Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Facial emotion recognition using convolutional neural networks

Ketan Sarvakar [a], R. Senkamalavalli [b], S. Raghavendra [c], J. Santosh Kumar [d], R. Manjunath [e], Sushma Jaiswal [f]

[a] Information Technology Department, U.V. Patel College of Engineering, Ganpat University, Mehsana, Gujarat 384012, India
[b] CSE Department, East Point College of Engineering, Bangalore, India
[c] Department of Computer Science and Engineering, School of Engineering and Technology, Christ Deemed to be University, Bengaluru 560074, India
[d] Department of Computer Science and Engineering, K S School of Engineering and Management, Bengaluru 56019, India
[e] Department of Computer Science and Engineering, R R Institute of Technology, Bengaluru 560090, India
[f] Department of Computer Science & Information Technology (CSIT), Guru Ghasidas Vishwavidyalaya, (A Central University)Koni, Bilaspur, (C.G.) 495009, India

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Emotion recognition
Facial expression
CNN

ABSTRACT

Emotional expressivity has always been a simple job for people, but computer programming is much harder to accomplish. Image emotions may be recognised by recent developments in computer vision and machine learning. In this article, we present a new method to detect face emotion. Use neural networks convolutionary (FERC). The FERC is based on a CNN network of two parts: the first portion removed the backdrop of the image, the second part removed the face vector. The expressional vector (EV) is utilised in the FERC model to detect the fve different kinds of regular facial expressions. The double-level CNN is continuous and the weights and exponent values of the final perception layer vary with each iteration. In that it improves accuracy, FERC varies from widely utilised CNN single-level technology. Moreover, EV generation prevents the development of a number of issues before a new background removal process is used (for example distance from the camera).

Selection and peer-review under responsibility of the scientific committee of the International Conference on Nanoelectronics, Nanophotonics, Nanomaterials, Nanobioscience & Nanotechnology.

1. Introduction

In understanding and recognising emotion, facial expressions play a crucial role. Even the word "interface" indicates how crucial the face of communication is Between two entities. between two entities. Studies have shown that reading Facial expressions may change the interpretation substantially what is being said and manage the flow of a discussion. The capacity of people to discern emotions is extremely essential for successful communication; for up to 93 percent of typical communication Talk relies on an entity's feeling. Ideal for children. Interfaces between human and computer (HCI) would want Man's emotion can be read by technology. This study is for this purpose how computers can correctly identify emotion from Its different sensors. This experiment was used as an experiment As a means to interpret human emotion, facial picture. The Human emotion studies may be traced back to Darwin's pioneering and has drawn a lot since then Researchers in this field. Seven fundamental emotions are universal To people. To people. Neutral, furious, disgusting, fearful, Happy, sad and surprising, and these fundamental emotions may be Recognized from the face of a human being. This study offers an efficient method to identify neutral, happy, sad and During the last decades, several techniques for emotional identification have been suggested. Many methods were proposed for the development of systems that can extremely effectively identify emotions. Computer applications may communicate better by altering answers in different encounters depending on human users' emotional state. A person's mood may be determined by words, expression or even gesture. The article examines the identification of expressions from the face. For the identification of facial emotions, conventional methods typically regard a face image that is separated from an information image and facial segments or milestones are identified in the facial districts. After that various spatial and worldly features are isolated from these face parts. Lastly, a classifier, for example, is trained at Keras library, the random forest, in order to provide recognition results.

This work is an applicable and profound model of learning. Deep learning is a well-established paradigm in the field of pattern

https://doi.org/10.1016/j.matpr.2021.07.297
2214-7853/
Selection and peer-review under responsibility of the scientific committee of the International Conference on Nanoelectronics, Nanophotonics, Nanomaterials, Nanobioscience & Nanotechnology.

Please cite this article as: K. Sarvakar, R. Senkamalavalli, S. Raghavendra et al., Facial emotion recognition using convolutional neural networks, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2021.07.297

recognition. It utilises an algorithm of the Convolutionary Neural Network (CNN) with the Keras library. CNN is a special kind of neural artificial network using a machine-learning device. CNN relates to detection of objects, facial recognition, picture processing, etc. Deep neural network (DCNN) composed of several layers of the neural network. This may also extract important characteristics from the data. Surprise these four frontal feeling emotions.

2. Literature review

The universal sign that all human beings communicate emotion is facial expression. Many efforts to create automated facial expression research methods have been made [3]. A number of applications include robotics, medical, driving systems and detectors of lies [4-6]. Each human person, independently of culture, has been developing seven fundamental emotions since Ekman et al., [7] defined seven basic emotions in the 20th century (fearful, frightened, joyful, Sad, disdain [8], disgust and astonishment). In the Facial Recognition Technology Research (FERET) dataset, Sajid et al. found the consequences of facial asymmetry as an age estimate marker [9]. Asymmetry on the right face is preferred than asymmetry on the left side, according to their results. In facial detection the look of the face remains an important problem. Ratyal et al., have offered a solution to the diversity of the face. They used a three-dimensional method invariant. Subject-specific descriptors are used [10,11]. Many problems are dealt with via coevolutionary networks, such as excessive makeup [12] and expressiveness [13].

3. Experimental setup

This part provides information for training and testing, preprocessing of the data, the various models and the evaluation of each kind.

3.1. Dataset

Neural networks tend to perform better as more training is given, and especially deep neural networks. In that regard, the most popular Expanded Cohn Canada (CK +) [9] and SFEW [2] were omitted. Instead, the data set was used for Facial expression (FER-2013). At the ICML Representation Learning Challenges [1] the 2013 FER-2013 data set was presented. There are 35,887 basic pictures: angry, outrageous, frightening, cheerful, sorrowful, indifferent and amazed. Fig. 1 shows the distribution of each expression. Each image displays a front perspective of one of seven sentences of a crazy topic. A sample of these words is shown in Fig. 2. The disgusted amount of individuals is important to remember. Expressions are much lower than the other terms. The predilection for pleasure was equally apparent. Expressions are employed because of the expressive bulk of the data. Researchers have made amazing accomplishments in face emotion recognition. Cognitive improvement of neuroscience Through research, science is progressed. Facial expression. The vision of growth of computers and machines. The accurate and generally available emotional recognition. Facial expression is thus categorised as an area of image processing. Some applications are meaningful. Interaction between people and computers, psychological observations, the identification of poisoned drivers etc. The need for a lie detector is essential.

3.2. Pre-processing

The dataset was created by converting numerous picture strings with a space separated number of 2304 into a 48*48 matrix. Each number represented a pixel value. The raw data of 35,887 pictures was divided into a training package of 28,709 photos and a test set of 7,178 photos - split at 80:20. In general, data is the most important element in deep learning. The greater the collection size, the better the performance. Improved performance. When there is less training data, the final outputs vary considerably more owing to a limited set. Carry on training. Carry on training. Carry on training. Carry on training. In this regard, a test set of 20% Total images may be deemed excessive. However, it is critical to prevent overfitting a large test set. Yes, yes, yes, yes, yes, yes, yes, yes, yes, yes It should also be noted that 60–20-20 splits are common for training, testing, and validation. Mollahosseini et al., [5] divided their 275 k image into 60% training data, 20% test data, and 20% training validation data. The validation set was expected in this instance. Please retrain the whole model whenever hyperparameters are updated. It took longer, and the computer power offered more instruction. For labels, a one-shot encoding technique was employed instead. Instead of categorising 0–6 numerals, consider categorising feelings. Using Haar Cascades [11], a face was recognised during live testing. This recognised face was then transformed into a picture. The greyscale image is reduced to 48*48 pixels. This converted the picture to the same format as Train.

3.3. Choosing a model

Three neural networks and a decision tree were used to choose the optimal foundation for building on. Keras [12] made use of neural networks. With a Python-based TensorFlow backend. In December, Sci-Kit Learn [13] was used for tree implementation. All Python 3.6 implementations use software to be created and reproduced for free. a contrast of Table 1 displays the final accuracy of all models. Because four distinct models are utilised, the training process for each model is included in the model description. Descriptions of descriptions This method comes with a test algorithm.

3.3.1. Decision tree

A decision-tab was selected initially because it requires little or no effort in data preparation and has a reputation for good performance. In nearly all circumstances. The first step was to instal the Sci-Kit Learn decision tree. Typical Decision Tree The classifier was employed, and its parameters were changed. However, the only metric that revealed a significant difference was The split emphmin sample parameter was tuned. With the setting set to 40, the total accuracy was just 0.309. When the difference between state-of-the-art models was higher than 0.6, a neural network was selected.

3.3.2. Feedforward neural network

The following experiment aimed to look into a feed-forward neural network in its most basic form. Three layers have been used: an input layer, a single layer, and an output layer. Input from the layer. Each was a densely connected (fully linked) layer. To avoid overfitting, a rate reduction of 0.2 was added for the input and the hidden layer. The Softmax output layer function is used, while the remaining layers rely on URLs (Rectified Linear Units). However, the model came to an end. The identical input expression - furious - is anticipated. Further Tweaking the hyperparameter produces no difference, thus it was decided that a convolutional neural network would be more effective.

3.3.3. Simple convolutional network

Following that, a rudimentary network was attempted. Simple convolution model had two convolutional layers of two sizes, one max pooling layer of two sizes, and the two layers (Dense). Before connecting the whole Layers entry, the output was flat-

Fig. 1. FER-2013 Expressions.

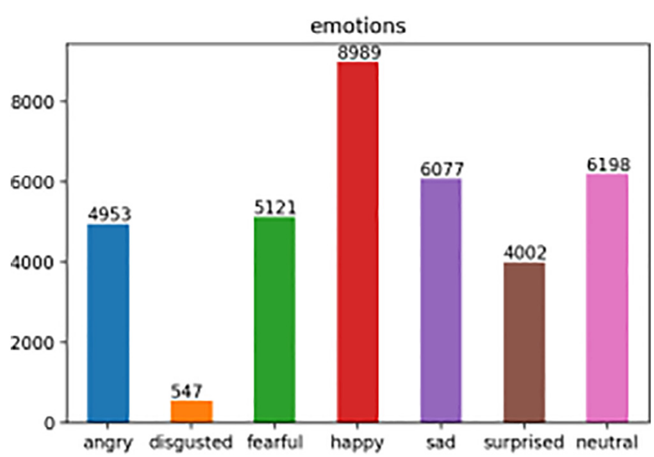

Fig. 2. FER-2013 Expression Distribution.

Table 1
Comparison of models.

Model	Accuracy
Decision Tree	31%
Simple CNN	18%
Proposed CNN	54%
Feed Forward NN	18%

tened. Layers are linked together. Dropout was used to solve the same problem as the feedforward neural network, but it was not unpleasant for all inputs. This is significant since one-fourth of the feedback is outstanding. Efforts have been made to deepen the model and create the first completely linked overfitting layer in order to enable the model to learn.

3.3.4. Final model

Six two-dimensional convolution layers, two maximum pooling levels, and two fully connected layers comprise the network. The preceding layer's Max Neurons use the maximum value for each cluster. This lowers the number of dimensions. Array of output values Array of output values Array of output values Array of output values The network input is pre-processed 48 by 48 pixel face. The model was created by studying the performance of earlier iterations. It It was chosen to go via a larger network. The benefit of having additional layers is that memorising is avoided. A large yet shallow network is readily stored, but it is not adequately generalised. Multi-layer networks provide abstraction functions that can be generalised effectively. The number of layers used to ensure high accuracy while being quick enough for real-time applications. The suggested CNN varies from a conventional CNN in that it employs four different coevolutionary and coevolutionary filter sizes. It also used maximum pooling and drop-out to reduce overfitting.

The network has two convolutional layers, each with a filter size of 64. Then comes a max layer of pooling. To minimise overfitting, a drop of 0.25 is utilised. A series of four convolutionary layers follows. The first two have 128 filters, whereas the second two contain 256 filters. A single layer of pooling After these four levels, there is a 0.25 decrease. The preceding output layers were smoothed to turn the result into a single vector dimension. After that, a fully connected 0.001 penalty controller L2 layer is utilised. Alogn reduces at a 0.5 percent faster pace. Finally, a complete layer with a softmax activation function is used. The output layer. The output layer. The kernel size is defined as the 2D width and height. Every convolutionary window is made up of three levels. Each maximum pooling layer has two dimensions and a pool size of two by two. This reduces output by 50% after each pooling layer. Every layer in the output layer utilised a ReLU. Activation of a Function Activation of a Function The ReLU activation function is now available. Because of advantages such as sparsity and a lower probability of gradient vanishing. The activation function Softmax has been used. The chance that the final production level will be reached. Every feeling. Every emotion. Every emotion. Every feeling. For the test, this model offered a 0.55 foundation. The hyperparameters, particularly the lot, were then set. Size, optimizer, and epoch count Each model was programmed to run for a total of 100 epochs. However, network time and computing power were

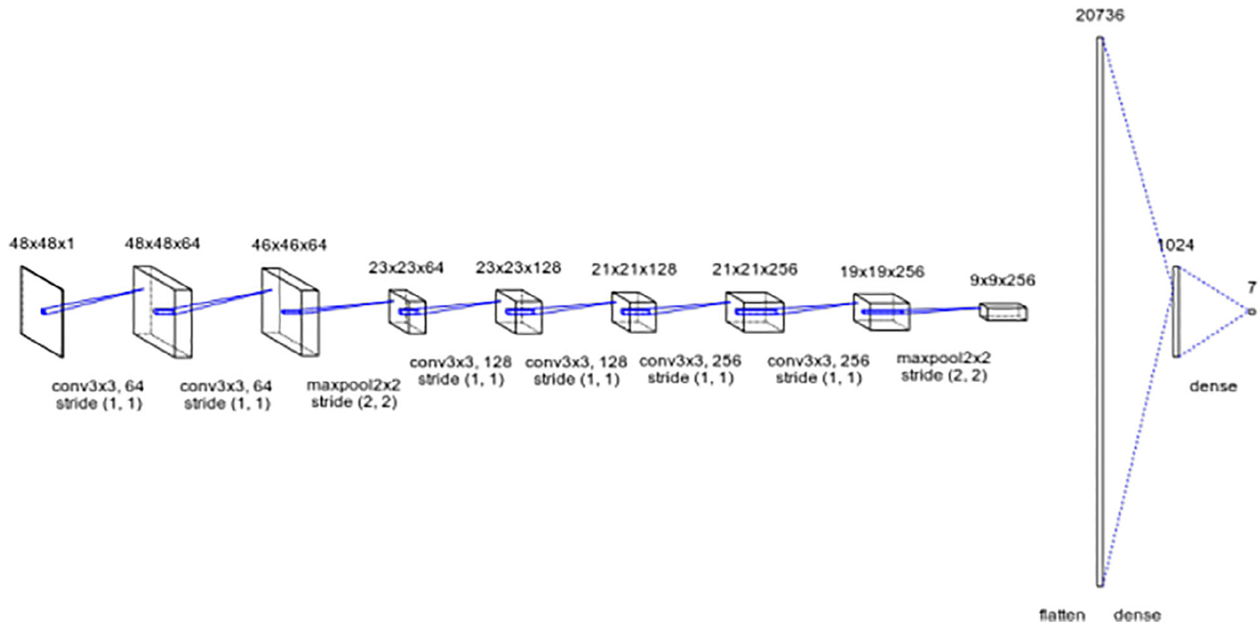

Fig. 3. Final Model Architecture.

allowed to be saved. Stop training if the precision of the epochs does not change. There have been no changes in accuracy across four consecutive epochs if the network stops training. Particularly when accuracy in previous periods saved time and processing resources. Earlier times. Earlier times. The choice was made since none of the models were more than 20 years old.

3.3.5. Testing

Initially, the dataset was split into an 80 percent training set and a 20% test set. During the testing phase, each trained network was loaded one image at a time, and the whole test set was loaded one picture at a time. There was a time. This was a brand-new photograph that the model had never seen before. Seen it all before, seen it all before. The image given to the model had been preprocessed. Exactly as stated in?? As a result, the model had no idea what the correct output was and had to be correct. Predict it based on your own experience. He attempted to categorise the emotion shown on the image purely based on what it has previously learned, as well as the image's characteristics. As a result, at the conclusion of each image's likelihood, it generated a list of categorise emotions. The most emotional possibility The actual emotions for each image were then compared with the connected with photos to count the precise quantity forecasts. The precision formula is provided below. It simply counts the number of items correctly predicted by the model as emotions and divides it by the total number of samples in the test set. See (Fig. 3).

4. Results

The first significant issue was the insufficient measurement of information to build a comprehensive framework. This must be resisted for the sake of nature's structure. Moves for learning This is the most common reaction. This is the method that has been started and changed from the original plan. Model that makes advantage of distant data A true reality. This is the actual world. A real world. The assumption was confirmed by start-up research. Recognition would be improved during highlight extraction. These systems are represented by models. The system correctly recognises a picture, classifies the image emotion, and selects the right image emotion. The use of deep learning classification is due to the classifier's use of many data layers. A deep learning algorithm, on the other hand, may be useful for less foreseeable problems since it is convinced of a vast amount of information. Over 14 million images are nearing the usual threshold for deep learning models for large-scale image recognition. To appropriately visualise the research of emotional patterns, a decision tree was employed. The character is represented in the decision tree by nodes and layers, and the experiment's result is likewise reflected by the branch. The decision tree has the benefit of making it extremely simple to see the feeling and comprehend the outcomes. The decision-making process is straightforward. When data is classified based on their mobility, responses, and sequences, the feelings are optimally distinct. This has also been divided into trees and subtrees, demonstrating that everything may be classified using these techniques, such as whether a person is sad, furious, or joyful, and so on. In order to do this, a retraining technique was utilised to remember and fulfil the pattern. If any of the conditions are fulfilled, it will proceed to the end of the tree. However, if the intermediate conditions are not met, the test terminates with the message, 'The emotion cannot be recognised.'

Everyone's emotions are unknown to them. Emotions are difficult to comprehend. The same feeling may be expressed in a variety of ways. Different individuals express the same feeling in different ways. Modern machine learning technologies may assist law enforcement in detecting emotions and understanding people's emotions. These emotional data are gathered from a variety of online and offline sources. kaggel.com is a website similar to Google. Family and friends, strange individuals, and so forth. Keras' library is used for categorising, evaluating, and analysing emotions. The emotion is then identified using hair and Numpy features. And with the assistance of the Anaconda platform. When the results are shown in real time, the raw data output is generated. The hierarchical data mining method, such as a decision tree, helps to generating probability choices by calculating different characteristics that are first used to determine the emotional pattern. It also conducted excellent field research, offline and online information gathering, and gathered more individuals, different types of people, and varied emotional expressions from various perspectives. The data set was obtained from kaggel.com through online data collecting. They offer high-quality data sets.

The pictures are transformed to pixel greyscale, and the number of shots is used. As a consequence, it provides outstanding data and batter outcomes. The specialists felt that this sentimental analysis might aid in more correctly detecting emotions and in taking the required measures to appropriately recognise emotions. It would give a better understanding of different kinds of sensations as well as the percentage of distinct emotions. During this study, we discovered that a large amount of test data and keywords are required to be more accurate. A substantial amount of raw data is also missing, which would help the study endeavour. To handle a huge amount of test data as fast as possible, a high-configuration graphics processing unit (GPU) is also required. If enough data is collected, it will be simpler to increase the accuracy with a high-performance computer to more than 97 percent. It may also utilise this approach to create a new platform for another result and aid in the discovery of a successful pattern of emotional expression.

5. Conclusion

An experienced person can often recognise another person's emotions by studying and gazing at them. Machines, on the other hand, are becoming more smart in our modern age. For the time being, machines are attempting to do more of the same. humans. If the computer is taught how to react in the present moment on behalf of human emotion. The computer may then behave like a human and respond accordingly. However, if the machine can identify emotions, data mining may aid in the discovery of proper expression patterns that computers can find and act as effectively as humans. This thesis is created or constructed via considerable research and fieldwork to discover the pattern of emotional expression. The framework then proceeded step by step to get the intended outcome. To adhere to the framework and more successfully identify emotional expressions, a deep learning CNN algorithm, as well as Keras, Tensorflow, and re-training concepts, will be used. Emotions, or types of emotions, may be recognised in the real image using these techniques. This also included decision-making tree techniques for determining which percentages of emotions are high and which percentages of emotions are low. The majority of emotions are now receiving the most exact sensations concedivable. And the minuscule percentage of emotions is unlikely to exist. Correct emotions may now be identified as a result of this discovery. And computers can more accurately identify emotion and react appropriately on its behalf, assisting in the avoidance of the same unfavourable event. This machine may also take the position of a human person. There are many examples as well.

CRediT authorship contribution statement

Ketan Sarvakar: Conceptualization, Methodology, Data curation. **R. Senkamalavalli:** Visualization. **S. Raghavendra:** Investigation. **J. Santosh Kumar:** Writing - original draft. **R. Manjunath:** Validation. **Sushma Jaiswal:** Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] F. De la Torre, J.F. Cohn, Facial expression analysis, Vis. Anal. Hum., pp. 377–410, 2011.

[2] A. Dhall, R. Goecke, S. Lucey, T. Gedeon, Static facial expressions in tough conditions: data, evaluation protocol and benchmark, First IEEE International Workshop on Benchmarking Facial Image Analysis Technologies BeFIT, IEEE International Conference on Computer Vision ICCV2011, Barcelona, Spain, 6-13, November 2011.

[3] A. Dhall, R. Goecke, S. Lucey, T. Gedeon, Collecting large, richly annotated facial-expression databases from movies, IEEE Multimedia, 2012, 19(3):3441, July 2012.

[4] Q.V. Le, N. Jaitly, G.E. Hinton, A simple way to initialize recurrent networks of rectified linear units. arXiv preprint arXiv:1504.00941, 2015.

[5] A. Kartali, M. Roglić, M. Barjaktarović, M. Đurić-Jovičić, M.M. Janković, Real-time algorithms for facial emotion recognition: a comparison of different approaches, in 2018 14th Symposium on Neural Networks and Applications (NEURAL), Nov 2018.

[6] Jonathan, A.P. Lim, Paoline, G.P. Kusuma, A. Zahra, Facial emotion recognition using computer vision, in Indonesian Association for Pattern Recognition International Conference (INAPR) IEEE, 31 January 2019.

[7] R.S. Deshmukh, V. Jagtap, S. Paygude, Facial emotion recognition system through machine learning approach, in International Conference on Intelligent Computing and Control Systems (ICICCS) IEEE, 11 January 2018.

[8] H.-J. Lee, K.-S. Hong, A study on emotion recognition method and its application using face image, in International Conference on Information and Communication Technology Convergence (ICTC) IEEE, 14 December 2017.

[9] D.Y. Liliana, C. Basaruddin, M.R. Widyanto, Mix emotion recognition from facial expression using SVM-CRF sequence classifier, 2017, pp. 27–31.

[10] 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.

[11] D.Y. Liliana, 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.

[12] Z. Ming, A. Bugeau, J.-L. Rouas, 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.

[13] J. Chen, D. Chen, L. Wang, Facial expression recognition using geometric and appearance features, Icimcs (2012) 29–33.

Further Reading

[1] Y.-L. Tian, T. Kanade, J.F. Cohn, Recognizing action units for facial expression analysis, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. CVPR 2000 Cat NoPR00662 1 (2) (2001) 1–19.

[2] A. Gudi, H.E. Tasli, T.M. Den Uyl, A. Maroulis, Deep learning based FACS Action unit occurrence and intensity estimation, vol. 2013, 2015.

[3] Z. Ming et al., Facial action units intensity estimation by the fusion of features with multi-kernel support vector machine to cite this version : facial action units intensity estimation by the fusion of features with multi-kernel support vector machine, 2015.

[4] S. Ebrahimi Kahou, V. Michalski, K. Konda, R. Memisevic, C. Pal, Recurrent neural networks for emotion recognition in video. Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, 2015, 467-474. ACM.

[5] D. Melanie, Emotional expression recognition using support vector machines, July 2001.

[6] M. Abdulrahman, Facial expression recognition using support vector machines, in 23nd Signal Processing and Communications Applications Conference (SIU) IEEE, 22 June 2015.