International Journal of Electrical and Computer Engineering (IJECE)
Vol. 12, No. 4, August 2022, pp. 3981~3993
ISSN: 2088-8708, DOI: 10.11591/ijece.v12i4.pp3981-3993

Analysis of student sentiment during video class with multi-layer deep learning approach

Imrus Salehin[1], Nazmun Nessa Moon[1], Iftakhar Mohammad Talha[1], Md. Mehedi Hasan[1], Farnaz Narin Nur[2], Md. Azizul Hakim[1], Farhan Al Haque[1]
[1]Faculty of Science and Information Technology, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh
[2]Faculty of Science and Information Technology, Department of Computer Science and Engineering, Notre Dame University, Dhaka, Bangladesh

Article Info

Article history:

Received Jun 19, 2021
Revised Mar 23, 2022
Accepted Apr 5, 2022

Keywords:

Convolutional neural network
Deep learning
Electronic-learning
Image data
Sentiment analysis

ABSTRACT

The modern education system is an essential part of the rise of technology. The E-learning education system is not just an experimental system; it is a vital learning system for the whole world over the last few months. In our research, we have developed our learning method in a more effective and modern way for students and teachers. For significant implementation, we are implementing convolutions neural networks and advanced data classifiers. The expression and mood analysis of a student during the online class is the main focus of our study. For output measure, we divide the final output result as attentive, inattentive, understand, and neutral. Showing the output in real-time online class and for sensory analysis, we have used support vector machine (SVM) and OpenCV. The level of 5*4 neural network is created for this work. An advanced learning medium is proposed through our study. Teachers can monitor the live class and different feelings of a student during the class period through this system.

Corresponding Author:

Imrus Salehin
Faculty of Science and Information Technology, Department of Computer Science and Engineering,
Daffodil International University
Dhaka 1207, Bangladesh
Email: imrus15-8978@diu.edu.bd

1. INTRODUCTION

The online-based direct education system is one of the most popular education systems of this era. Teaching through the web, app, and software based video conferencing is a very popular teaching medium. It gives the teachers and students the opportunity to come closer to teach and take classes in a virtual way if any critical situation creates where the live class is impossible. Google Meet, Zoom, Skype are the very popular video conferencing medium for teaching. About 30-35 students can participate together on this learning platform, which can be conducted by one teacher. But for a teacher, understand teaching flexibility and need a clear idea of the best study environment. For more efficient learning, the development of virtual platforms is essential for all students. A structured virtual environment has been created in the United States, where teachers can enter virtual classrooms and they have designed the schedule as if they were doing the class on that typical day [1]. In our research, we created a model to find out student's impressions and expressions through a live face detection module. For the student's sentimental analysis, we have used the multi-task cascade convolutional neural network algorithm. Multiple face detection and multiple network training models have been used to complete our study. For multiple network models, we have used log-likelihood and hinge loss method. We first run a single model, then we average each model of a single network and we train

and test all the concatenated networks for the output response. In the face detection step, we used face cropping and slicing methods. In this method, the system will slice the facial image into different parts which are very important for expression identification. In the final result, support vector machine (SVM) is used to classify the genres of student's feelings that are marked as attentive, pleasant, understanding, and neutral. Datasets are created from multiple sources so that we can cover all aspects of online live video. After all the proceedings, the system will provide a ratio through a test and the training database. We have focused on OpenCV for multiple face detection and at the last position we applied support vector machine [2]. Our main goal is to find out a percentage of sentiment and multiple face detection. The main contribution of the research which is summarized as the following: i) we have designed a new rapid student sentiment approach (RSSA) model and an advanced backend framework for detecting facial expressions from running the virtual classroom, ii) develop e-learning to apply rapid methods to a combination of advanced image processing and deep learning, and iii) E-learning advanced method construction and outcome-based online learning system are the main key-point in our study.

The rest of the paragraph is complete in all parts as follows. In section 2, we are talking about existing associated works. Section 3 discusses the proposed model for the working protocol. The major contribution of the work is "live video facial sentiment analysis" which is illustrated in section 4. Section 5 contains the data processing and analysis for image accuracy. Finally, sections 6 and 7 mention the main result and conclude in the article.

2. LITERATURE REVIEW

The researcher worked on a comprehensive face detection system that is capable to detect multiple faces randomly. In their research, they have presented an effective method-based model which is the combination of four methods. They divided their method into two parts. For the first part they use Haar-AdaBoost, local binary pattern (LBP)-AdaBoost the booting algorithm to extract the data for the selection method, on the other hand, they use multiple kernel SVM with group features (GF-SVM), Gaussian fuzzy-neural network (GF-NN) to extract the characteristic for the classification method. They trained a large number of images to detect the face using neural networks in their model. They train the data to set both supervised and unsupervised learning to show the results in terms of detection rate and false detection rate. Among the four methods, the Haar-AdaBoost method performs better compared to other methods used in their methodology [3]. In this research, appearance-based face detection and tracking system from different video sequences is introduced through this work. In their proposed methodology they perform their work combining of APF-based tracking algorithm and AdaBoost face detection algorithm, namely boosted adaptive particle filter (BAPF) to perform automatic verification of the face, to detect random face from video track. They track the faces using the particle filtering into the input image frame and then the existing tracking faces detect the AdaBoost algorithm in their model. For their data set, they use 6000 face image size in pixels 20 by 20. Nonface images are taken from video sequences and use 6598 images to train the model. They use 25 layers to detect multi-view face detection in their proposed system [4]. To solve manual initialization of the tracking and face detection they have come up with an effective technique of the combination of both features based and appearance-based face detection in real-time. They present their system into two-folds. Face skin color detection is the first phase of their work. They use the Haar-like features AdaBoost algorithm on the other part of their work they extended the continuously adaptive mean shift (CAMSHAFT) algorithm to detect the same kind of objects like facial scale and lighting to compare the robustness of the system. They work with both static and dynamic face images from a complex background into their proposed model. Their detection system firstly classifies the given pixel image secondly recognizes various skin image detection. Finally, their model decided a face or non-face from the given input image frame [5].

3. PROPOSED MODEL

Universities, colleges, schools are very much affected with online learning and it is hard to find out the sentiment and circumstance of a student [6]. In our study, we have designed the RSSA model which is works through convolutional neural network and impression analysis techniques. In this section, we represent it with a geographical image view.

From the Figure 1 at first, Figure 1(a) is a device with a webcam for connecting a video call. Figure 1(b) is a full screen of video call mentors and active students. In this model, Figure 1(c) indicates the full image data store which contains a full chunk of visualization data program, and it is using for the next image processing experiment. After that, Figure 1(d) indicates the multiple face detection for impression analysis that means the programmable machine learning image training data. For single-frame face analysis

in Figure 1(e) is the main section of this model. Then from the database in Figure 1(f) we train and test the image dataset. Figures 1(g) and 1(h) is the result part of this model. Getting a visualization of the sentiment analysis calculation and measuring the model perfection decision is the main purpose of creating the model. In our system, we divide our output result as attentive, inattentive, understand, and neutral.

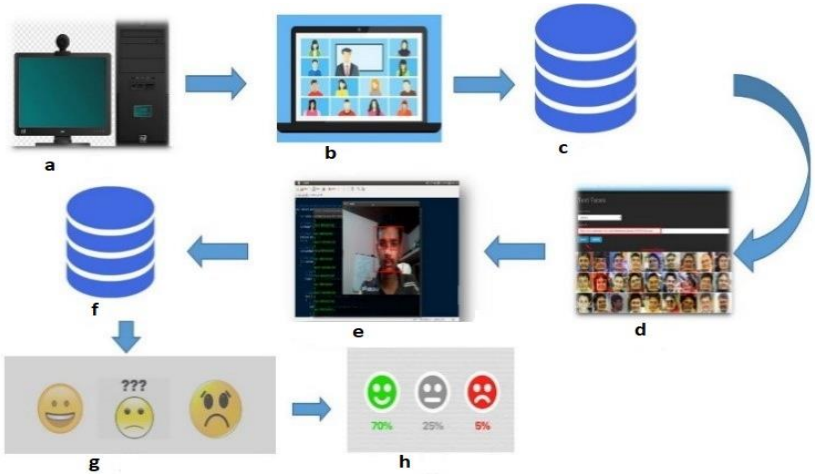


Figure 1. A short model of RSSA: (a) device with a webcam, (b) screen of video call, (c) image data storage, (d) multiple face detection, (e) main image detection and analysis, (f) train and test the image data, (g) analysis result, and (h) analysis result percentages

4. METHOD AND EXPERIMENTAL ANALYSIS

This section discusses the rapid structure and development of its algorithm. From Figure 2, the analysis process will start with a single frame and multi-frame. Using OpenCV and convolutional neural network (CNN), this model trains its intelligent system through training datasets. After completion of the training, it will produce output results as attentive, inattentive, understand, and neutral with a percentage measurement. Then we get the accuracy of the model using the SVM algorithm.

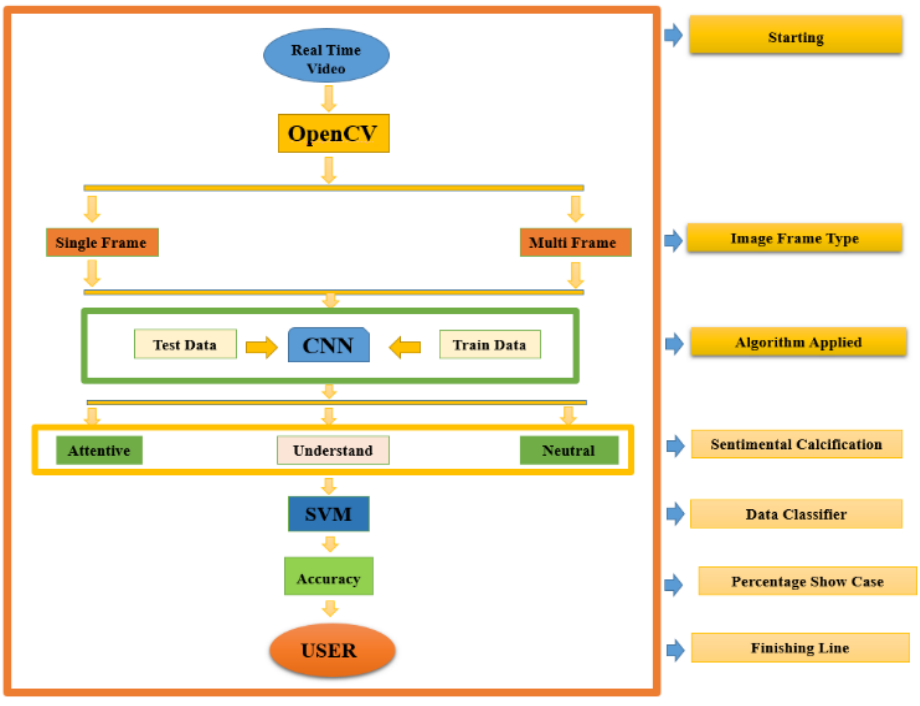


Figure 2. Working process of RSSA model

4.1. Real time video sequence pre-processing

Real-time face recognition to be a process that studies the input image, and works on the size, position, amount, and orientation of the face [7]. In the second phase of the framework, accurate face recognition is required to ensure that faces are extracted from each frame of the video sequence. Automatic and dynamic face detection (ADFD) is the basic programming framework that works with the new techniques OpenCV and Keras [8]. We represent this framework with a proper dimensional composite working process. From the online real-time video, we select the main multiple facial parts and store them for testing. We run our algorithm to train the ADFD framework. Deep metric learning is the main backend function to calculate accuracy.

$$D(x, x') = \sqrt{(x - x')\top M(x - x')} \quad (1)$$

The (1) implies that a Mahalanobis distance, which is used for calculating metric learning. In our study, accept a single input image and output a classification for that image is the main mechanism to recall the multiple face detection. In this system we apply the next method to data classifier. For the video sequence pre-processing, we collect many sample datasets from online video shown in Figure 3. Better accuracy and testing data level, we observe this model proper way.

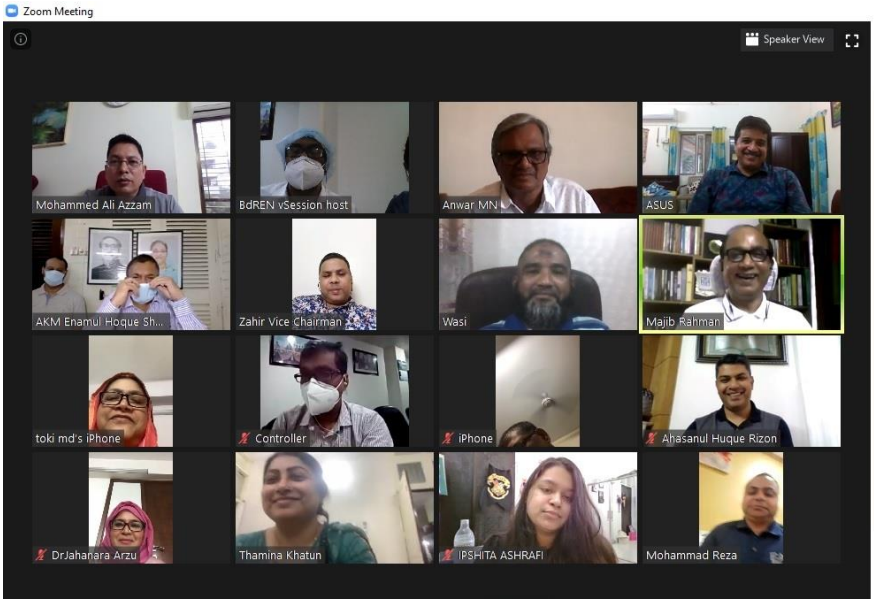


Figure 3. Sample facial data set of online

4.2. CNN for single image facial expression target matching

When we get images from the real-time video, we have used a face detection algorithm and also train by this image with a dataset. So that we get the best output and accuracy from a single image processing system. In the convolution approach, we see the shift-invariant artificial neural networks due to the shared-weights architecture for different types of characteristics. It also contains some hidden layers in image matching with mathematical calculations [9]. Here, *j* is the output node, data point is represented as *n*, *d* is called the target vale in here, and *y* is the value produced by the perceptron.

$$e_j(n) = d_j(n) - y_j(n) \quad (2)$$

And, another side we can write minimize the error in the entire output:

$$\varepsilon(n) = \frac{1}{2}\sum_j^n e^2(n) \quad (3)$$

It is more difficult to analyze the hidden node (layer) for the weight change, but we have shown the relevant formula (4).

$$-\frac{\partial \varepsilon(n)}{\partial v_j(n)} = \emptyset(v_j(n)) \sum_k^n -\frac{\partial \varepsilon(n)}{\partial v_k(n)} w_{kj}(n) \quad (4)$$

Derivative of the activation function is $\emptyset$ and local field V_j, W_k is weight. Starting from inputs, this model has three main stages, including convolution and max-pulling layer, a flat layer, fully connected two layers, and a Softmax layer for the output shown in Figure 4.

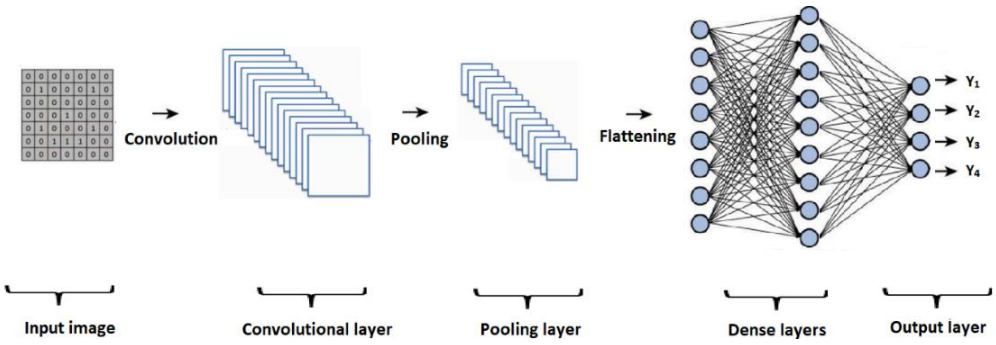


Figure 4. Convolution three major stages

4.3. Multiple network training

In our model, we used multiple network functionality. We need to perform multiple learning models. In order to learn the ensemble weights w, we individually train multiple CNN model and output their responses. A loss is defined on the weighted ensemble response, W adapted to reduce such loss. In this test, W is used to submit test responses.

$$\min_W -\sum_{i=1}^N log \sum_{k=1}^K P_k(y_i \mid \boldsymbol{X}_i) w_k + \lambda \sum_{k=1}^K w^2{}_k \quad (5)$$

$$s.t. \sum_{k=1}^K w_k = 1, w_k \geq 0, \forall k$$

In this function, the number of training samples is denoted as *N*, and the number of networks is denoted as *K*. $P_k(y_i|X_i)$ is the k^{th} network output response on the y^{th} category. For maximizing the validation accuracy λ is used.

$$\min_W \sum_{i=1}^N \sum_{y \neq y_i} \left[1 - \frac{\sum_{k=1}^K \left(P^{i,\ y_i}{}_k - P^{i,\ y}{}_k\right) w_k}{\gamma}\right]_+ + \lambda \sum_{k=1}^K w^2{}_k \quad (6)$$

$$s.t. \sum_{k=1}^K w_k = 1\ , w_k \geq 0, \forall k$$

The intuition is that the embedded output response combined with the ground truth should be larger than others with a margin of γ. Again, both γ and λ are determined with respect to the accuracy of the validation set [10].

4.4. Convolutional network architecture for proper impression identification

Convolutional network architecture is a major factor in research. In this section, we see some processing and identification of robust structures. In the first section, we are inputting real-time video preprocess image sets into the processing portion so that it can be converted with a neural network. To convert the neural network, this pre-process image has been extracted with the fracture extraction method. This is an advanced method to image data analysis with the networking process. Here, we are using n-number of dataset as shown in Figure 5 for vital and accurate measurement.

4.5. Facial expression landmark

In our study, when this system gets a face, it will detect that face. After identifying the face, this algorithm will first identify the important features of the face that are necessary for the extraction of expression of the face. The algorithm for facial expressions will detect eyebrows, eyes, nose, and lips [11], [12]. After identifying the important features, this system will slice the image into smaller parts: i) identify all

features including facial expressions, ii) lip detection, iii) eyes and eyebrows, iv) nose identification, and v) fragmenting the image into smaller pieces. From these small feature images, the algorithm will model the expression face with the important features identified like as Figures 6(a)-(e), which will then be used for expression detection.

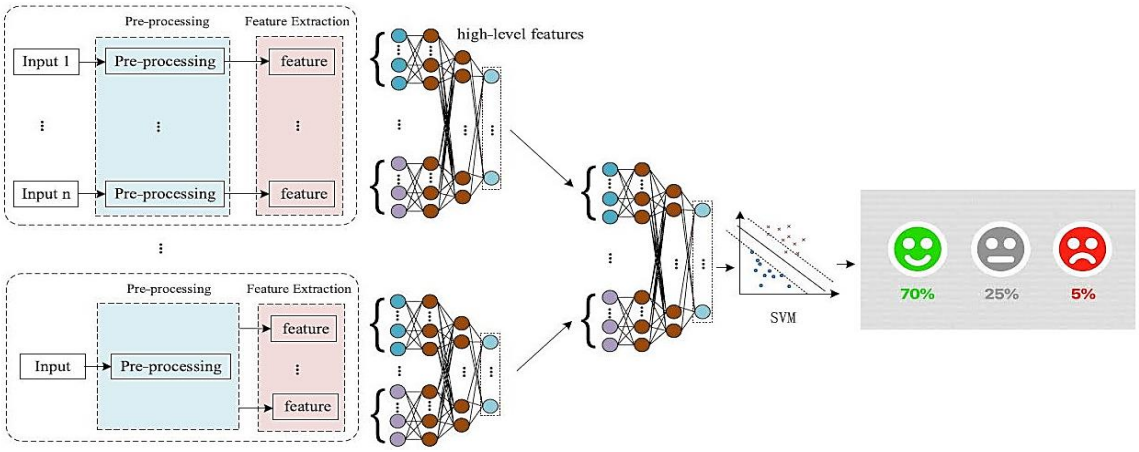


Figure 5. CNN and data fracture extraction architecture

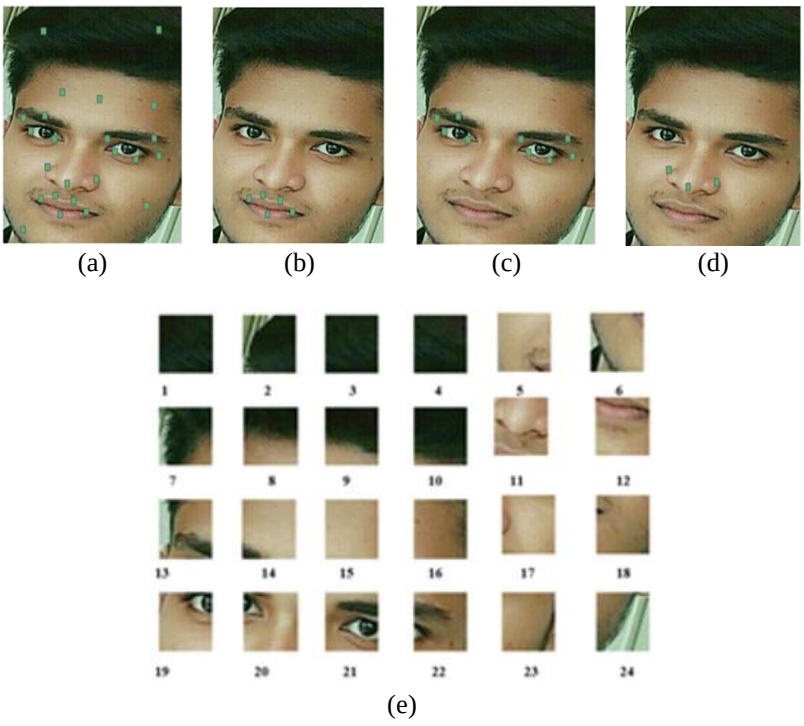


Figure 6. Identification of face features and image slicing (a) full face detection (b) lip detection (c) eye and eyebrows detection (d) nose detection (e) fragmenting the image into smaller pieces

4.6. SVM for sentimental classification

SVM classifier is applied in this study, which is widely used in sentiment recognition [13]. A simple and efficient recognition model is used in this study, which is similar to Chi-Jen [14]. This package provides a way to automate parameter selection and grid searches that we apply for data classification and sensing analysis. Support vector machine is using for that data classification. At the last stage of data identification from the image, we are applying it and get a result with accuracy.

5. DATA PROCESSING SYSTEM

This section explains yet which form of the dataset is used. How it is modified is to be used for the experiment and several mechanisms. Which will play an effective rule on our proposed model. Data processing system is very important and become more challenging to find out the sentiment of a student form the huge educational databases [15]. It is difficult to determine how best to prepare image data when a convolution neural network is being trained. The algorithmic conception and data processing flowchart have demonstrated in Figure 7.

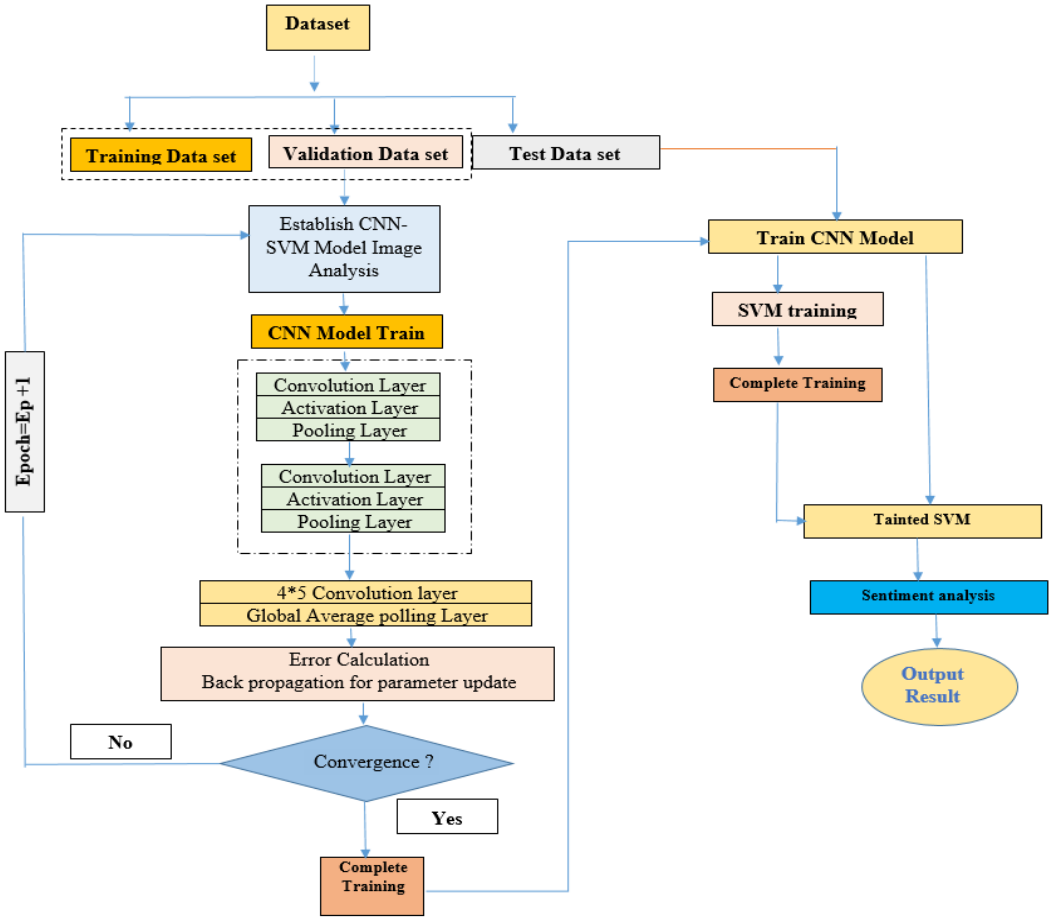


Figure 7. Data processing flowchart with algorithm conception

5.1. Data set

We divided our work into two parts. We proposed a multi-task learning network in the first part, which will explicitly revert the 3D morph able model (3DMM) parameters from a well-cropped 2D single face image and we named this as single face network (SFN). We designed our architecture for the multiple face network (MFN), which takes 2D image as input predicts the center location, the dimensions of the bounding box, and the 3DMM parameters for each face in the image through inspiring by you only look once (YOLO) [16] and its variants [17], [18]. We integrated multiple datasets to have a good training set for accurate prediction of each group of 3DMM parameters for single face retargeting. 300 W-LP comprises several large poses and the Face warehouse is a rich expression dataset. As static image test sets, labeled faces in the wild (LFW) and AFLW2000-3D are used and 300 VW is used for video tracking as a test set. Annotated faces in the wild (AFW) has pose angles, ground truth bounding boxes and 6 landmarks and is used as a test set for static images for multiple face retargeting, whereas face detection data set and benchmark (FDDB) and web image dataset for event recognition (WIDER) provide only ground truth regarding bounding boxes [19] and hence are used for training. Other video dataset is used to test our MFN performance on videos.

Table 1 shows that different types of data collection and its modalities, data type and participate list. Also, it shows the different types of data collection and its modalities, data type and participant list. This is a

different type of video data set for our analysis which is mentioned with proper information. More dataset details are summarized in Figure 8.

Table 1. Sentimental calcification dataset

Dataset	participant	Modalities	Emotional data type
Facebook Video Chat	10	Audio, Face Video	Discrete emotions (happy, sad, anger, disgust, fear)
Zoom	16	Audio, Physiological Video	Discrete emotions
Google Meet	15	Audio, Face Video	Discrete emotions
Skype	25	Audio, Face Video	Discrete emotions
Cisco WebEx	40	Audio, Face Video	Discrete emotions (amusement, sadness, anger, disgust, fear)

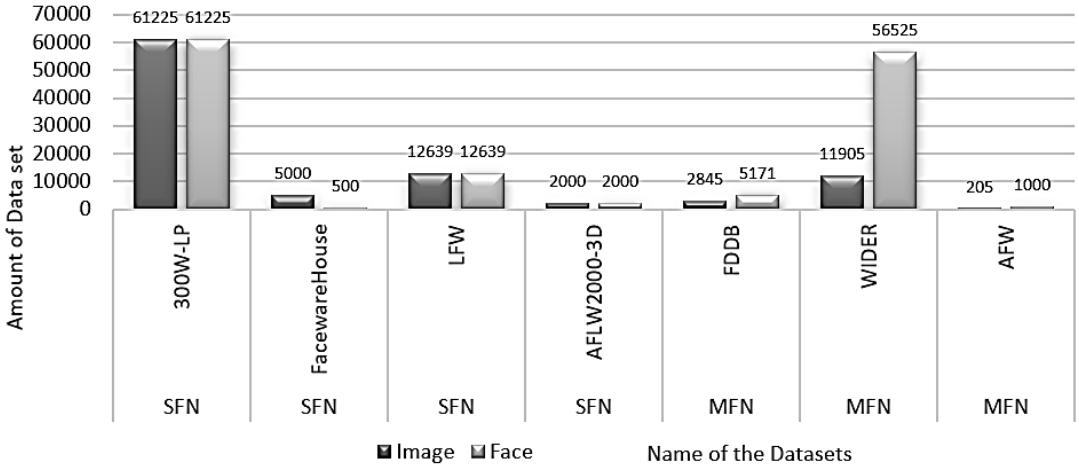


Figure 8. Number of images and faces for each training and testing dataset

5.2. Data analysis

We have studied various classifiers for the best algorithm for our research. We have performed their functional analysis by performing different sentimental analyzes as shown in Tables 2 to 4 in our database. For the facial expression identification, the output is showing us that we mainly focused on some machine learning classifier. The SVM accuracy rate is very high then the other classifier. We also have some very specialized algorithms for facial object detection or image processing so that we also applied the scale-invariant feature transform (SIFT) algorithm for checking accuracy ratio. SIFT algorithm has a special mechanism for an image processing system [20]–[22].

Table 2. A comparison of the methods used for classification on (attentive) training and test dataset

Classifier	Precision	Recall	F1-Score
Linear Regression	0.51	0.78	0.69
SVM	0.70	0.83	0.79
CNN	0.74	0.89	0.77
SIFT	0.56	0.72	0.68

Table 3. A comparison of the methods used for classification on (inattentive) training and test dataset

Classifier	Precision	Recall	F1-Score
Linear Regression	0.48	0.58	0.62
SVM	0.71	0.83	0.79
CNN	0.77	0.87	0.69
SIFT	0.56	0.71	0.79

Table 4. A comparison of the methods used for classification on (understand) training and test dataset

Classifier	Precision	Recall	F1-Score
Linear Regression	0.49	0.57	0.69
SVM	0.71	0.81	0.85
CNN	0.78	0.83	0.86
SIFT	0.51	0.78	0.62

From Table 4, we can see that the value of precision in case of linear regression and SIFT is 0.51 and 0.56. Where the precision values of SVM and CNN are 0.70 and 0.74. Which is higher than the value of linear regression and SIFT algorithm. We can also see that the value of Recall and F1 score is higher for

CNN and SVM in comparison of linear regression and SIFT algorithm. We can see further from Figure 9 that the accuracy of CNN and SVM is higher than linear regression and SIFT algorithm.

Table 3 also implies that the SVM and CNN algorithms are providing the best value for precision, recall, and F1-score. Also, the accuracy rate of SVM and CNN are quite high. Compared to Tables 2 to 4, we can see that the SVM and CNN algorithms provide the highest accuracy among all the other classification algorithms. From Figures 10 to 12, we can see that the accuracy rate of SVM and CNN algorithms [23], [24] is higher for attentive, inattentive, and understand genre. To do this separate calculation, we have divided our dataset into 3 genres (attentive, inattentive, and understand).

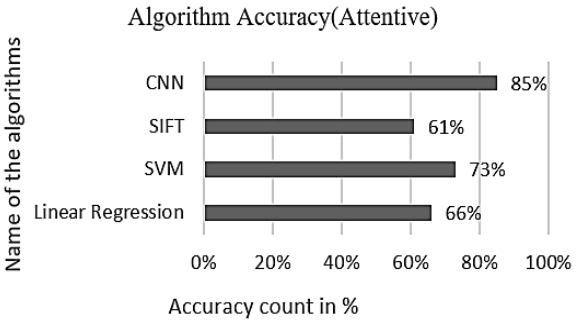


Figure 9. Algorithm accuracy for the attentive genre

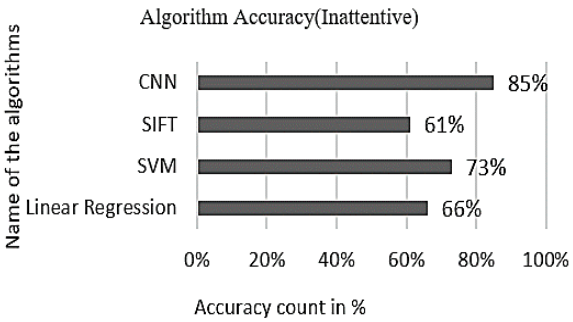


Figure 10. Algorithm accuracy for the inattentive genre

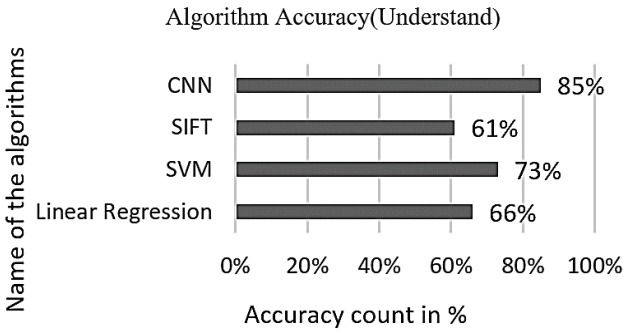


Figure 11. Algorithm accuracy for the understand genre

6. RESULT AND DISCUSSION

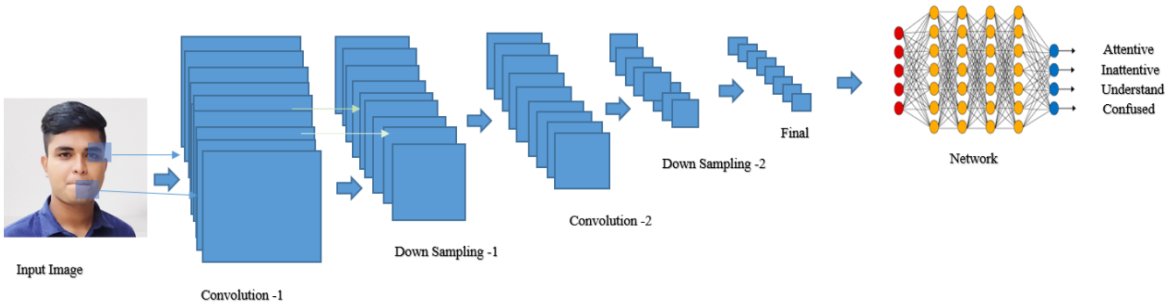
When our model receives the input image the system detects the important image features and then it will crop the image into smaller parts. The image cropping process will work step by step in the convolution and sampling stages which is shown in Figure 12. When the system receives the pre-processed image for identifying the expression, then the neural network algorithm processed the image and make a decision about the expression which is shown in Figure 12(a). By using the neural network algorithm our model system will display the output result as attentive, inattentive, understand, and neutral as well as with the percentage value described in Figure 12(b).

In this experiment we mainly run our system with single the CNN model. Then we are averaging each of the single CNN model, which is randomly initialized, pre-trained with randomly chosen dataset. Last of all each single CNN model assembled and then SVM is trained and tested on the concatenated network output responses.

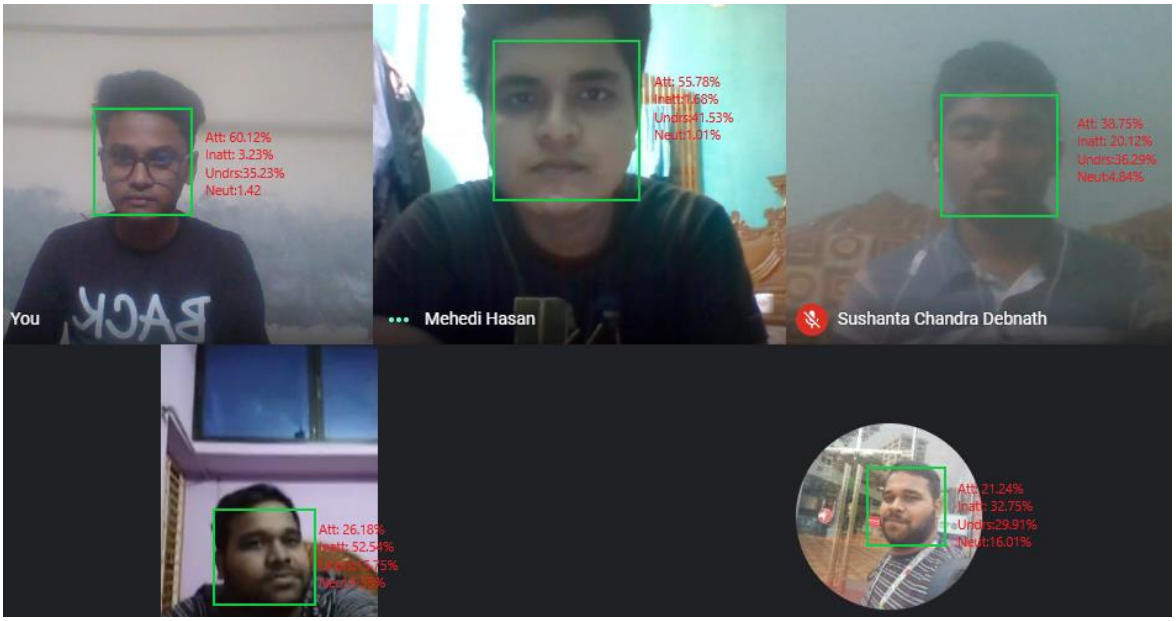
In this experiment we mainly run our system with single the CNN model. Then we are averaging each of the single CNN model which is randomly initialized, pre-trained with randomly chosen dataset. Last of all each single CNN model assembled and then SVM is trained and tested on the concatenated network output responses. The proposed work achieves the best performance. From Tables 5 and 6 we can determine that this model achieved 87.41% and 88.29% accuracy on the test set. Here, λ in the log likelihood loss and log likelihood loss is denoted as LL. Also, γ and λ in the hinge loss. In here hinge loss is denoted as HL.

From Table 7, we can see the total model precision, recall, and F1 Score. Among the 4-classification algorithm correctly predicted positive observations are 82% for SVM and 85% for CNN algorithm. We also

got recall and F1 score 86% and 81% for SVM algorithm and 87% and 84% for CNN algorithm. Which is a pretty good measure for a good prediction. From Figure 13 we got the accuracy of our proposed model and the predicted accuracy [25] of algorithms. Here CNN achieved accuracy is 88.29 and SVM achieved accuracy is 87.41.



(a)



(b)

Figure 12. Neural network model final results (a) Convolution model for final output and (b) generating attentive, inattentive, understand and neutral value

Table 5. Weights for the learned ensemble networks

	N # 1	N # 2	N # 3	N # 4	N # 5
LL (Log likelihood)	0.2371	0.2491	0.2987	0	0.1119
HL (Hinge Loss)	0.2308	0.2345	0.2806	0	0.1067

Table 6. Classification accuracy (%) of datasets

Accuracy	Single	AVG	SVM	LL	HL
Value	71.28	76.31	77.15	78.45	80.19
Test	74.08	80.24	85.16	87.41	88.29

Table 7. Compression of multiple classification algorithm

Classifier	Precision	Recall	F1 Score
Linear Regression	0.76	0.79	0.82
SVM	0.82	0.86	0.81
CNN	0.85	0.87	0.84
SIFT	0.73	0.77	0.80

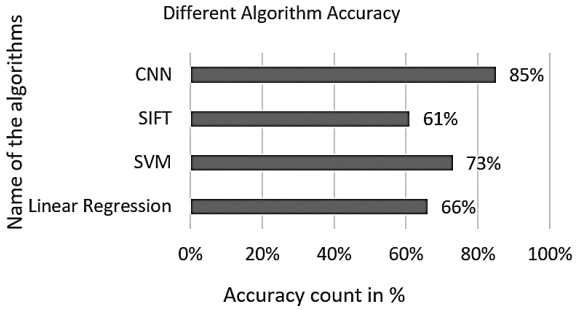


Figure 13. Model accuracy

7. CONCLUSION

Online based education is increasing in a vast way. Nowadays people are more reliable with online media and anything can be found in online. With the development of this modern technology, the online based education system becomes more famous and placed as more reliable for both students and teachers. In the physical class system, teachers can easily interact with students very easily and effectively. But in this online education system, it is not an easy task to understand a student's behavior and his conditions. Finding out if a student is attentive or inattentive in class or whether students are understanding the teacher's discourse is really a big problem. Whether the class should be made more interesting or informative. Through all of this analysis, we can make a prediction about a student how much he/she is active in class or depressed. A depressed mind cannot concentrate in studies. For this reason, expression analysis is a crucial part of the online education system. To do all this work we introduce this model so that we can determine whether a student is attentive or inattentive during the class period. We can determine all those matters through this proposed model and can solve problems. Through our proposed model it will produce a result that will inform that a student how much attentive or inattentive or understand the topic during the class. In this paper, we used a convolution neural network and support vector machine algorithm for predicting the student sentiment and producing the required accuracy with 4 genres (attentive, inattentive, understand, and neutral). For better experiments, we also perform several classification algorithms for finding better results.

ACKNOWLEDGEMENTS

The author acknowledges the support of the Daffodil International University Innovation Lab and Department of Computer Science Lab to develop this study. Especially we acknowledge the support in our honorable supervisor and faculty to develop our system.

REFERENCES

[1] G. Basilaia and D. Kvavadze, "Transition to online education in schools during a SARS-CoV-2 coronavirus (COVID-19) pandemic in Georgia," *Pedagogical Research*, vol. 5, no. 4, Apr. 2020, doi: 10.29333/pr/7937.

[2] A. Talele, A. Patil, and B. Barse, "Detection of real time objects using tensor flow and openCV," *Asian Journal of Convergence in Technology*, vol. 5, no. 1, pp. 1–4, 2019.

[3] H. Filali, J. Riffi, A. M. Mahraz, and H. Tairi, "Multiple face detection based on machine learning," in *2018 International Conference on Intelligent Systems and Computer Vision (ISCV)*, Apr. 2018, pp. 1–8, doi: 10.1109/ISACV.2018.8354058.

[4] W. Zheng and S. M. Bhandarkar, "Face detection and tracking using a boosted adaptive particle filter," *Journal of Visual Communication and Image Representation*, vol. 20, no. 1, pp. 9–27, Jan. 2009, doi: 10.1016/j.jvcir.2008.09.001.

[5] L.-K. Lee, S.-Y. An, and S.-Y. Oh, "Efficient face detection and tracking with extended CAMSHIFT and haar-like features," in *IEEE International Conference on Mechatronics and Automation*, Aug. 2011, pp. 507–513, doi: 10.1109/ICMA.2011.5985614.

[6] E. Corvee and F. Bremond, "Combining face detection and people tracking in video sequences," in *3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009)*, 2009, pp. P43----P43, doi: 10.1049/ic.2009.0271.

[7] D. Kabakchieva, "Predicting student performance by using data mining methods for classification," *Cybernetics and Information Technologies*, vol. 13, no. 1, pp. 61–72, Mar. 2013, doi: 10.2478/cait-2013-0006.

[8] G. H. P. Kumar, M. Ashwini, G. N. Divya, and B. N. Manjushree, "Multiple face detection and recognition in real-time using open CV," *International Journal of Engineering Research & Technology (IJERT)*, vol. 3, no. 27, 2015.

[9] S. Sawhney, K. Kacker, S. Jain, S. N. Singh, and R. Garg, "Real-time smart attendance system using face recognition techniques," in *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Jan. 2019, pp. 522–525, doi: 10.1109/CONFLUENCE.2019.8776934.

[10] T. Rao, X. Li, H. Zhang, and M. Xu, "Multi-level region-based convolutional neural network for image emotion classification," *Neurocomputing*, vol. 333, pp. 429–439, Mar. 2019, doi: 10.1016/j.neucom.2018.12.053.

[11] Z. Yu and C. Zhang, "Image based static Facial expression recognition with multiple deep network learning," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, Nov. 2015, pp. 435–442, doi: 10.1145/2818346.2830595.

[12] M. Lee, Y. K. Lee, M.-T. Lim, and T.-K. Kang, "Emotion recognition using convolutional neural network with selected statistical photoplethysmogram features," *Applied Sciences*, vol. 10, no. 10, May 2020, doi: 10.3390/app10103501.

[13] I. M. Talha, I. Salehin, S. C. Debnath, M. Saifuzzaman, N. N. Moon, and F. N. Nur, "Human behaviour impact to use of smartphones with the python implementation using naive bayesian," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Jul. 2020, pp. 1–6, doi: 10.1109/ICCCNT49239.2020.9225620.
[14] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, doi: 10.1109/72.991427.
[15] Y. Li, J. Zeng, S. Shan, and X. Chen, "Occlusion aware facial expression recognition using CNN with attention mechanism," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2439–2450, May 2019, doi: 10.1109/TIP.2018.2886767.
[16] W. Liu, W.-L. Zheng, and B.-L. Lu, "Emotion recognition using multimodal deep learning," in *International conference on neural information processing*, 2016, pp. 521–529.
[17] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, Apr. 2011, doi: 10.1145/1961189.1961199.
[18] A. M. Shahiri, W. Husain, and N. A. Rashid, "A review on predicting student's performance using data mining techniques," *Procedia Computer Science*, vol. 72, pp. 414–422, 2015, doi: 10.1016/j.procs.2015.12.157.
[19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.
[20] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 6517–6525, doi: 10.1109/CVPR.2017.690.
[21] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," *arXiv:1804.02767*, vol. 2018, pp. 1–6.
[22] A. Bulat and G. Tzimiropoulos, "How far are We from solving the 2D & 3D face alignment problem? (and a dataset of 230,000 3D faciallandmarks)," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 1021–1030, doi: 10.1109/ICCV.2017.116.
[23] I. Salehin, I. M. Talha, M. Saifuzzaman, N. N. Moon, and F. N. Nur, "An advanced method of treating agricultural crops using image processing algorithms and image data processing systems," in *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, Oct. 2020, pp. 720–724, doi: 10.1109/ICCCA49541.2020.9250839.
[24] I. Salehin, I. M. Talha, N. Nessa Moon, M. Saifuzzaman, F. N. Nur, and M. Akter, "Predicting the depression level of excessive use of mobile phone: decision tree and linear regression algorithm," in *2020 IEEE International Conference on Sustainable Engineering and Creative Computing (ICSECC)*, Dec. 2020, pp. 113–118, doi: 10.1109/ICSECC51444.2020.9557394.
[25] N. N. Moon *et al.*, "Natural language processing based advanced method of unnecessary video detection," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 6, pp. 5411–5419, Dec. 2021, doi: 10.11591/ijece.v11i6.pp5411-5419.

BIOGRAPHIES OF AUTHORS

Imrus Salehin completed BSc. In Computer Science and Engineering from Daffodil International University, Dhaka, Bangladesh. His interested research fields are Image Processing, Machine Learning and Data Mining. His also research interests include Data science and Computer vision. He has published several journals and conference papers. He is also a well-known reviewer of many journals and conferences. He can be contacted at email: imrus15-8978@diu.edu.bd.

Nazmun Nessa Moon is the Assistant Professor of the Department of Computer Science and Engineering at Daffodil International University. She received the B. Sc. degree in computer science and engineering from Rajshahi University of Engineering & Technology and M.Sc. in Information and Communication Technology from Bangladesh University of Engineering and Technology (BUET). From 2004 to 2007, she was a Lecturer. Her interested research fields are IoT, Digital Image Processing and Machine Learning. She can be contacted at email: moon@daffodilvarsity.edu.bd.

Iftakhar Mohammad Talha studied Computer Science and Engineering from Daffodil International University, Dhaka, Bangladesh. Main Interested research fields are Image Processing, Machine Learning and Data Mining, Data science and Computer vision. Mr. Iftakhar Mohammad Talha is main author of one paper in the international conferences and one Co-author in one Journal and 3 conferences. He has published several journals and conference papers. He is also a well-known reviewer of many journals and conferences. He can be contacted at email: iftakhar15-9019@diu.edu.bd

Md. Mehedi Hasan studied Computer science and engineering at Daffodil International University, Dhaka, Bangladesh. Main Interested research fields are image processing, machine learning and data mining, natural language processing. His also research interests include data science, internet of things (IoT) and Computer Vision. He can be contacted at email: mehedi15-9021@diu.edu.bd.

Farnaz Narin Nur is the Associate Professor of the Department of Computer Science and Engineering at Notre Dame University. She completed Master in Information Technology, IIT, DU. And PhD (Enrolled) in DU. Her interested research fields are network, database, digital logic design, computer fundamentals, and object oriented programming. She can be contacted at email: fernazcse@ndub.edu.bd.

Md. Azizul Hakim has completed his graduation from Ahsanullah University of Science and Technology, Dhaka, Bangladesh and post-graduation from United International University, Dhaka, Bangladesh in Computer Science and Engineering. Currently he is working as a Lecturer (Senior Scale) at Daffodil International University, Dhaka, Bangladesh. His research interests include deep learning, computer vision, data mining, and machine learning. He can be contacted at email: azizul.cse@diu.edu.bd.

Farhan Al Haque Al Haque is a young researcher who has completed his Bachelor's in Computer Science and Engineering with flying colors from Ahsanullah University Science and Technology, Dhaka, Bangladesh. He is now doing his post-graduation on the same subject from United International University, Dhaka, Bangladesh. Currently he has been employed as a lecturer at Daffodil International University, Dhaka, Bangladesh. His research interests include machine learning specially computer vision (deep learning). He can be contacted at email: Farhan.cse@diu.edu.bd.