Facial Emotion Detection using Convolutional Neural Network

Pooja Bagane, Shaasvata Vishal, Rohit Raj, Tanushree Ganorkar, Riya
Department of Computer Science and Engineering, Symbiosis Institute of Technology (SIT), Pune, India

***Abstract*—Non-verbal specialized strategies, e.g. look, eye development, and motions are utilized in numerous uses of human-PC connection, among them facial feeling is generally utilized as it conveys the enthusiastic states and sensations of people. In the machine learning calculation, a few significant separated highlights are utilized for displaying the face. As a result, it won't get a high accuracy rate for acknowledging that the highlights rely on prior knowledge. Convolutional Neural Network (CNN) has created this work for acknowledgment of facial feeling appearance. Looks assume an essential part in nonverbal correspondence which shows up because of the inner sensations of an individual that thinks about the countenances. This paper has utilized the calculation to distinguish features of a face such as eyes, nose, etc. This paper identified feelings from the mouth, and eyes. This paper will be proposed as a viable method for distinguishing outrage, hatred, disdain, dread, bliss, misery, and shock. These are the seven feelings from the front-facing facial picture of people. The final result gives us an accuracy of 63% on the CNN model and 85% on the ResNet Model.**

Keywords—Feature extraction; convolutional neural network; resnet; emotion recognition; emotion detection; facial recognition

I. Introduction

For ideal human-PC interfaces (HCI) would likely want that equipment to produce usable and safe systems [1]. This examination is definitely about how precisely PCs could distinguish feelings correctly from the distinct sensors. This try-out has been applied as a cosmetic picture as a new medium to follow human inclination [2]. Seven important feelings are most inclusive to men and women. Specifically, unbiased, mad, disdain, dread, happy, miserable, and impact and these vital feelings can always be perceived from a new human's look. This specific examination proposes a new compelling way for figuring out unbiased, cheerful, gloomy, and shocking these kinds of four feelings by front-facing cosmetic inclination. During many previous years, different techniques have been proposed for feeling acknowledgment. Numerous calculations were recommended to foster framework applications that can identify feelings well indeed [3]. PC applications could better convey by changing reactions as per the enthusiastic condition of human clients in different connections. The feeling of an individual is not set in stone by discourse or his face or even one's signal [4]. The work introduced in this paper investigates the acknowledgment of looks from the face. Concerning feeling suggestion, the customary procedures ordinarily consider another face picture that is perceived coming from your information picture, and facial broken expressions or accomplishments as a rule are seen from the specific face regions. Just after that novel space and normal qualities are disengaged out of their facial helpings. At last dependent on the secluded attributes a classifier, with respect to example, the Keras index, sporadic forest, is typically ready to supply affirmations results.

Each of our facial thoughts is expressed through account activation of specific systems of facial muscles. These are sporadic, subtly complex alarms in an expression that frequently contains a wealth of data [5]. Through facial sentiment recognition, we are able to assess the effects that content and services have on the audience/users. We used a machine learning model like CNN to achieve these aims [6].

The essential motive of this research is to discover facial feelings. Emotion recognition will be a significant step in identifying many thoughts on the face. Through face emotion recognition, all of us can gauge the effects content materials and services bestow on users. The opportunity that devices can interpret human beings' emotions is exactly what motivated us to generate this particular research.

Convolutional Neural Network is one type of Deep Learning algorithm which is used for working with images and videos. Images are used as inputs, and the system collects their features before categorizing those using previously learned features. The various filters available in the model help extract features at each layer such as edges, and shapes (vertical, horizontal, round), and combine them all to be used as identification for the image.

The CNN model works in mostly two stages:

Feature Extraction – The phase where filters and layers work for extracting information from the images to be passed on to the next layer for classification.

Classification – Classifying the images based on the target variable is carried out.

In Section II, we describe the literature review followed by the methodology and model specifications in Section III that have been implemented in this paper. Moving forward we have discussed the outcome of our model in Section IV and have concluded that in Section V.

II. Literature Review

1) Bartlett et al. carried out the initial work that outlines a method to complete the task [7]. The goal of the article was to develop a system that could automatically identify frontal faces in video streams and categorize the emotions those faces were expressing based on their facial expressions [8]. The article defined facial expressions as expressing happiness, sorrow, surprise, contempt, fear, rage, or neutral emotions.

2) With the help of the Radial Basis Function network (RBFN) for classification and Fisher's Linear Discriminant (FLD), Singular Value Decomposition (SVD), for feature selection, an algorithm for identifying a person's emotional state through facial expressions like anger, disgust, and happiness is demonstrated. It attempts to represent the face analytically in order for the feature vector can be input into a classifier. The system's overall success mostly hinges on the accurate recognition of the face or specific facial features like the eyes, brows, and lips [9].

3) Authors proposed to use of a revolutionary EEG-based emotion recognition method. This method uses multi-channel EEG data to examine the usage of three-Dimensional Convolutional Neural Networks (3D-CNN) for emotion identification. The proposed 3D-CNN approach's performance is improved by the data augmentation step. And using the multi-channel EEG signals as the data input for the proposed 3D-CNN model, a 3D data representation is created [10].

4) The VGG Net architecture was adopted by Yousif Khaireddin and Zhuofa Chen, who tuned its hyperparameters and experimented with other optimization techniques. To their experience, the model does not require additional training data to attain state-of-the-art single-network accuracy of 73.28% on FER2013 [11].

5) The initial step of visual processing, the encoding of facial features, and the decoding of facial expression connotation can all be considered as stages in the process of facial emotion identification, which was seen in schizophrenia patients [12].

6) In their study, M. Kalpana Chowdary, D. Jude Hemanth, and Tu N. Nguyen discuss the use of transfer learning techniques to recognise emotions. Pre-trained Resnet50, vgg19, Inception V3, and Mobile Net networks were employed in this study. After removing the completely connected layers from the pre-trained ConvNets, their own fully connected layers that are appropriate for the number of instructions in our task were added. The CK + database was used to conduct the experiment [13].

7) The technology of convolutional neural network model is used in an effective deep learning technique for face emotion detection to classify emotions from facial photographs and accurately identify age and gender from facial expressions [14].

8) A pre-trained DCNN model is adopted by replacing its dense higher layer(s) compatible with FER in one study's proposal for very Deep Convolution Neural Network (DCNN) modelling through Transfer Learning (TL) technique. The model is then tweaked using facial emotion data. A new approach is shown, which leads to a progressive increase in the model's accuracy by first tweaking each of the pre-trained DCNN blocks individually, after training the dense layer(s). Eight alternative pre-trained DCNN models (VGG-16, VGG-19, ResNet-18, ResNet-34, ResNet-50, ResNet-152, Inception-v3, and DenseNet-161) as well as the well-known KDEF and JAFFE facial image datasets are used to verify the proposed FER system [15].

9) A pre-trained DCNN model is adopted by replacing its dense higher layer(s) compatible with FER in one study's proposal for very Deep Convolution Neural Network (DCNN) modelling through Transfer Learning (TL) technique. The model is then tweaked using facial emotion data. A new approach is shown, which leads to a progressive increase in the model's accuracy by first tweaking each of the pre-trained DCNN blocks individually, after training the dense layer(s). Eight alternative pre-trained DCNN models (VGG-16, VGG-19, ResNet-18, ResNet-34, ResNet-50, ResNet-152, Inception-v3, and DenseNet-161) as well as the well-known KDEF and JAFFE facial image datasets are used to verify the proposed FER system [16].

*10)*The authors of the cited paper have suggested a two-tower CNN architecture to categorize images into seven fundamental types of emotion in light of the significance of edges in an image. In addition to the other tower, the planned CNN contains a tower called the edge-tower, which has a more straightforward architectural design and employs edge pictures as inputs [17].

III. Methodology and Model Specifications

This activity looks at the ongoing challenge facing machine learning and the whole program is part of the training. The location where the system should educate using real-time individual face replies to data. For example, if the system has to discover a furious face the very first system must learn about an angry face and if the program should see a happy face, then the first program should get acquainted with a happy face. To precede a program with these types of emotions, a retraining method was used. Retraining data is then collected from surveys and people. The most difficult part of the program was the iterating component out of various other components of the system. Machine learning works as an advanced tool to analyze data from a wide range of expertise and slow motion [18]. This gives you the ability to get more accurate emotions [19] and provides real-time feedback. The machine did not hold out for a future result, not to be preserved. With the help of modern personal computers, neoteric data exploration techniques can examine thousands of data quickly and save several hours [20].

Other than that, using and installing such programs is very costly. Properly covered, it will withstand a high deal of adverse situations. This work offended the common and possible framework of emotional data mining to verify emotional patterns with the use of machine learning. This kind of paper proposes software based on typically the in-depth reading style and emotional conception of computer eyesight. This proposed approach uses the CNN algorithm in this document. This has triggered a much even more advanced approach as compared to the one which observed only seven feelings about CNN [21].

CNN Implementation: Convolutional Neural Network is one type of Deep Learning algorithm which is used for working with images and videos. Images are used as inputs, and the system collects their features before categorizing those using previously learned features [22]. The various filters available in the model help extract features at each layer such as edges, and shapes (vertical, horizontal, round), and combine them all to be used as identification for the image [23].

The CNN model works in mostly two stages as seen in Fig. 1:

Feature Extraction – The phase where filters and layers work for extracting information from the images to be passed on to the next layer for classification [24].

Classification – Classifying the images based on the target variable is carried out.

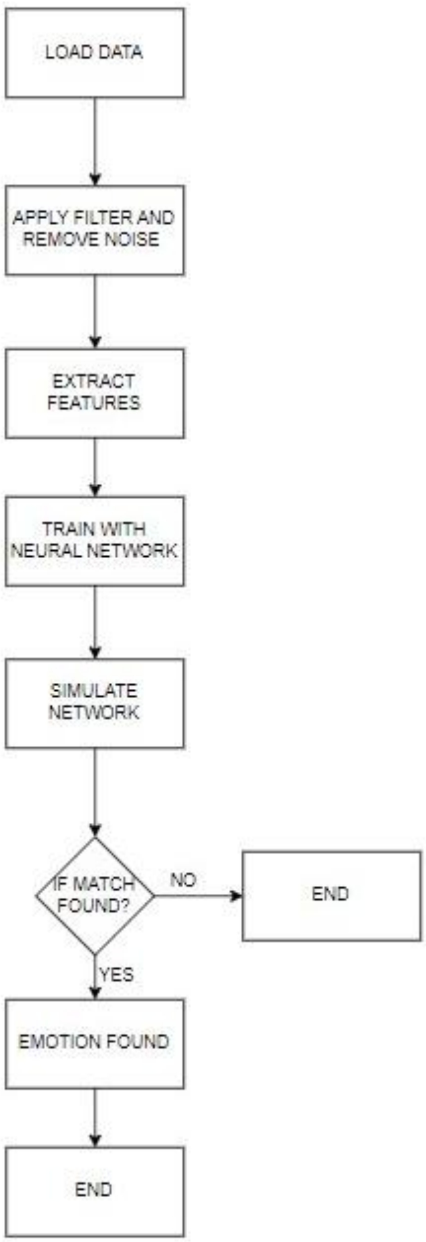
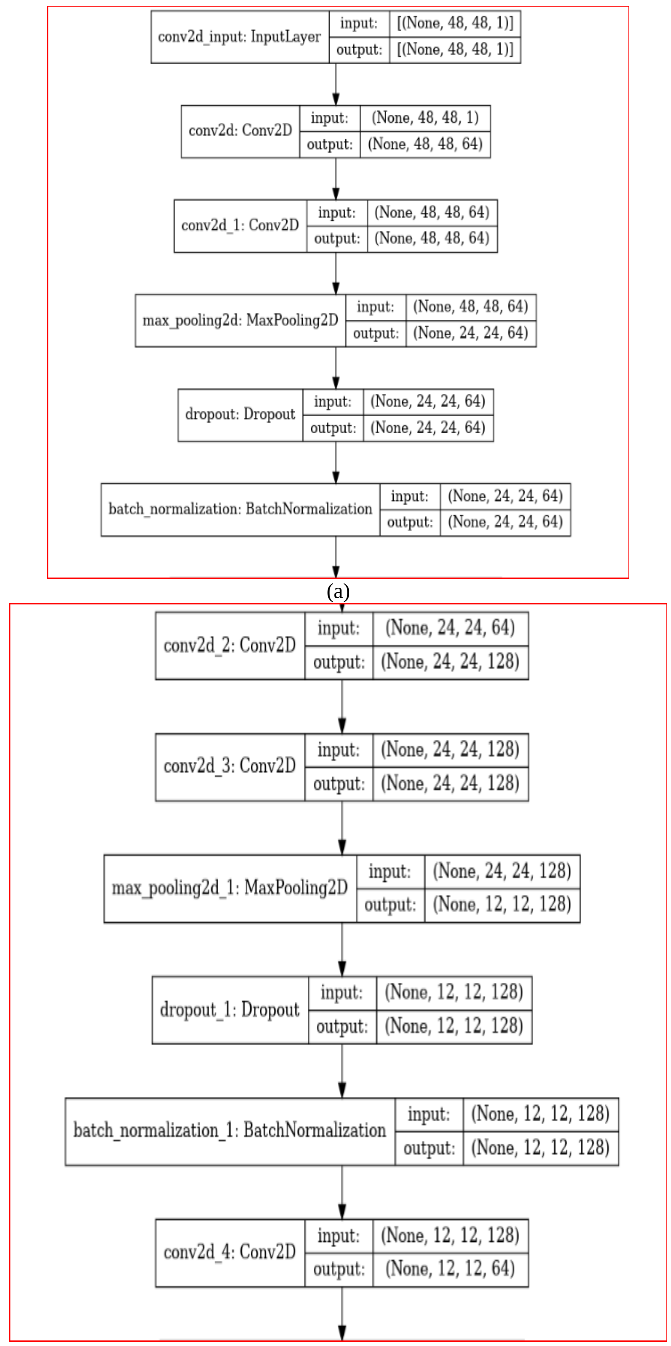


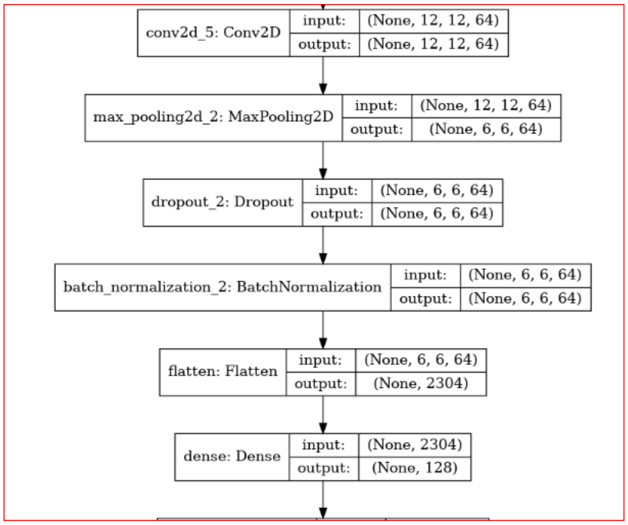
Fig. 1. Flowchart.

CNN Algorithm: The convolution neural network works by taking input images and assigning them weights and biases according to the importance of the objects identified. Here, we take the Facial Emotion Recognition 2013 dataset available publicly as our input data. Various pre-processing of input data is carried out as converting the images to grayscale, resizing them, reshaping the data, etc. The data is then divided into training and testing sets for model implementation. Various accuracy parameters are then taken into consideration for checking model performance and prediction.

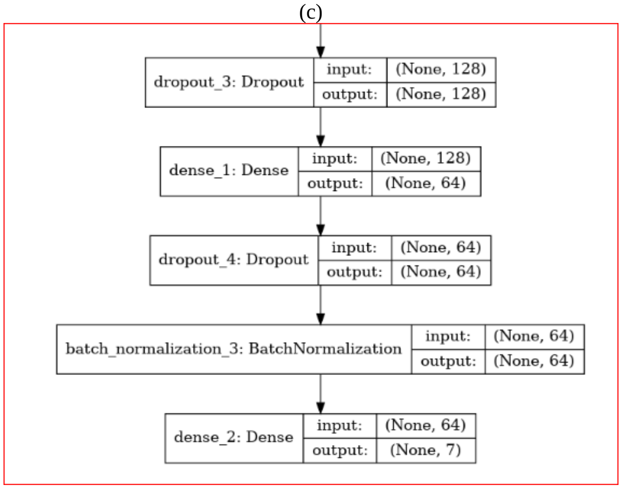
The architecture of the CNN Model Fig. 2:



(b)



(c)



(d)

Fig. 2. Model Architecture (a) - (d).

1) ResNet implementation: Residual Networks, or ResNet, learn residual functions with reference to the layer inputs, instead of learning unreferenced functions. The residual is the amount or number by which you have to change your prediction to meet the actual value. Rather than skipping every few stacked layers directly, to fit a desired underlying mapping, residual nets let these layers fit a residual mapping. Residual blocks are stacked on top of each other to form a network. The model ResNet-50 has fifty layers using these blocks.

To produce classification results, ResNet-50 first performs a convolution operation on the input, then it applies residual blocks, and finally, it applies a full connection operation. CNNs typically end with fully connected (FC) layers that summarize the characteristics of the preceding connected layers [25]. The Fully Connected layer can be viewed as feature weighting and uses earlier convolution and pooling as feature engineering, local amplification, and local feature extraction.

The structure of the Fully Connected layer is a way to quickly learn the non-linear combinations of advanced attributes generated by the convolutional layer. The layer learns a possible non-linear function. First, the image, which has been converted into a suitable format, is flattened into column vectors and fed back to the feed-forward neural network. Every training iteration uses flattened data. In this approach, the model is able to discriminate between some low-level aspects in the image and the image's main features, classifying those using methods like SoftMax [26]. The categorization outcomes for the seven expressions are output here.

2) ResNet Algorithm: The ResNets work on the principle of building deeper networks compared to other existing network architectures and simultaneously finding an optimized number of layers to negate the vanishing gradient problem. Here, we take the Facial Emotion Recognition 2013 dataset available publicly as our input data. Various pre-processing of input data is carried out as converting the images to grayscale, resizing them, reshaping the data, etc. The data is then divided into training and testing sets for model implementation. The model works on a 50-layer deep residual neural network that stacks residual blocks on top of each other. Various accuracy parameters are then taken into consideration for checking model performance and prediction performance.

IV. RESULTS

The very first challenge that we faced here was the smaller number of images in the dataset for emotions like disgust and fear. As a result of this, the model would be biased towards the emotions that have a larger number of images while training. Machine learning algorithms work well with datasets having a large number of segments and information. Image data generator helps deal with this by creating augmented images by making some modifications to the images in the form of flipping or rotating at various angles.

A. CNN Result

The model used by us here uses a convolution neural network (CNN) where we work by classifying images according to the emotion that it represents [27]. This is done with the help of feature extraction and various layers of the model such as max pool and dense [28]. The dropout layers help decrease the complexity of the models. This model is helpful in classifying the images with an accuracy level of approximately 62%. This however is not useful and very accurate so we try handling this problem with the help of transfer learning [29]. We can see the prediction outcome in Fig. 3.

In Fig. 4, we see how accurately our model performs by plotting the loss and accuracy metrics for the training and validation datasets. The loss that occurs, can be seen as decreasing as we increase the number of epochs the model iterates over. The accuracy for the model can be seen increasing as the model learns and the validation data accuracy can be seen as that of training i.e., 62%.

Fig. 3. Expression Predicted: Sad.

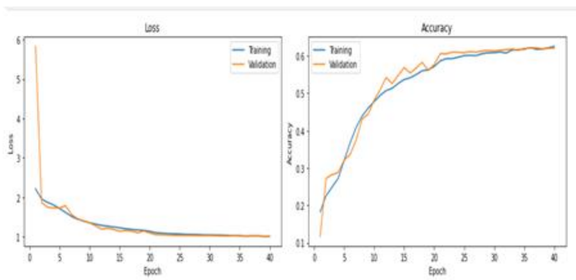


Fig. 4. Loss and Accuracy.

B. *ResNet Result*

Transfer learning involves the use of pre-trained models or information that is already stored and with the help of it we train new models. It helps in the reduction of time complexity as well as space complexity. We have chosen Resnet50 as the transfer learning model to solve our problem. Resnet50 makes use of its pre-existing information for feature extraction in determining emotions [30]. The model is then fine-tuned on different parameters and is able to reach an accuracy of 85% for both training and validation data. With an AUC score of approximately 70%, the model using Resnet50 seems to be more promising as compared to CNN. We can see the prediction outcome in Fig. 5.

In Fig. 6, we see how accurately our ResNet model performs by plotting the loss and accuracy metrics for the training and validation datasets. The loss that occurs, can be seen as decreasing as we increase the number of epochs the model iterates over. The losses for validation data can also be seen as decreasing. The accuracy for the model can be seen increasing as the model learns and the validation data accuracy can be seen as that to training after finetuning as well to reach around 85%.

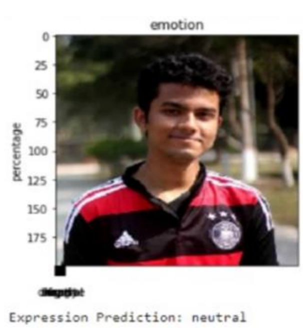


Fig. 5. Expression Predicted: Neutral.

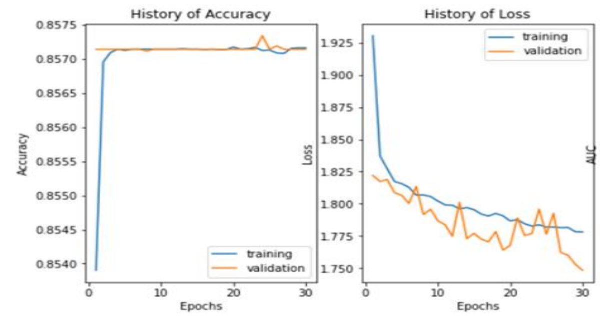


Fig. 6. History of Accuracy and History of Loss.

V. Conclusion

A seasoned human can often identify another human's thoughts by inspecting and searching his/her face. Yet, machines are becoming more intelligent day by day as we are moving towards more advancements. Machines with capacities for acting similar to humans are introduced. Machines are being trained to behave on behalf of human opinion. Alternatively, if machines can identify the sentiment, it can prevent lots of incidents too. Data exploration can facilitate appropriate expression patterns by allowing machines to find an action more like humans with additional proficiency and error-free computation sentiment.

We have discussed a method for expression recognition in static pictures in this paper. In order to develop this emotion analysis system, CNN and Haarcascade were used for feature selection, and CNN and ResNet models for classification. This study uses average values generated from training samples to identify emotional expressions on human faces. We tested the system's ability to recognize the photographs and correctly interpret the expressions from the images.

Acknowledgment

It gives us great pleasure in presenting the research report on "FACIAL EMOTION DETECTION USING CNN MODEL". I am grateful to Dr. Deepali Vora, Head of Department, SIT, Pune for her indispensable support, and suggestions. In the end our special thanks to Symbiosis Institute of Technology for providing various resources such as laboratories with all needed software platforms, and continuous Internet connection, for our research.

References

[1] Hashim Abdulsalam, Wisal & al-hamdani, dr & Al Salam, Mohammed. (2018). Facial Emotion Recognition: A Survey. 7. 771-779. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 7, Issue 11, November 2018, ISSN: 2278 – 1323.

[2] Liu Y, Li Y, Ma X, Song R (2017) Facial expression recognition with fusion features extracted from salient facial areas. Sensors 17(4):712.

[3] Anggraeni Pitaloka, Diah & Wulandari, Ajeng & Basaruddin, T. & Liliana, Dewi Yanti. (2017). Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition. Procedia Computer Science. 116. 523-529. 10.1016/j.procs.2017.10.038.

[4] R. Cowie et al., "Emotion recognition in human-computer interaction," in IEEE Signal Processing Magazine, vol. 18, no. 1, pp. 32-80, Jan 2001, doi: 10.1109/79.911197.

[5] Jonathan, Jonathan & Lim, Andreas & Paoline, & Kusuma Negara, I Gede Putra & Zahra, Amalia. (2018). Facial Emotion Recognition Using Computer Vision. Indonesian Association for Pattern Recognition. 46-50. 10.1109/INAPR.2018.8626999.

[6] Mehendale, N. Facial emotion recognition using convolutional neural networks (FERC). SN Appl. Sci. 2, 446 (2020). Springer.https://doi.org/10.1007/s42452-020-2234-1.

[7] M. S. Bartlett, G. Littlewort, I. Fasel and J. R. Movellan, "real time face detection and facial expression recognition: development and applications to human computer interaction," Computer Vision and Pattern Recognition Workshop, vol. 5, July 2003.

[8] Viola, Paul & Jones, Michael. (2001). Robust Real-time Object Detection. International Journal of Computer Vision. 57. 137-154.

[9] Arumugam, Devi & Srinivasan, Purushothaman. (2011). Emotion Classification Using Facial Expression. International Journal of Advanced Computer Science and Applications - IJACSA. 2. 10.14569/IJACSA.2011.020714.

[10] Shawky, Elham & El-Khoribi, Reda & Shoman, Mahmoud & Wahby Shalaby, Mohamed. (2018). EEG-Based Emotion Recognition using 3D Convolutional Neural Networks. International Journal of Advanced Computer Science and Applications. 9. 329.

[11] Yousif Khaireddin, Zhuofa Chen,Cornell University, arXiv:2105.03588

[12] Gao Z, Zhao W, Liu S, Liu Z, Yang C and Xu Y (2021) Facial Emotion Recognition in Schizophrenia. Front. Psychiatry 12:633717. doi: 10.3389/fpsyt.2021.633717.

[13] Chowdary, M.K., Nguyen, T.N. & Hemanth, D.J. Deep learning-based facial emotion recognition for human–computer interaction applications. Neural Comput & Applic (2021). https://doi.org/10.1007/s00521-021-06012-8.

[14] Khattak, A., Asghar, M.Z., Ali, M. et al. An efficient deep learning technique for facial emotion recognition. Multimed Tools Appl 81, 1649–1683 (2022). https://doi.org/10.1007/s11042-021-11298-w.

[15] Akhand, M.A.H.; Roy, S.; Siddique, N.; Kamal, M.A.S.; Shimamura, T. Facial Emotion Recognition Using Transfer Learning in the Deep CNN. Electronics 2021, 10, 1036. https://doi.org/10.3390/electronics10091036.

[16] Bodavarapu PNR, Srinivas PVVS (2021) Facial expression recognition for low resolution images using convolutional neural networks and denoising techniques. Indian Journal of Science and Technology 14(12): 971-983. https://doi.org/10.17485/IJST/v14i12.14.

[17] Arkaprabha Bhandari, Nikhil R. Pal, Can edges help convolution neural networks in emotion recognition?, Neurocomputing, Volume 433, 2021, Pages 162-168, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2020.12.092.

[18] Rafaqat Hussain Arain, Abdullah Maitlo, Sadaquat Ali Ruk, Hidayatullah Shaikh. Madan Lal, Kamlesh Kumar Study of Face Recognition Techniques: A Survey. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 6, 2018.

[19] Kak, Shakir & Mustafa, Firas & Valente, Pedro. (2018). A Review of Person Recognition Based on Face Model. 4. 157-168. 10.23918/eajse.v4i1sip157. Eurasian Journal of Science & Engineering ISSN 2414-5629 (Print), ISSN 2414-5602 (Online).

[20] X. Han and Q. Du, "Research on face recognition based on deep learning," 2018 Sixth International Conference on Digital Information, Networking, and Wireless Communications (DINWC), 2018, pp. 53-58, doi: 10.1109/DINWC.2018.8356995.

[21] Amal Azazi, Syaheerah Lebai Lutfi, Ibrahim Venkat, Fernando Fernández-Martínez, Towards a robust affect recognition: Automatic facial expression recognition in 3D faces, Expert Systems with Applications, Volume 42, Issue 6,2015, Pages 3056-3066, ISSN 0957-4174.

[22] Huang, H.-F & Tai, S.-C. (2012). Facial expression recognition using new feature extraction algorithm. Electronic Letters on Computer Vision and Image Analysis. 11. 41-54.

[23] Morshed, Golam & Ujir, Hamimah & Hipiny, Irwandi. (2021). Customer's spontaneous facial expression recognition. Indonesian Journal of Electrical Engineering and Computer Science. 22. 1436. 10.11591/ijeecs.v22.i3.pp1436-1445.

[24] Prof. S. Graceline Jasmine, Shivam Singh. Face Recognition System. International Journal of Engineering Research & Technology (IJERT) http://www.ijert.org ISSN: 2278-0181 IJERTV8IS050150 Vol. 8 Issue 05, May-2019.

[25] KH Teoh et al 2021 J. Phys.: Conf. Ser. 1755 012006. Face Recognition and Identification using Deep Learning Approach. 5th International Conference on Electronic Design (ICED) 2020 Journal of Physics: Conference Series 1755 (2021) 012006 IOP Publishing doi:10.1088/1742-6596/1755/1/012006.

[26] Nazeer, Shahrin & Omar, Normah & Khalid, Marzuki. (2007). Face Recognition System using Artificial Neural Networks Approach. Proceedings of ICSCN 2007: International Conference on Signal Processing Communications and Networking. 420 - 425. 10.1109/ICSCN.2007.350774.

[27] Sajid M, Ali N, Dar SH, Iqbal Ratyal N, Butt AR, Zafar B, Shafique T, Baig MJA, Riaz I, Baig S (2018) Data augmentation-assisted makeup-invariant face recognition. Math Probl Eng 2018:1–10.

[28] TANOY DEBNATH et al. Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity. Scientific Reports, [s. l.], v. 12, n. 1, p. 1–18, 2022. DOI 10.1038/s41598-022-111730.

[29] Zhang H, Jolfaei A, Alazab M. A face emotion recognition method using convolutional neural network and image edge computing. IEEE Access. 2019; 7: 159081-159089. 10.1109/ACCESS.2019.2949741.

[30] Deshmukh, S.; Patwardhan, M.; Mahajan, A. Survey on real-time facial expression recognition techniques. IET Biom. 2016,5, 155–163. Sensors (Basel). 2019 Apr; 19(8): 1863.Published online 2019 Apr 18. doi: 10.3390/s19081863. Multidisciplinary Digital Publishing Institute (MDPI).