Temple University

Face Expression Recognizer

Course Code: CIS 5603

Submitted By: Smriti Roy

Email: smriti.roy@temple.edu

Abstract:

Emotion is one of the very few words in the English language that do not have a concrete definition and it is understandable. It is abstract. Yet almost every decision we have ever made in our lives is driven by emotion. Marketing research has proven that predicting sentiments correctly can be a huge source of growth for businesses and that's what we will be working on today — Reading Emotions. The facial expressions for Happiness, Sadness, anger, Surprise, fear, and disgust are the same across cultures.

Automatic emotion recognition based on facial expression is an interesting research field, which has presented and applied in several areas such as safety, health and in human machine interfaces. Researchers in this field are interested in developing techniques to interpret, code facial expressions, and extract these features to have a better prediction by computer. With the remarkable success of deep learning, the different types of architectures of this technique are exploited to achieve a better performance. The purpose of this paper is to make a study on automatic facial emotion recognition FER via deep learning.

Introduction:

Human facial expressions communicate a lot about their behavior and their interest in the products and services. Emotion is a powerful feeling that connects one human to other. Probably that's a reason why humans like to express their feelings through emotions. Human emotions are mainly listed into 6 types-Anger, Joy, Love, Sadness, Fear, and Surprise. In this technology era, everything that is unimaginable is possible. That goes with leveraging human emotions with the use of technology for learning about the industry end-customers responses, demands, and needs. Since a lot of businesses and industries come across threat barriers, safety and protection remain a matter of worry which now can be recognized and identified with the potential of technologies. Automatic emotion recognition powered by Machine learning that takes up the human facial expressions for dealing with multiple things is getting rushed in multiple industries such as safety, health, and human-machine interface.

In this project we are going to classify different expressions of faces and implementing them on real time streaming data. This Can be utilized in various industries like Gaming, Market Research, Analyze the emotions of job candidates in online interview.

Related Work:

Here is some related literature on this topic.

Emotion Recognition Based on Facial Expressions: The process of human communication is inextricably linked to the fluctuation of various emotions. When people are experiencing basic emotions, their faces will display a variety of expression patterns, each with its own set of characteristics and distribution scale. Facial expression recognition is a crucial part of human-computer interaction that allows computers to understand facial expressions based on human thinking. According to the processing of facial expression recognition process can be divided into three important face detection, feature extraction and classification module, face detection as the key technology of face **recognition (Adjabi et al., 2020; Zhang et al., 2021)** with its rapid development has basic mature, which can effectively extracted from the original face image of excellent characteristics and the characteristics of correct classification becomes key factor affecting the recognition result. For example, **Gao and Ma (2020)** obtained facial expression attributes from facial images to predict emotional states according to facial expression changes.

Emotion Recognition Based on Physiological Signals: The basis of emotion recognition based on physiological signals is that humans will produce different responses under different stimuli. For example, physiological signals such as brain electricity, electrocardiogram, pulse, and skin electrical response can all reflect emotions. **Momennezhad (2018)** used EEG signals for emotion recognition, extracting features from the time domain and frequency domain of EEG signals. Although the changes of physiological signals are not controlled by humans, they can most objectively reflect human emotional conditions.

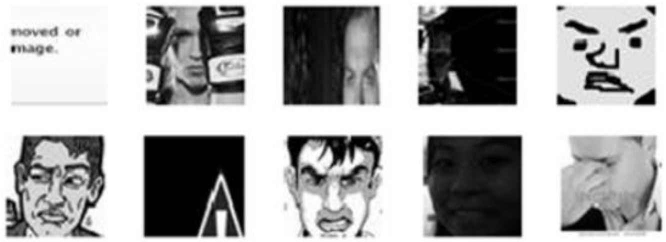
Emotion Recognition Based on Gestures: People will involuntarily undergo some posture changes in different environmental states and moods, and judge human emotions based on physical information such as the time and frequency of these posture changes, according to gesture-based emotion recognition. **Ajili et al. (2019)** used human movement analysis to identify motion, then evaluated and evaluated the emotions expressed by human motion posture. However, the single use of human gestures for emotion recognition has certain limitations because many gestures do not have emotional significance or the same gestures have different emotional meanings in different background environments, so human gestures are usually different from others. The modalities (such as expressions, speech, etc.) are combined for emotion recognition. Expressions are the most intuitive way to convey emotions among several ways to express human emotion information, such as facial expressions, voices, physiological signals, and gestures, and expression information is relatively easy to obtain in most environments, so I use them. Human emotional states are studied using facial expressions as objects.

Methodology:

The data set used for training is the Kaggle FER2013 emotion. The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is centered and occupies about the same amount of space in each image. The task is to categorize each face based on the emotion shown in the facial expression in to one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

Dataset Challenges

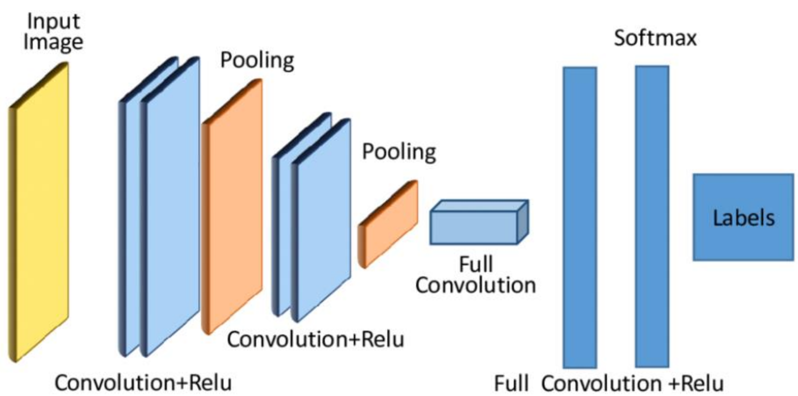
- **Imbalance**: Imbalance is when one class has many more images than another class. This results in the model being biased towards one class. For example, if there are 2000 images for the happy expression and 500 images for the fear expression, then the model will be biased towards the happy expression. Data augmentation is done to avoid this problem. Data augmentation increases the amount of data using techniques like cropping, padding, and horizontal flipping.
- **Intra-class variation**: Some images in the dataset are not human faces as there are drawings and animated faces. Model performance will be better if all images in the dataset are human faces so other images should be removed.
- **Occlusion**: Occlusion is when part of the image is covered. This can occur when a hand covers a part of the face such as the right eye or nose. A person wearing sunglasses, or a mask also creates occlusion. Eyes and noses have primary features which are important to extract and recognize emotions. Thus, occluded images should be removed from the dataset as the model cannot recognize emotions from these images
- Fer2013 is a challenging dataset. The images are not aligned and some of them are incorrectly labeled as we can see from the following images. Moreover, some samples do not contain faces.



Models:

In this paper I have tried to build the model using deep learning. I started with a baseline CNN model and gradually improved the model with different parameters also used transfer learning to achieve best results.

Proposed 4 models based on CNN

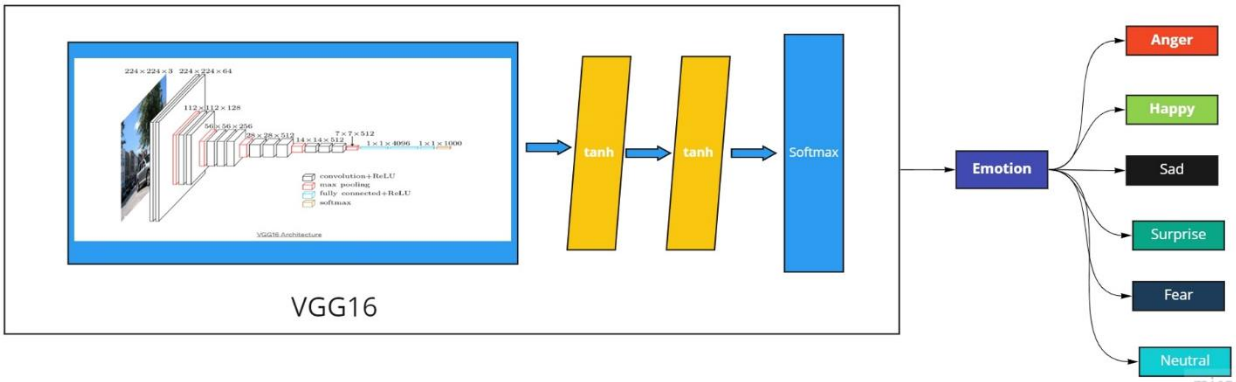


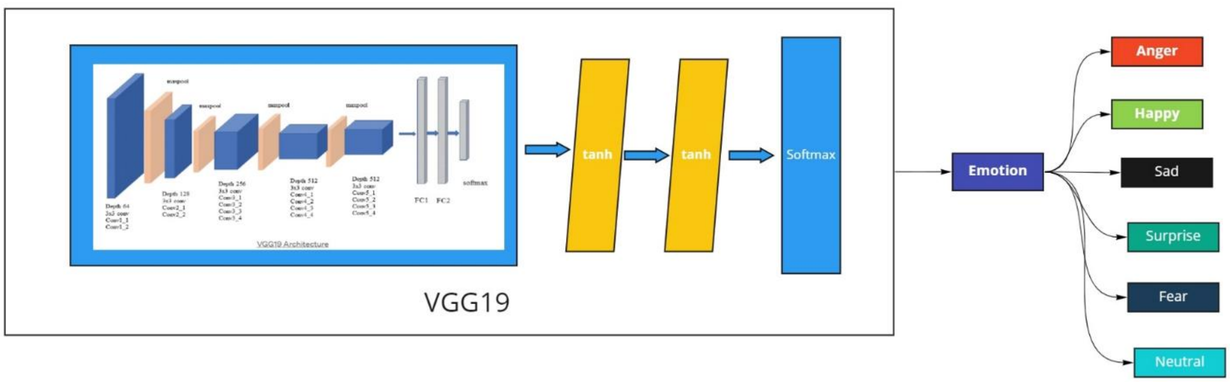
A generic CNN Architecture.

Proposed Models	Model Archietecture
Model 1	4-layer model using relu and softmax
Model 2	6-layer CNN model of tanh, MaxPooling2D, and softmax
Model 3	6-layer CNN model of tanh, MaxPooling2D, and softmax, normalized data, batch-size = 10 early callback
Model 4	Added one more tanh layer to last model 3

Below Two models have been Implemented using Transfer learning for getting best results:

Proposed Models	Model Archietecture
Model 5	VGG16 with 2 tanh layers and softmax output layer
Model 6	VGG19 with 2 tanh layers and softmax output layer





Results:

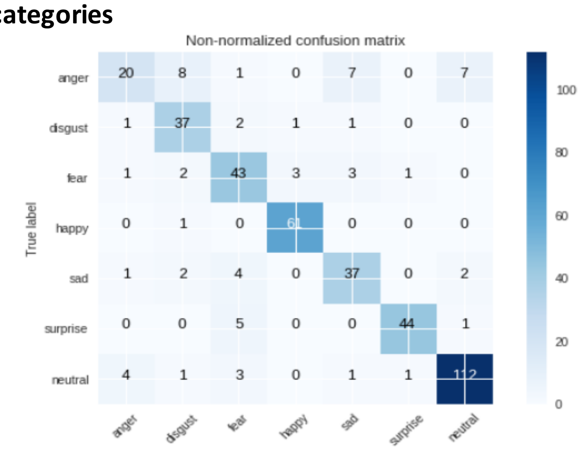
Model	epochs	validation accuracy
1	15	0.09

We will not test or visualize these results because this is our base case (the accuracy is bad)

Model	epochs	f1	precision	recall
2	100	0.448027843	0.582199014	0.437799

f1, recall, and precision for each of the categories

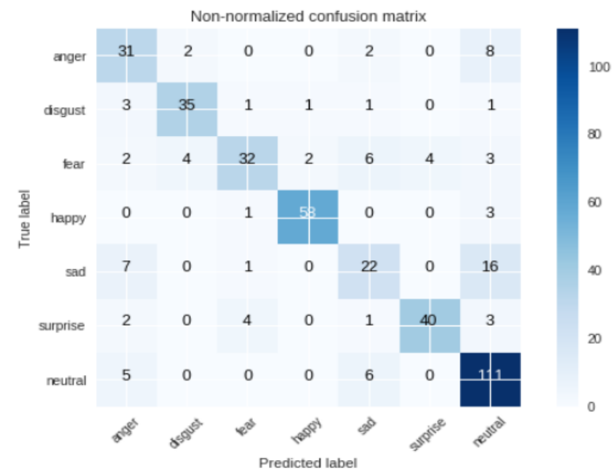
	precision	recall	f1-score	support
0	0.74	0.47	0.57	43
1	0.73	0.88	0.80	42
2	0.74	0.81	0.77	53
3	0.94	0.98	0.96	62
4	0.76	0.80	0.78	46
5	0.96	0.88	0.92	50
6	0.92	0.92	0.92	122
micro avg	0.85	0.85	0.85	418
macro avg	0.83	0.82	0.82	418
weighted avg	0.85	0.85	0.84	418
samples avg	0.85	0.85	0.85	418



Model	epochs	Batch_size	f1	precision	recall
3	35	10	0.783531328	0.790505	0.78708134

F1, recall, and precision for each of the categories

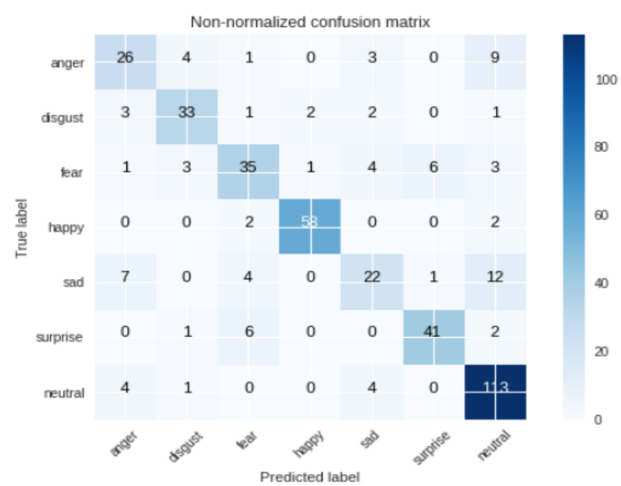
	precision	recall	f1-score	support
0	0.62	0.72	0.67	43
1	0.85	0.83	0.84	42
2	0.82	0.60	0.70	53
3	0.95	0.94	0.94	62
4	0.58	0.48	0.52	46
5	0.91	0.80	0.85	50
6	0.77	0.91	0.83	122
micro avg	0.79	0.79	0.79	418
macro avg	0.79	0.75	0.77	418
weighted avg	0.79	0.79	0.78	418
samples avg	0.79	0.79	0.79	418



Model	epochs	Batch_size	f1	precision	recall
4	35	10	0.779251015	0.779386	0.784689

F1, recall, and precision for each of the categories

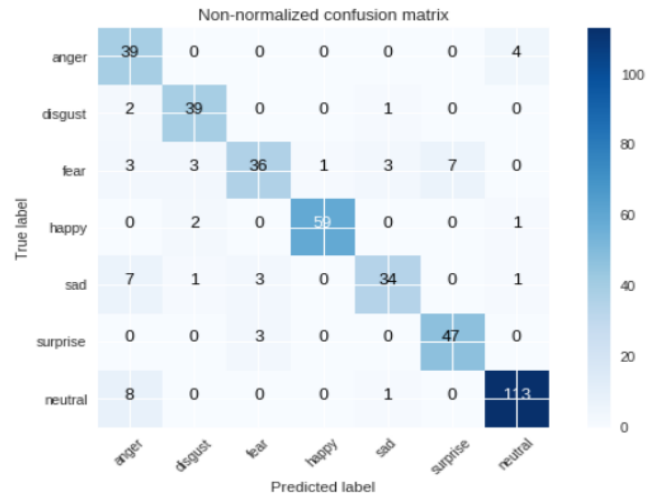
	precision	recall	f1-score	support
0	0.63	0.60	0.62	43
1	0.79	0.79	0.79	42
2	0.71	0.66	0.69	53
3	0.95	0.94	0.94	62
4	0.63	0.48	0.54	46
5	0.85	0.82	0.84	50
6	0.80	0.93	0.86	122
micro avg	0.78	0.78	0.78	418
macro avg	0.77	0.74	0.75	418
weighted avg	0.78	0.78	0.78	418
samples avg	0.78	0.78	0.78	418



Transfer Learning (VGG16)	Model	epochs	Batch_size	f1	precision	recall
	5	15	10	0.878162934	0.886815	0.87799043

f1, recall, and precision for each of the categories

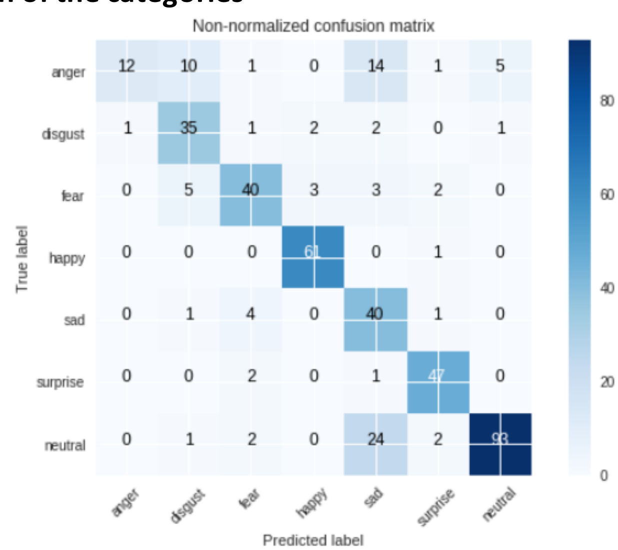
	precision	recall	f1-score	support
0	0.66	0.91	0.76	43
1	0.87	0.93	0.90	42
2	0.86	0.68	0.76	53
3	0.98	0.95	0.97	62
4	0.87	0.74	0.80	46
5	0.87	0.94	0.90	50
6	0.95	0.93	0.94	122
micro avg	0.88	0.88	0.88	418
macro avg	0.87	0.87	0.86	418
weighted avg	0.89	0.88	0.88	418
samples avg	0.88	0.88	0.88	418



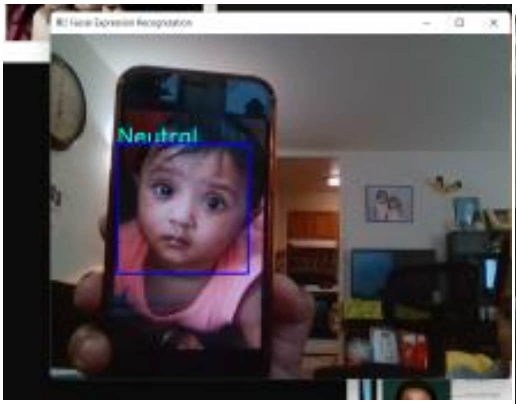
Transfer Learning (VGG19)	Model	epochs	Batch_size	f1	precision	recall
	6	15	10	0.780245691	0.831804	0.784689

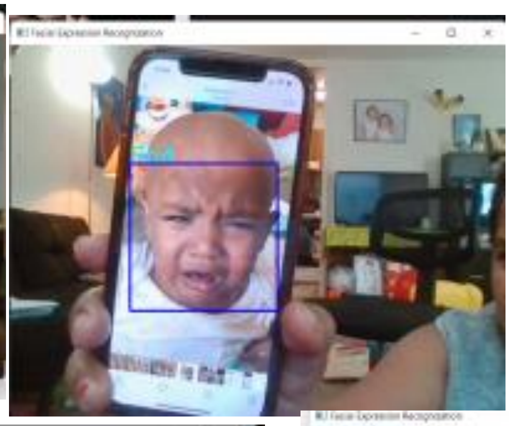
f1, recall, and precision for each of the categories

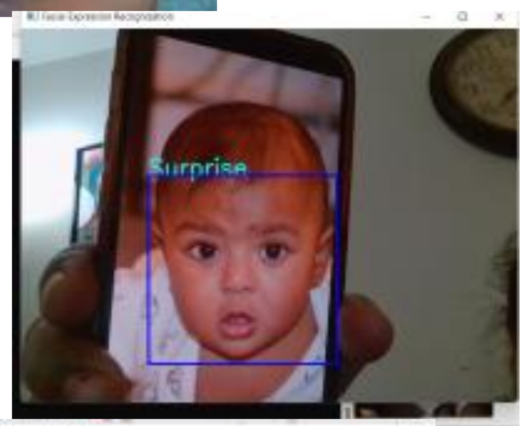
	precision	recall	f1-score	support
0	0.92	0.28	0.43	43
1	0.67	0.83	0.74	42
2	0.80	0.75	0.78	53
3	0.92	0.98	0.95	62
4	0.48	0.87	0.62	46
5	0.87	0.94	0.90	50
6	0.94	0.76	0.84	122
micro avg	0.78	0.78	0.78	418
macro avg	0.80	0.77	0.75	418
weighted avg	0.83	0.78	0.78	418
samples avg	0.78	0.78	0.78	418

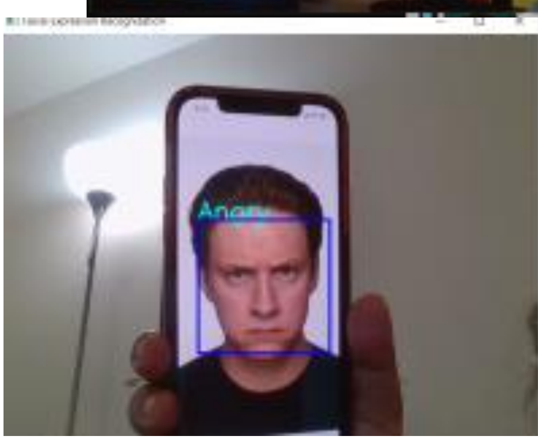


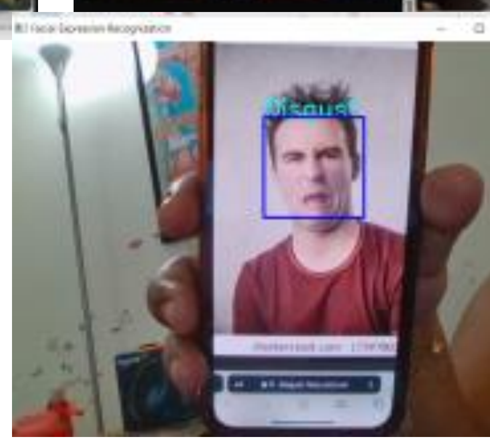
Test Results from Web Feed:











Discussion:

Model 5 (VGG16 with 2 tanh layers and softmax output layer) with F1-score = 0.88 gave best results among all the models. Our model is very good for predicting happy and neutral faces. However, it predicts quite poorly sad faces because it confuses them with fear and surprised faces. With more research and more resources this model could certainly be improved

	Model	epochs	validation accuracy			
	1	15	0.09			
	Model	epochs	f1	precision	recall	
	2	100	0.448027843	0.582199014	0.437799	
	Model	epochs	Batch_size	f1	precision	recall
	3	35	10	0.783531328	0.790505	0.78708134
	4	35	10	0.779251015	0.779386	0.784689
Transfer Learning (VGG16)	5	15	10	0.878162934	0.886815	0.87799043
Transfer Learning (VGG19)	6	15	10	0.780245691	0.831804	0.784689

Adding another dense layer or changing some parameter with more epochs we can achieve more accuracy in our model. Facial emotion recognition is an emerging field so considering other Neural Networks such may improve the accuracy. The feature extraction is like pattern recognition which is used in intelligence, military, and forensics for identification purposes. Thus, techniques such as the Capsnet algorithm for pattern recognition can be considered. Deep Learning based approaches require a large, labeled dataset, significant memory and long training and testing times which makes them difficult to implement on mobile and other platforms with limited resources. Thus, simple solutions should be developed with lower data and memory requirements

Acknowledgement:

At the outset, I would like to express my sincerest gratitude to Professor Dr. Pei Wang, a respected, responsible, and resourceful scholar, who has provided me with valuable guidance throughout the writing process. I would not have been able to complete my thesis without his enlightening instructions, impressive kindness, and patience. I am enlightened by his keen and vigorous academic observation not just in this thesis, but also in my future studies. Finally, I would like to thank my family for their encouragement and support.

References:

[1] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," in *IEEE Transactions on Affective Computing*, doi: 10.1109/TAFFC.2020.2981446.

[2] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic and K. Scherer, "The first facial expression recognition and analysis challenge," *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2011, pp. 921-926, doi: 10.1109/FG.2011.5771374.

[3] Caifeng Shan, Shaogang Gong, Peter W. McOwan,Facial expression recognition based on Local Binary Patterns: A comprehensive study,Image and Vision Computing,Volume 27, Issue 6,2009,Pages 803-816

[4] S. Berretti, A. Del Bimbo, P. Pala, B. B. Amor, and M. Daoudi. A set of selected sift features for 3d facial expression recognition. In Pattern Recognition (ICPR), 2010 20th International Conference on, pages 4125–4128. IEEE, 2010

[5] H. Ding, S. K. Zhou, and R. Chellappa. Facenet2expnet: Regularizing a deep face recognition net for expression recognition. In Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on, pages 118–126. IEEE, 2017

[6] Z. Meng, P. Liu, J. Cai, S. Han, and Y. Tong. Identity-aware convolutional neural network for facial expression recognition. In Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on, pages 558– 565. IEEE, 2017

[7] H. Jung, S. Lee, J. Yim, S. Park, and J. Kim. Joint fine-tuning in deep neural networks for facial expression recognition. In Proceedings of the IEEE International Conference on Computer Vision, pages 2983–2991, 2015

[8]M. Liu, S. Shan, R. Wang, and X. Chen. Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1749– 1756, 2014.

[9] K. Kaulard, D.W. Cunningham, H.H. Bulthoff, C. Wallraven, The MPI facial expression database: A validated database of emotional and conversational facial expressions, PLoS One, vol. 7, no. 3, art. e32321, (2012).

[10] G.E. Hinton et al., Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups, IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82-97, (2012).

[11] A. Pentland, Social signal processing, IEEE Signal Processing Magazine, vol. 24, no. 4, pp. 108-111, (2007).

[12] M. Xie, Development of artificial intelligence and effects on financial system, Journal of Physics: Conference Series 1187, art. 032084, (2019).

[13] A. Nandi, F. Xhafa, L. Subirats, S. Fort, Real time emotion classification using electroencephalogram data stream in e-learning contexts, Sensors, vol. 21, no. 5, art. 1589, (2021).

[14] A. Raheel, M. Majid, S.M. Anwar, M. Alnowami, Physiological sensors based emotion recognition while experiencing tactile enhanced multimedia, Sensors, vol. 20, no. 14, art. 04037, (2020).

[15] D. Keltiner, P. Ekrman M. Lewis, J.H. Jones, Handbook of Emotions (2nd ed.), Guilford Publications (New York), pp. 236-249, (2000).

[16] B.C. Ko, A Brief review of facial emotion recognition based on visual information, Sensors, vol. 18, no. 2, art. 401, (2018).

[17] B. Zafar, R. Ashraf, N. Ali, M. Iqbal, M. Sajid, S. Dar, N. Ratyal, A novel discriminating and relative global spatial image representation with applications in CBIR, Applied Sciences, vol. 8, no. 11, art. 2242, (2018).

[18] M. Liu, S. Li, S. Shan, R. Wang, and X. Chen. Deeply learning deformable facial action parts model for dynamic expression analysis. In Asian Conference on Computer Vision, pages 143–157. Springer, 2014. [19] M. Liu, S. Shan, R. Wang, and X. Chen. Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition. In Proceedings of the IEEE Conference on Computer

[20] D. Y. Liliana, C. Basaruddin, and M. R. Widyanto, "Mix Emotion Recognition from Facial Expression using SVM-CRF Sequence Classifier," 2017, pp. 27–31.

[21] V. Sudha, et.al., "A fast and robust emotion recognition system for real-world mobile phone data," in 2015 IEEE International Conference on Multimedia Expo Workshops (ICMEW), 2015, pp. 1–6.

[22] D. Y. Liliana and C. Basaruddin, "A review on conditional random fields as a sequential classifier in machine learning," in 2017 International Conference on Electrical Engineering and Computer Science (ICECOS), 2017, pp. 143–148.

[23] Z. Ming, A. Bugeau, J.-L. Rouas, and T. Shochi, "Facial Action Units intensity estimation by the fusion of features with multi-kernel Support Vector Machine," in Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on, 2015, vol. 6, pp. 1–6.

[17] J. Chen, D. Chen, and L. Wang, "Facial Expression Recognition using Geometric and Appearance Features," Icimcs, pp. 29–33, 2012.

[24] D. A. Pitaloka, A. Wulandari, T. Basaruddin, and D. Y. Liliana, "Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition," Procedia Comput. Sci., vol. 116, pp. 523–529, 2017.

[25] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," ISCAS 2010 - 2010 IEEE Int. Symp. Circuits Syst. Nano-Bio Circuit Fabr. Syst., pp. 253–256, 2010. [20] J. Zraqou, W. Alkhadour, and A. Al-Nu'aimi, "An efficient approach for recognizing and tracking spontaneous facial expressions," 2013 Second Int. Conf. E-Learn. E-Technol. Educ. ICEEE, pp. 304–307, 2013.