURE 3.** Wave plot of audio sample.

FIGURE 4. Mel spectrogram as images.

For textual data, we applied natural language processing and data cleaning approaches [35]. Firstly, we deal with anomalies within the text by removing extra space, removing punctuation marks, removing stop words, and removing numbers [36]. Further, we tokenized the corpus using a word tokenizer from the NLTK library [37]. As tokens would be a key element while the generation vocabulary list would be further fed into the model training. To normalize the text we used Porter Stemmer and passed all tokens through it and gather a new list of tokens [38]. PorterStemmer uses the suffix stripping technique to produce root (stem) words [39]. In this analysis, we prefer to use PorterStemmer for its simplicity,
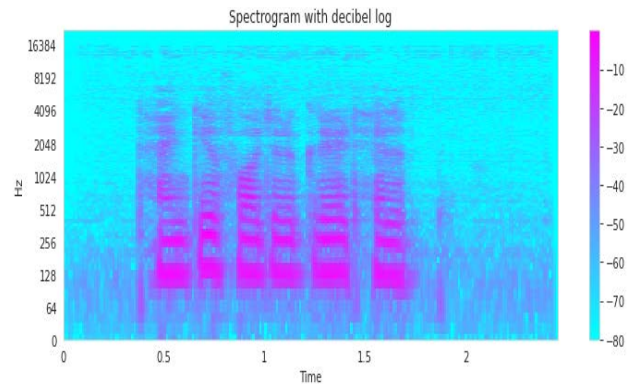

FIGURE 5. Spectrogram with Decibel log.

fast speed, and utility in an Information retrieval environment [40]. Further, we convert the corpus sentences into an array of numbers. Here numbered array is formed from the vocabulary that is get by using the fit_transformer operation in the CoutVectorizer function of the Sklearn library [41]. It converts data into a vectorized form. By getting corpus in this form now we have data in an organized form where each sentence will of the same length i.e., size of the vocabulary and we are all set to feed the data into the machine learning model.

Algorithm 1 Stress Prediction and Alert Generation Model at the Cloud Layer

---

Input: Preprocess data set from fog layer. (facial_data U speech_data U tect_data)

Step 1: for i=1 to N; **Repeat**
  (a) Extract the dataset from the cloud database.
  (b) Determine the data against the trained data model for facial expression detection.
  (c) Classify the stress emotion results to the respective class.
  (d) Apply the audio emotion analysis model against the collected data set.
  (e) Classify the speech analyzer results to the respective stress class.
  (f) The data is processed against the content analyzer for emotion detection.
  (g) Placed data under the respective positive, negative, and neutral classes.
  End for loop

Step 2: Normalise the results to a single scale of range +1 to −1.

Step 3: Filter out the records which are closer to the negative side.

Step 4: Generate a real-time alert to the students and their belongings for the stress results using the Ubidot platform.

Step 5: Store the results in the cloud database for future analysis.

Step 6: Exit

D. CLOUD LAYER

This layer consists of the model training and analysis of the data against the trained model. The predicted model is trained to extract meaningful information related to the stress activities from the data set in real-time. The stress monitoring

is performed on the selected three parameters. The whole workflow of the purposed prediction model is discussed in Algorithm 1. The data files that are generated from the IoT sensors dataset are stored at the Amazon EC2 cloud, where it is analyzed using various trained models. Further, the Ubidots IoT development platform is used for sending alerts to caregivers and family members at this layer. Ubidots is a free-to-use and secure IoT platform to provide IoT-based facilities and solutions, especially to students and researchers[Φ]. The services like sending data from any internet-enabled device to a cloud database, implementing operations or actions on the data, alert generating facilities based on data activity, and data visualization are provided under this platform[φ].

1) EMOTION ANALYSIS

Emotion analysis is expressed as a measure of bringing off the emotions out of the data [42]. The extraction process is based on the feelings like stress, happiness, and anger that are conveyed in the input data. In our proposed work, we have utilized the services of three different emotion extricate models; facial emotion detection, speech emotion detection, and content-based emotion analysis. These models are first trained over a large sample of data for understanding the emotional or psychological state of the student in real-time while they undergo counseling.

a: FACIAL EMOTION ANALYSIS

For analyzing the emotions from the face of the candidate/student we had trained our machine learning model over the Pretrained VGG16 model extended by Flatten layer, one Dense layer, and an output layer giving output over 6 categories [43] (refer to Fig. 6). Simonyan and Zisserman [44] proposed VGG16, a convolutional neural network for the recognition of images from a large dataset.

Further, we used the transfer learning method to train our facial expression model to foster the accuracy of predicted data [45]. We employed ImageNet weights and the VGG16 model with the image input size as $224 \times 224 \times 3$ [46]. The dense layer after the flattened layer is fed with 256 units with an activation function as Relu (refer to (1)) [47]. In the output block, the Dense layer is fed with the 6 units and softmax as an activation function (refer to (2)). The model is compiled with an Adam optimizer with a learning rate of 0.001 and categorical_crossentropy as a loss function with accuracy metrics (refer to (3)) [48]. Further, the model is trained over 50 epochs with 32 as batch size and categorical as class mode.

$$R(z) = \max(0, z) \qquad (1)$$

$$\sigma(\vec{z})_i = \frac{e^{Z_i}}{\sum_{J=1}^{K} e^{Z_i}} \qquad (2)$$

[Φ]https://www.thethingsindustries.com/docs/integrations/cloud-integrations/ubidots/#:~:text=Ubidots%20provides%20a%20secure%20and,that%20data%2C%20and%20visualizing%20it.
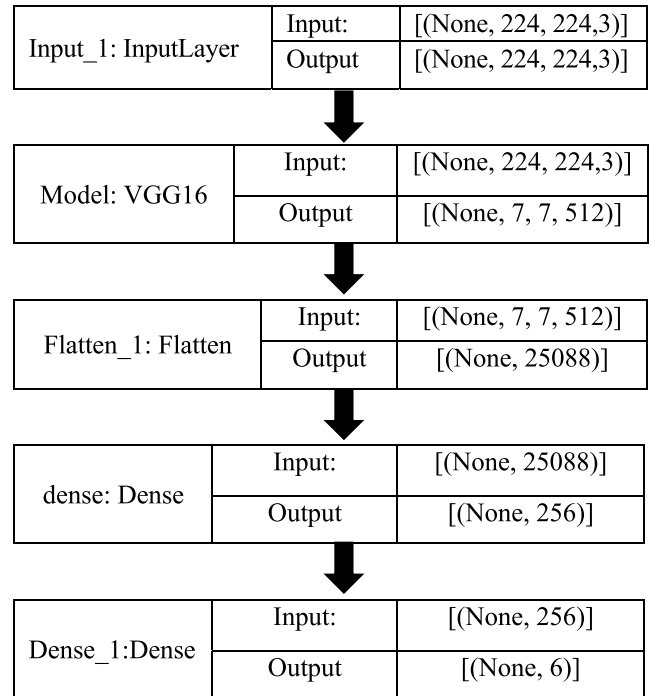
[φ]https://iot-fpms.fandom.com/wiki/Ubidots

Input_1: InputLayer	Input:	[(None, 224, 224,3)]
	Output	[(None, 224, 224,3)]

Model: VGG16	Input:	[(None, 224, 224,3)]
	Output	[(None, 7, 7, 512)]

Flatten_1: Flatten	Input:	[(None, 7, 7, 512)]
	Output	[(None, 25088)]

dense: Dense	Input:	[(None, 25088)]
	Output	[(None, 256)]

Dense_1:Dense	Input:	[(None, 256)]
	Output	[(None, 6)]

FIGURE 6. The architecture of the trained model.

$$Loss = -\sum_{i=1}^{outputsize} y_i . \log \hat{y}_i \qquad (3)$$

Here in (1), $-Z$ is the input value, in (2): σ is softmax, $\vec{z}$ is the input vector, e^{Z_i} is a standard exponential function for input vector, K is the number of classes (in our case it is 6), e^{Z_i} standard exponential function for output vector.

b: SPEECH EMOTION ANALYSIS

For analyzing the emotions or sentiment, we had proposed layered architecture of the model (in Fig. 7) using a Bidirectional Long Short Term Memory network (Bi -LSTM) [49]. In this network, we feed the features in the form of a Mel scaled spectrogram. Bi-LSTM is used when we want our model to observe the sequential information in both forward and backward directions. We had formed a model using 2 Bidirectional LSTMs and 2 blocks of Dense and dropout layers. Further Dense layer and output layer complete the model architecture (refer to Fig. 7). The model is compiled with rmsprop optimizer and sparse_categorical_crossentrophy as a loss function [50]. Further, the model is trained over 100 epochs and has a batch size of 64. The trained model will take the speech data in the form of a Mel scaled spectrogram and feed it into the prediction pipeline to detect the sequence and classify the features to their resultant category.

c: CONTENT-BASED EMOTION ANALYSIS

For analyzing the emotions from the written text data, we proposed a multinomial naïve Bayes approach [51]. For feeding

the data in this model we first generated the Bags of words in unigram form. Processed data is converted to vector form using the count vectorization technique (refer to (4) and (5)). Once we are ready with the bag of words, we fit the whole data into vector form. Now vectors are fed to multinomial naïve bayes classifier to train the data.
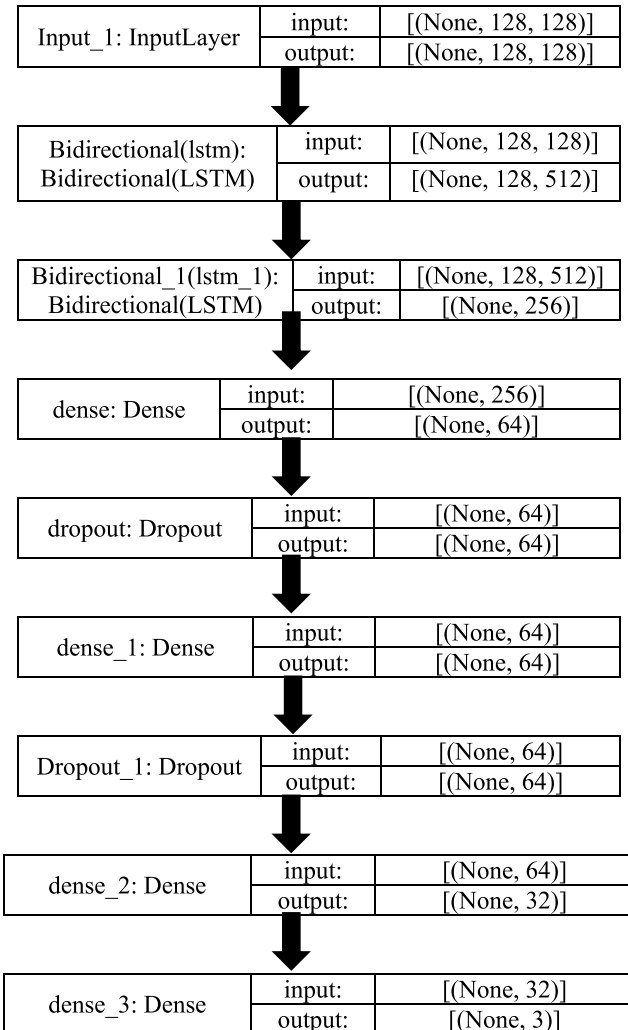
| Input_1: InputLayer | input: | [(None, 128, 128)] |
| | output: | [(None, 128, 128)] |

| Bidirectional(lstm): Bidirectional(LSTM) | input: | [(None, 128, 128)] |
| | output: | [(None, 128, 512)] |

| Bidirectional_1(lstm_1): Bidirectional(LSTM) | input: | [(None, 128, 512)] |
| | output: | [(None, 256)] |

| dense: Dense | input: | [(None, 256)] |
| | output: | [(None, 64)] |

| dropout: Dropout | input: | [(None, 64)] |
| | output: | [(None, 64)] |

| dense_1: Dense | input: | [(None, 64)] |
| | output: | [(None, 64)] |

| Dropout_1: Dropout | input: | [(None, 64)] |
| | output: | [(None, 64)] |

| dense_2: Dense | input: | [(None, 64)] |
| | output: | [(None, 32)] |

| dense_3: Dense | input: | [(None, 32)] |
| | output: | [(None, 3)] |

FIGURE 7. The architecture of the trained model.

Model is trained over 3 different classes i.e., positive, negative, and neutral.

$$p\left(C_k \mid x\right) = \frac{p\left(C_k\right) p\left(x \mid C_k\right)}{P(x)} \qquad (4)$$

$$p\left(x \mid C_k\right) = \frac{\left(\sum_{i=1}^{n} x_i\right)!}{\prod_{i=1}^{n} x_i!} \prod_{i=1}^{n} p k_i^{x_i} \qquad (5)$$

Here $p\left(C_k \mid x\right)$ is the posterior probability, $p\left(C_k\right)$ is the prior probability of C_k, $p\left(x \mid C_k\right)$ is the likelihood, $P(x)$ is the probability of x, x is a feature vector $(x_1, x_2, x_3, \ldots, x_n)$.

The data flow layout of the model used in the current scenario is illustrated in Fig. 8.

2) NORMALIZATION CRITERIA

Normalization is the process of rescaling or transforming the data such that each category type has a uniform distribution [52]. Many techniques have been deployed for the normalization process of data distribution over a specified range [53]. For our purpose, we used the max normalization technique (a variant of min-max) where each feature is scaled over a range of $+1$ to -1. The rescaling is performed by dividing each feature value by its maximum value. The following equation 6 is used to obtain rescaling data.

$$x'_{i,n} = \frac{x_{i,n}}{\max\left(|x_i|\right)} \qquad (6)$$

where $i \in f$, and $n \in N$, f is the features and N is the number of instances. This approach is selected based on its simplicity, for preserving the relationship among the original inputs [54].
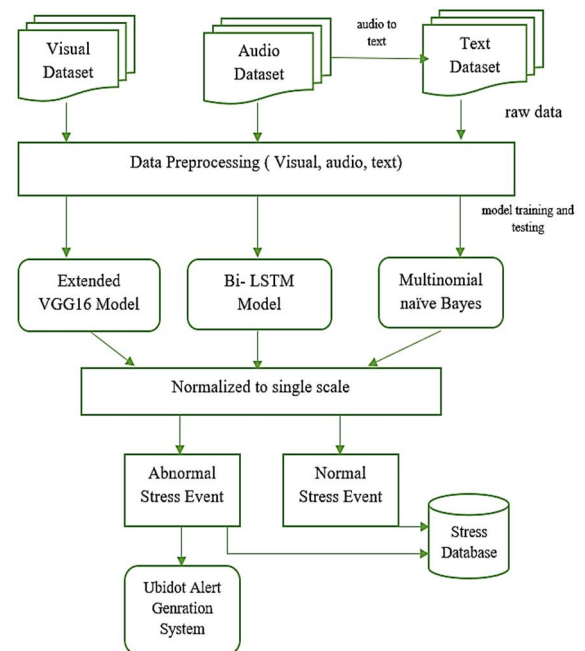

FIGURE 8. Data flow layout of the proposed model.

3) STRESS IDENTIFICATION

The stress identification is a binary function, able to provide two outputs; activate and deactivate. The stress event is activated when the results from all three emotion analysis models are more on the negative side. The stress identification process also takes up the historical data and compares it with the current results, and generates output based on the current and previous results. If the results continue on the negative side, an activation message is sent to the alert generation system for updating the stress outcomes to the students and their belongings.

4) ALERT GENERATION

The stress identification process sends an activation message to the alert generating system whenever the stress results

of a student exceed the threshold value. The ubidots IoT development platform is used for sending alerts to caregivers and family members at the cloud layer. ubidots is a free-to-use and secure IoT platform to provide IoT-based facilities and solutions, especially to students and researchers. The ubidot contains the students' information like name, address, contact number of student and belongings.

5) STUDENT DATABASE

The student stress database contains the results of the previous counsellings of the students. Each entry in the database contains name, personal information, counseling date, facial analysis, vocal, and content analysis results. The database provides the facility to check the results at any time and is able to perform comparisons also. Further, the database also has the features of plotting results on graphs and provides visualization facilities.

IV. RESULT AND ANALYSIS

The proposed model is tested against a valid set of inputs collected from a series of students' counselings. The raw data after preprocessing stage (refer to section III.C) is fed to the prediction model at the cloud layer. This categorical data is analyzed with the respective predictor models (as explained in sections III.D.1.a to III.D.1.c) and is expected to generate an alert based on the activation of the stress event. This section provides an overview of the training and validation results of the proposed IoT-Fog-based system. The acceptance of the presented work depends upon the accuracy and validation of the trained model♠. The training and validation accuracy of the facial emotion detection method is visualized in Fig. 9. It is observed that the model shows a significant amount of satisfactory results in both the testing and validation phases, thus yielding a good overall model fitting. It is also noticed that the performance of the model is growing over time employing the model improvement with the experience (learning).

Another important metric to measure the fitness of the model is by observing its validation loss and training loss over the data$^\alpha$. The training loss indicates how well the model is fitting the training data, while the validation loss indicates how well the model fits new data$^\beta$. The training loss and validation loss for the visual emotion analysis model are illustrated in Fig. 10. The training loss going down over time is an indication of achieving low error values. It is seen that the gap between the training loss and validation loss is comparatively lower, in the beginning, and slightly increases over time indicating the goodness of model fitting.

Another aspect of emotion detection is by analyzing the voice samples of the students during counseling. The results

♠https://towardsdatascience.com/train-validation-and-test-sets-72cb40c ba9e7

αhttps://machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-performance/

βhttps://rstudio-conf-2020.github.io/dl-keras-tf/notebooks/learning-curve-diagnostics.nb.html
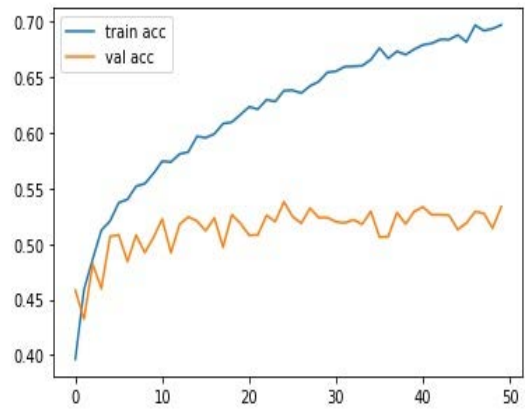
FIGURE 9. Training and validation accuracy for facial detection.

of the speech emotion analysis models are visualized in Fig. 11. It represents the training accuracy, validation accuracy, training loss, and validation loss metrics. The results reveal that the model yields a low degree error specifying the model is a good fit over the testing and training data set.
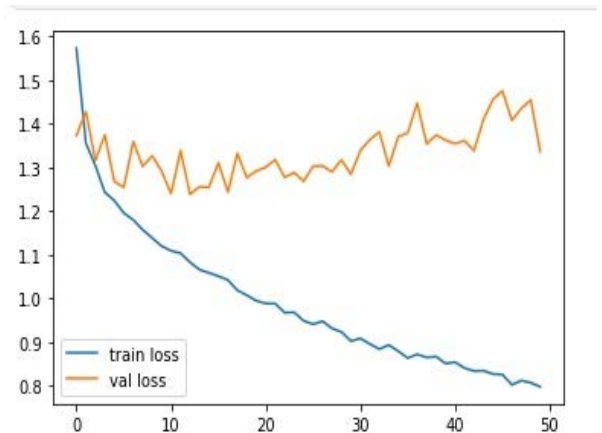
FIGURE 10. Training and validation loss for facial emotion detection.

For the content-based analysis, we implemented a multinomial naïve Bayes model approach. This text categorization model is tested over the content of the students' speech to extract meaningful emotions out of the textual data. During the training process of the model, on average it yields around 94.23% accuracy and for validation, the model results in 90.11% accuracy.

We generalize the final output received from all the models on a uniform scale of +1 to −1. The detection of the positive emotions (happy or joy) rectify the results more on the positive side of the scale. On the other hand, the extraction of a great amount of negative stress-based emotions will take the final results more on the negative side of the scale. The consequence of obtaining a negative result will activate
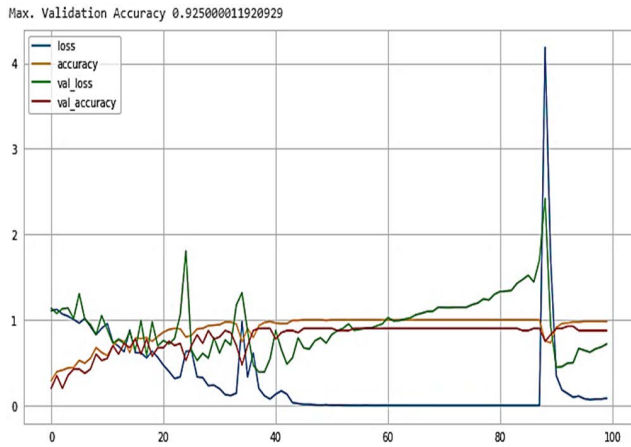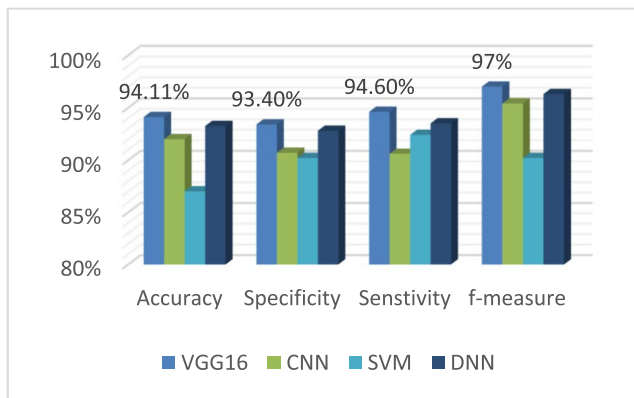
FIGURE 11. Metrics for speech emotion analysis.

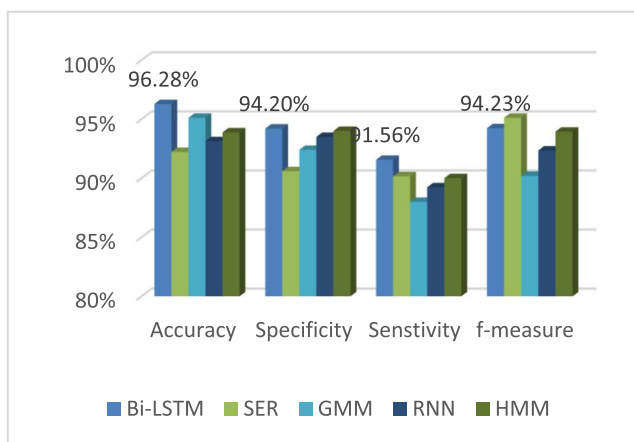**FIGURE 12. Comparison of facial emotion detection models.**

FIGURE 13. Comparison of vocal emotion detection models.

the stress event. In case of the activation of the stress event, the Ubidot IoT platform which has pre-saved information about the students' belongings will send an alert to them. The alert will send via SMS, and also an email will be sent to the caretakers and the student themselves.

The comparison results of our selected model against a few well-known models are shown in Fig. 12, Fig. 13, and Fig. 14 respectively. The comparison is performed based on the performance parameters accuracy, specificity, sensitivity, and f-measure. It is observed from the results that the models used in this study are far more efficient than other techniques.
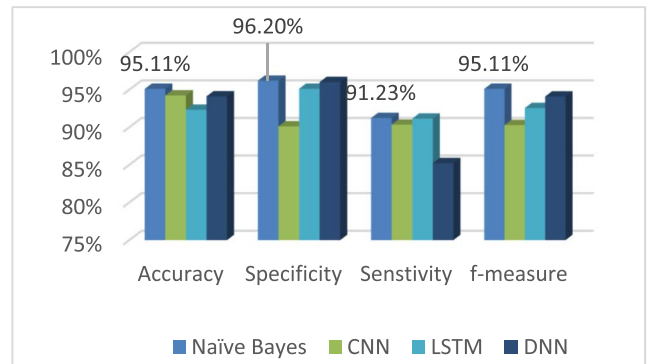

FIGURE 14. Comparison of content-based emotion detection models.

V. DISCUSSION

The IoT framework is utilized is to recognize the students under severe stress conditions at the earliest stages to decrease the dropout rates in the education system along with other severe consequences to the physical and mental health of the learners. Stress and anxiety issues are among the primary sources of the decreasing grades, absenteeism, and dropout of the students. The conducted counseling is measured based on the three parameters: facial expressions, speech measures, and the content of the student's speech. The emotion detection parameters are selected in a way to get the analysis from every aspect of the emotion measure.

(a) The facial expressions are the initial representer of stress and depression feelings. The effective study of the facial expression will benefit the overall evaluation of the stress monitoring system. The conducted study monitors the change in the expression on the face during the counseling and computes results based on some known stress emotion parameters. For this purpose, we trained and used a three-layered machine learning model VGG16. The model is selected based on performance, compilation time, simplicity, training capability, and learning rate. The results section depicts the performance of the facial expression detection system against the accuracy of the training data set and the validation accuracy when run on the real-world data set. It was observed that there are only slight significant variations in the accuracy performance of the training and testing dataset.

(b) The second parameter is the analysis of the speech texture of the candidate. The voice data input is captured with help of a microphone installed on the candidates. The difference in the speech pitch, the hesitation sounds, time to respond, etc. are all counted from the

obtained voice sample and then processed against the stress or depression parameters. For analysis purposes, we used a layered architectural model employing a Bidirectional Long Short Term Memory network. The model is selected for its feature of a sequential information processing mechanism in both forward and backward directions which makes it effective for lateral changes. It was observed from the results that a fine range of accuracy has been obtained by the model when implemented on the testing data set. The voice sample along with the facial expression detection model yields a high degree of effective results as compared when counted individually.

(c) The final analysis is based on the content of the spoken data. The data collected using voice samples are first converted in the textual form (using Google API) to process against the emotion analyzer based on the textual information. The data is matched against three sentimental categories namely; positive, negative, and neutral. These sentiments are derived from their respective emotion classes. The emotion investigation process is performed by the multinomial Naïve Bayes method. The input to the model is primarily converted into data vectors for training and testing purposes. The idea of the content analysis is to check the texture of the written discourse of the student under counseling. The presence of negative emotions signifies the chances of the student being under stressful conditions. Moreover, the content analysis will aid in concluding the validation of the overall stress monitoring system.

(d) The three measures of stress monitoring techniques follow different aspects of stress detection. To conclude the overall outcomes, each phase is needed to be brought on a single scale. The normalized scale is selected between $+1$ to -1, where the positive side indicates positive emotions, and the negative side is for negative emotions. An alert generation system has been employed to generate the alert when a threshold value is crossed. The threshold value is set up based on the number of negative emotions that corresponds to the total number of emotions.

VI. CONCLUSION AND FUTURE RECOMMENDATIONS

In this study, the IoT-Fog-based framework is proposed to generate real-time alerts depicting the level of stress while a student undergoes psychological counseling. The objective is to predict the stress emotions in earlier stages, to avoid undesirable situations like a decrease in grading, an increase in absentee, school/college dropouts. The three-phase model depicts the endangered students on a normalized scale of $+1$ to -1. The results demonstrate the working of the proposed model against the provided data set and it shows a high degree of effective outcomes when compared to the testing and validation results. The model is significant to implement in the education system and will aid the students, their parents, teachers, etc. to get the prior intimation of stress-prone

students. The main limitation of the study is to focus on the current sample while ignoring any historical record. In the future, the model is promised to extend over the inclusion of large sample data along with the addition of historical data, and student health records.

REFERENCES

[1] G. Hägg and J. Gabrielsson, "A systematic literature review of the evolution of pedagogy in entrepreneurial education research," *Int. J. Entrepreneurial Behav. Res.*, vol. 26, no. 5, pp. 829–861, May 2019.

[2] H. J. Smith, "A critique of the teaching standards in England (1984–2012): Discourses of equality and maintaining the status quo," *J. Educ. Policy*, vol. 28, no. 4, pp. 427–448, Jul. 2013.

[3] J. K. Behera, "Digital transformation and its impact: An analytical study," in *Digitization of Economy and Society*. Palm Bay, FL, USA: Apple Academic Press, 2021, pp. 27–49.

[4] N. P. Stromquist and K. Monkman, "Defining globalization and assessing its implications for knowledge and education, revisited," *Globalization Educ., Integr. Contestation Across Cultures*, vol. 1, pp. 1–21, Mar. 2014.

[5] J. Komljenovic, "The rise of education rentiers: Digital platforms, digital data, and rents," *Learn., Media Technol.*, vol. 46, no. 3, pp. 320–332, 2021.

[6] M. Escueta, "Education technology: An evidence-based review," Tech. Rep., 2017.

[7] J. Accuosti, "Factors affecting education technology success," in *Proc. ASEE Zone I Conf.*, Apr. 2014, pp. 1–6.

[8] B. Bruggeman, J. Tondeur, K. Struyven, B. Pynoo, A. Garone, and S. Vanslambrouck, "Experts speaking: Crucial teacher attributes for implementing blended learning in higher education," *Internet Higher Educ.*, vol. 48, Jan. 2021, Art. no. 100772.

[9] R. A. Rasheed, A. Kamsin, and N. A. Abdullah, "Challenges in the online component of blended learning: A systematic review," *Comput. Educ.*, vol. 144, Jan. 2020, Art. no. 103701.

[10] M. Bond, "Facilitating student engagement through the flipped learning approach in K-12: A systematic review," *Comput. Educ.*, vol. 151, Jul. 2020, Art. no. 103819.

[11] A.-J. Moreno-Guerrero, R. Soler-Costa, J.-A. Marín-Marín, and J. López-Belmonte, "Flipped learning and good teaching practices in secondary education," *Comunicar*, vol. 29, no. 68, pp. 107–117, Jul. 2021.

[12] K. Ratih, M. F. J. Syah, N. Nurhidayat, S. Jarin, and J. Buckworth, "Learning patterns during the disruptive situation in informal education: Parents' efforts and challenges in the adjustment of progressive learning," *Indonesian J. Learn. Adv. Educ.*, vol. 3, no. 3, pp. 180–193, 2021.

[13] S. Hidi, "An interest researcher's perspective: The effects of extrinsic and intrinsic factors on motivation," in *Intrinsic and Extrinsic Motivation*, 2000, pp. 309–339.

[14] M. Pulido-Martos, J. M. Augusto-Landa, and E. Lopez-Zafra, "Sources of stress in nursing students: A systematic review of quantitative studies," *Int. Nursing Rev.*, vol. 59, no. 1, pp. 15–25, Mar. 2012.

[15] M. S. Cebelin and C. S. Hirsch, "Human stress cardio-myopathy-myocardial lesions in victims of homicidal assaults without internal injuries," in *Human Pathology*, vol. 2, 1980, pp. 123–132.

[16] J. B. Jemmott and S. E. Locke, "Psychosocial factors, immunologic mediation, and human susceptibility to infectious diseases: How much do we know?" *Psychol. Bull.*, vol. 95, no. 1, pp. 78–108, 1984.

[17] C. Dunkel-Schetter and M. Lobel, "Stress among students," *New Directions for Student Services*, vol. 1990, no. 49, pp. 17–34, Mar. 1990.

[18] N. K. Giang, M. Blackstock, R. Lea, and V. C. M. Leung, "Developing IoT applications in the fog: A distributed dataflow approach," in *Proc. 5th Int. Conf. Internet Things (IoT)*, Oct. 2015, pp. 155–162.

[19] P. Verma and S. K. Sood, "Fog assisted-IoT enabled patient health monitoring in smart homes," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1789–1796, Jun. 2018.

[20] M. Mozafari, F. Firouzi, and B. Farahani, "Towards IoT-enabled multi-modal mental stress monitoring," in *Proc. Int. Conf. Omni-layer Intell. Syst. (COINS)*, Aug. 2020, pp. 1–8.

[21] S. Uday, C. Jyotsna, and J. Amudha, "Detection of stress using wearable sensors in IoT platform," in *Proc. 2nd Int. Conf. Inventive Commun. Comput. Technol. (ICICCT)*, Apr. 2018, pp. 492–498.

[22] T. A. Shaikh and R. Ali, "Fog-IoT environment in smart healthcare: A case study for Student stress monitoring," in *Fog Computing for Healthcare 4.0 Environments*, 2021, pp. 211–250.

[23] R. Kocielnik, N. Sidorova, F. M. Maggi, M. Ouwerkerk, and J. H. D. M. Westerink, "Smart technologies for long-term stress monitoring at work," in *Proc. 26th IEEE Int. Symp. Comput.-Based Med. Syst.*, Jun. 2013, pp. 53–58.

[24] O. Oti, I. Azimi, A. Anzanpour, A. M. Rahmani, A. Axelin, and P. Liljeberg, "IoT-based healthcare system for real-time maternal stress monitoring," in *Proc. IEEE/ACM Int. Conf. Connected Health, Appl., Syst. Eng. Technol.*, Sep. 2018, pp. 57–62.

[25] P. Verma and S. K. Sood, "A comprehensive framework for Student stress monitoring in fog-cloud IoT environment: M-health perspective," *Med. Biol. Eng. Comput.*, vol. 57, no. 1, pp. 231–244, Jan. 2019.

[26] 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. Informat.*, vol. 15, no. 1, pp. 481–489, Jan. 2019.

[27] L. Rachakonda, S. P. Mohanty, E. Kougianos, and P. Sundaravadivel, "Stress-lysis: A DNN-integrated edge device for stress level detection in the IoMT," *IEEE Trans. Consum. Electron.*, vol. 65, no. 4, pp. 474–483, Nov. 2019.

[28] Z. Sheng, S. Yang, Y. Yu, A. Vasilakos, J. McCann, and K. Leung, "A survey on the IETF protocol suite for the Internet of Things: Standards, challenges, and opportunities," *IEEE Wireless Commun.*, vol. 20, no. 6, pp. 91–98, Dec. 2013.

[29] A. Alberdi, A. Aztiria, and A. Basarab, "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review," *J. Biomed. Inform.*, vol. 59, pp. 49–75, Feb. 2016.

[30] Cui, Zhang, Li, Luo, Zhang, and Fu, "WSMS: Wearable stress monitoring system based on IoT multi-sensor platform for living sheep transportation," *Electronics*, vol. 8, no. 4, p. 441, Apr. 2019.

[31] G. Li, K. Ota, M. Dong, J. Wu, and J. Li, "DeSVig: Decentralized swift vigilance against adversarial attacks in industrial artificial intelligence systems," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3267–3277, May 2020.

[32] H. Qassim, A. Verma, and D. Feinzimer, "Compressed residual-VGG16 CNN model for big data places image recognition," in *Proc. IEEE 8th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2018, pp. 169–175.

[33] F. Chollet. (2016). *The Keras Blog. The Keras Blog ATOM*. Accessed: Dec. 7, 2021. [Online]. Available: https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html

[34] R. Rajak and R. Mall, "Emotion recognition from audio, dimensional and discrete categorization using CNNs," in *Proc. TENCON IEEE Region Conf. (TENCON)*, Oct. 2019, pp. 301–305.

[35] S. Gurbuz, J. N. Gowdy, and Z. Tufekci, "Speech spectrogram based model adaptation for speaker identification," in *Proc. IEEE SoutheastCon. Preparing New Millennium*, Apr. 2000, pp. 110–115.

[36] M. Saravanan, P. C. R. Raj, and S. Raman, "Summarization and categorization of text data in high-level data cleaning for information retrieval," *Appl. Artif. Intell.*, vol. 17, nos. 5–6, pp. 461–474, May 2003.

[37] S. Vijayarani, M. J. Ilamathi, and M. Nithya, "Preprocessing techniques for text mining—An overview," *Int. J. Comput. Sci. Commun. Netw.*, vol. 5, no. 1, pp. 7–16, 2015.

[38] M. Wang and F. Hu, "The application of NLTK library for Python natural language processing in corpus research," *Theory Pract. Lang. Stud.*, vol. 11, no. 9, pp. 1041–1049, Sep. 2021.

[39] W. B. A. Karaa and N. Gribâa, "Information retrieval with porter stemmer: A new version for english," in *Advances in Computational Science, Engineering and Information Technology*, 2013, pp. 243–254.

[40] B. Issac and W. J. Jap, "Implementing spam detection using Bayesian and porter stemmer keyword stripping approaches," in *Proc. TENCON IEEE Region Conf.*, Nov. 2009, pp. 1–5.

[41] W. B. A. Karaa, "A new stemmer to improve information retrieval," *Int. J. Netw. Secur. Appl.*, vol. 5, no. 4, pp. 143–154, Jul. 2013.

[42] A. Kulkarni and A. Shivananda, "Converting text to features," in *Natural Language Processing Recipes*, 2021, pp. 63–106.

[43] T. Balomenos, A. Raouzaiou, S. Ioannou, A. Drosopoulos, K. Karpouzis, and S. Kollias, "Emotion analysis in man-machine interaction systems," in *Proc. Int. Workshop Mach. Learn. Multimodal Interact.*, 2004, pp. 318–328.

[44] H. Qassim, A. Verma, and D. Feinzimer, "Compressed residual-VGG16 CNN model for big data places image recognition," in *Proc. IEEE 8th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2018, pp. 169–175.

[45] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[46] T. Akter, M. H. Ali, M. I. Khan, M. S. Satu, M. J. Uddin, S. A. Alyami, S. Ali, A. Azad, and M. A. Moni, "Improved transfer-learning-based facial recognition framework to detect autistic children at an early stage," *Brain Sci.*, vol. 11, no. 6, p. 734, May 2021.

[47] S. Masood, A. Rai, A. Aggarwal, M. N. Doja, and M. Ahmad, "Detecting distraction of drivers using convolutional neural network," *Pattern Recognit. Lett.*, vol. 139, pp. 79–85, Nov. 2020.

[48] E. Pranav, S. Kamal, C. S. Chandran, and M. H. Supriya, "Facial emotion recognition using a deep convolutional neural network," in *Proc. 6th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, 2020, pp. 317–320.

[49] A. Ilapakurti, S. Kedari, J. S. Vuppalapati, S. Kedari, and C. Vuppalapati, "Artificial intelligent (AI) clinical edge for voice disorder detection," in *Proc. IEEE 5th Int. Conf. Big Data Comput. Service Appl. (BigDataService)*, Apr. 2019, pp. 340–345.

[50] L. He, D. Jiang, L. Yang, E. Pei, P. Wu, and H. Sahli, "Multimodal affective dimension prediction using deep bidirectional long short-term memory recurrent neural networks," in *Proc. 5th Int. Workshop Audio/Visual Emotion Challenge*, Oct. 2015, pp. 73–80.

[51] D. Indira, "An enhanced CNN-2D for audio-visual emotion recognition (AVER) using Adam optimizer," *Turkish J. Comput. Math. Educ. (TURCOMAT)*, vol. 12, no. 5, pp. 1378–1388, Apr. 2021.

[52] A. M. Kibriya, E. Frank, B. Pfahringer, and G. Holmes, "Multinomial naive Bayes for text categorization revisited," in *Proc. Australas. Joint Conf. Artif. Intell.*, 2004, pp. 488–499.

[53] B. M. Bolstad, R. A. Irizarry, M. Åstrand, and T. P. Speed, "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias," *Bioinformatics*, vol. 19, no. 2, pp. 185–193, 2003.

[54] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 105524.

[55] S. G. K. Patro and K. K. Sahu, "Normalization: A preprocessing stage," 2015, *arXiv:1503.06462*.

MADANJIT SINGH is currently a Research Student with the Department of Computer Science, Guru Nanak Dev University, Amritsar, India. He has four journal publications. His research interests include data analytics, educational data mining, and machine learning.

SARVESHWAR BHARTI received the Master of Computer Applications (M.C.A.) degree from the University of Jammu, Jammu, India, and the Ph.D. degree from Guru Nanak Dev University, Amritsar, India. He is currently an Assistant Professor at the Department of Computer Science, Guru Nanak Dev University. His research interests include software engineering, machine learning, sentiment analysis, and data science.

HARJIT KAUR has been working as an Assistant Professor with the Department of Computer Science, Guru Nanak Dev University, Amritsar, Punjab, India, since 2021. Her main research interests include machine learning, data analytics, and open source software.

MANEVPREET KAUR received the Master of Computer Applications degree from Guru Nanak Dev University, Amritsar, Punjab, India, in 2015. She has been working as an Assistant Professor with the Department of Computer Science, Guru Nanak Dev University, since 2015. She has three years of teaching. Her main research interests include software security and open source software.

VAIBHAV ARORA is currently pursuing the bachelor's degree with the Department of Computer Engineering and Technology, Guru Nanak Dev University, Amritsar, India. His research interests include data analytics and machine learning.

MUNISH SAINI received the Ph.D. degree in computer science and engineering from Guru Nanak Dev University, India. He is currently an Assistant Professor with the Department of Computer Engineering and Technology, Guru Nanak Dev University. His research interests include educational data mining, open source software, machine learning, and deep learning.

JASWINDER SINGH received the Ph.D. degree in computer science from Guru Nanak Dev University, Amritsar, India. He is currently an Assistant Professor with the Department of Computer Science, Guru Nanak Dev University. His research interests include distributed computing, machine learning, and data analytics.

● ● ●