An experimental study in Real-time Facial Emotion Recognition on 3RL dataset

Rahmeh Abou Zafra (201620003@aiu.edu.sy)
Arab International University https://orcid.org/0000-0001-9092-1644
Lana Ahmad Abdullah (201620013@aiu.edu.sy)
Arab International University
Rouaa Alaaraj
Arab International University
Rasha Albezreh
Arab International University
Tarek Barhoum
Arab International University
Khloud Al Jallad
Arab International University https://orcid.org/0000-0001-9474-9204

Research Article

Keywords: Facial Emotion Recognition, Facial Emotion Recognition, Dataset, Deep Learning, Computer Vision

Posted Date: March 11th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1439248/v1

An experimental study in Real-time Facial Emotion Recognition on 3RL dataset

Rahmeh Abou Zafra*, Lana Ahmad Abdullah, Rouaa Alaaraj, Rasha Albezreh, Tarek Barhoum, Khloud Al Jallad

Arab International University
Daraa, Syria
*Correspondence: 201620003@aiu.edu.sy
*ORCID https://orcid.org/0000-0001-9092-1644

Abstract— Although real-time facial emotion recognition is a hot topic research domain in the field of human-computer interaction, state-of-the-art available datasets still suffer from various problems such as some unrelated photos like document photos, unbalanced number of photos in each class, and some misleading images that can affect negatively on correct classification. 3RL dataset was created which contains about 24K images and will be publically available, to overcome previously available datasets problems. 3RL dataset is labelled with five basic emotions: happiness, fear, sadness, disgust, and anger. Moreover, we compared 3RL dataset with other most famous state-of-the-art datasets (FER dataset, CK+ dataset), we have applied the most common used algorithms in previous works, SVM and CNN. Results have shown a noticeable improvement of generalization on 3RL dataset. Experiments have shown an accuracy of up to 91.4% on 3RL dataset using CNN where results on FER2013, CK+ are respectively (approximately from 60% to 85%).

Keywords—Facial Emotion Recognition; Facial Emotion Recognition; Dataset; Deep Learning; Computer Vision.

Declarations
Funding
The authors declare that they have no funding.

Conflicts of interest/Competing interests
The authors declare that they have no competing interests.

1. INTRODUCTION

Facial Emotion Recognition (FER) is the technology that analyses facial expressions from both static images and videos in order to reveal information on one's emotional state. The shape of eyebrows, lips, nose, chin plays an important role in determining facial expressions.

This paper proposed a new dataset to overcome those problems. Moreover, two models are applied to compare our dataset with previously available datasets.

The first model is SVM using landmark and HOG feature descriptors extracted from images. The second model is CNN using images and landmarks as features.

Paper is organized as follows, section 1 is an introduction, section 2 is related works. Section 3 is about previously available datasets Experiments are shown in section 4 and then setup, and finally section 6 contains conclusion and future work.

2. RELATED WORKS

SVM is one of the most used machine learning models in FER as in (Alshamsi and Këpuska)[1]1, Alshamsi et al use SVM for classification, COG and landmarks for feature extraction. In (Patwardhan)[2], (Youssef, Aly and Ibrahim)[3], (Zhang , Cui and Liu)[4] a geometric, kinematic and extracted features from daily behavioral patterns used in feature extraction part, whereas SVM algorithm used to recognize the emotions., constructed a facial emotion recognition dataset which contains 84 samples. SVM and k-NN are used to classify emotions.

Support vector machines (SVMs) have proved its qualification of multi class classification. As in (Alshamsi and Këpuska)[1] , (Joseph and P. Geetha)[5], (Lucey, Cohn and Kanade)[6], (Duan and Keerthi)[7], (Chandran and Dr. Naveen S)[8], (GLAUNER)[9] to classify emotions.

SVMs are usually implemented by combining several two-class SVMs. The hyperplane SVMs use in n dimensional space distinguish points to correct classes by their labels. The optimal hyperplane is the one that is able to form the biggest distance between dataset points in this case images. SVMs avoid the "curse of dimensionality" that arises in high-dimensional spaces by tuning the regularization parameter c in linear problems or by kernel trick along with tuning the kernel parameters in non-linear problems.

The second approach is the adoption of deep learning algorithms like CNN is the most used deep learning model used for emotion recognition.

Hafiz et al in (Ahamed, Alam and Islam)[10] proposed CNN and with HOG as feature extraction in (Ahamed, Alam and Islam)[10], But in (Li and Deng)[11], Li & Deng, they used handcrafted HOG features and the deep learned features, and linear SVM for

classification, in (GLAUNER)[9], Patrick also used CNN and for feature extraction handcrafted features was used. to determine the appropriate feeling for the input image and by reading many articles and researches, finding that the best algorithms used and the most accurate in machine learning to describe the feeling is the algorithm of the SVM Support Vector Machine and that the best deep learning algorithm in this field is the Convolution Neural Network algorithm, so these two algorithms are used mainly in this study and made some structural improvements and adjustments for both networks to get the best results, and then we did a simple comparison to find out the differences between using deep learning algorithms and machine learning to determine the feeling from the human face. In the following Table 1, a brief search on the state of the art was done in order to know the experiments were before and what are the most effective methods to start with.

	Network	Features	Dataset	Accuracy
(GLAUNER)[9]	CNN Particular smile recognition	Hand-crafted	DISFA	99.45%
(Li and Deng)[11]	ECAN Classified seven emotion classes	Deep learning & HOG	JAFFE, MMI, CK+ Oulu-CASIA FER2013 SFEW	61.94% 69.89% 89.69% 63.97% 58.21% 58.21%
(Minaee and Abdolrashidi)[12]	CNN	Hand-crafted	CK+ FER2013 FERG JAFFE	98% 70.02% 99.3% 92.8%
(Ghaffar)[13]	CNN Classified seven emotion classes		JAFFED + KDEF	78%
(Liu, Cheng and Lee)[14]	SVM with genetic algorithm	Geometric feature (landmark curvature, victories landmark)	8-class CK+ 7-class CK+ 7-class MUG	93.57% 95.58% 96.29%
(Maw, Thu and Mon)[15]	SVM		JAFFE	80%
(Alshamsi and Këpuska)[1]	SVM	Facial Landmarks, BRIEF		96.27%

Table 1: State of the art

3. DATSETS

3RL dataset consists of 24394 images collected and edited from three famous datasets.

As Table 1 shows, the results on CK+ are better than the results on other datasets in terms of accuracy, but it is the smallest dataset between them. “FER-2013”, “CK+” and a dataset for facial expression recognition which contains more images to work on, mixing them will lead to sufficient number of images. However, many misleading images were found in the mentioned datasets, so manually these images were checked one by one and some deleting, replicating and replacing processes were done during checking them and it was noticed a huge number of misleading images that may effects immediately and negatively on the output of the applied network. This filtering process has a noticeable improvement on model performance as well as balancing between classes was applied to obtain fair results. The resulted dataset (3RL) will be available at 1.

The three datasets used to create 3RL dataset are

1-Facial Expression Recognition (FER-2013) (Facial Expression Recogntion,(FERc), 2013)[16]

2- Extended Cohn-Kanade Dataset (CK+48) (Cohn-Kanade: (CK+), 2010)[17]

3-a dataset for facial expression recognition (Facial expressions)[18].

The tables 2, 3, 6 Show statistics about previously available datasets.

3.1 FER-2013

	Test	Train	sum
Angry	958	3995	4953
Disgust	111	436	547
Fearful	1024	4097	5121
Happy	1774	7215	8989
Neutral	1233	4965	6198
Sad	1247	4830	6077
Surprised	831	3171	4002
Sum	7178	28709	35888

Table 2: FER-2013 dataset

3.2 CK+48

	Images number
Angry	135
Disgust	177
Fearful	75
Happy	207
Neutral	54
Sad	84
Surprised	249
Sum	981

Table 3: CK+48 dataset

	Test	Train	Sum
Angry	474	3190	3664
Disgust	86	399	485
Fearful	493	1015	1508
Happy	1547	2678	4225
Neutral	805	2552	3357
Sad	1134	4228	5362
Surprised	504	2342	2846
Sum	5043	16404	21447

Table 4: Merged & Edited dataset

3.3 Facial expression recognition dataset

This dataset contains 13718 images collected without classification.

Dataset creation steps

1) In order to obtain better results including the biggest number of facial situations, merging datasets maybe a good practice so, after merging the CK+ 48 dataset which contains (981 images) with Fer2013 dataset which contains (35887images), the final merged dataset was 36858 images. The dataset was edited manually by removing all misclassified images and some images that contain hands which cover the face completely. The edited merged dataset was in table 4.

2) Next, dataset (CK+ & FER-2013) after deleting all replicated images or images that contain confusion and splitting data, the dataset was as in table 5

3) A third dataset was added (dataset for facial expression) with the previous datasets (CK+ & FER (2013)). This dataset contains 13718 images with (350*350) size which is available at (Facial expressions)[18] and the dataset was merged and concise from the last seven emotions to five classes (merging between anger & disgust, sad & fear) and almost divided to (80% train and 20% test) with balancing between classes (each class contains approximately the same number of images) in each train class 2000 images and in each test class 400 as in the table 6

4) Finally, the dataset was maximized by adding some images and replicating others that may contain strong features - which benefit the model to classify better- with low number of images that include this feature, so that each train class had 4000 images, and each test class 800 images to obtain the 3RL dataset, as shown in table 7

	Test	Train	Sum
Angry	418	1908	2326
Disgust	74	299	373
Fearful	233	936	1169
Happy	1435	5744	7179
Neutral	671	2688	3359
Sad	445	1782	2227
Surprised	418	1676	2094
Sum	3694	15033	18727

Table 5: Cleaned dataset

	Test	Train	Sum
Angry	484	2069	2553
Happy	492	2039	2531
Neutral	401	2000	2401
Sad	403	2031	2432
Surprised	400	2046	2446
Sum	2180	10185	12365

Table 6: concise dataset

	Test	Train	Sum
Angry	838	4040	4878
Happy	822	4061	4883
Neutral	854	4070	4924
Sad	822	4057	4879
Surprised	816	4014	4830
Sum	4152	20242	24394

Table 7: 3RL dataset available here

The whole process explained earlier in all three datasets was concise in the conceptual diagram in figure 1.

A random collection of photos is shown from the final 3RL dataset in the figure 2.

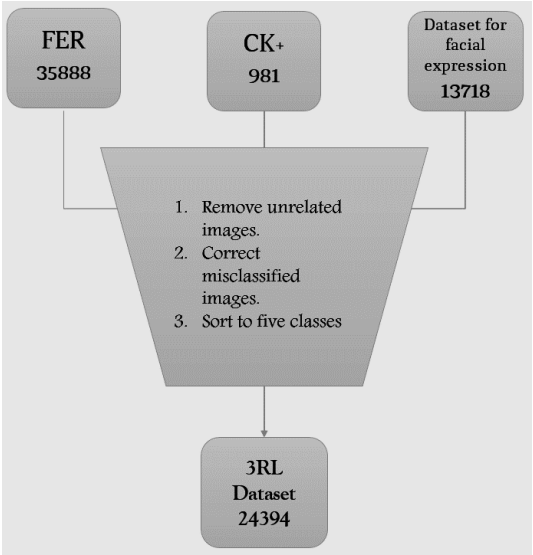

Figure 1: 3RL dataset conceptual diagram

Figure 2: 3RL dataset samples

4. SETUP

All experiments were done on core i7 from seventh generation, 2.8 GHz processor and 16 Giga RAM memory.

OpenCV was used to permit using the camera, also TensorFlow and Keras were used.

Most of the experiments had been done with 50 epochs because of a repetitive process was noticed when increasing the number of epochs without obvious difference in accuracy.

In the last experiment in Table 11, 50 epochs of training were taking about seven hours.

5. EXPERIMENTS

1. SVM

The first experiment on SVM was implemented on FER-2013 as in Table 2 dataset after extracting Landmark and Histogram of Oriented Gradients (HOG) features using 16*16 pixels per cell and 8 bins pixels' gradients will be split up to in histogram. Using Gridsearch, the best hyper parameters we got are: parameter c equals 1, gamma parameter for regularization equals 0.1 with max iterations 10,000 and decision-function one-vs-rest strategy accuracy achieved:

Kernel	Acc	Test acc
RBF	99%	29%
Linear	21%	14%
Sigmoid	24%	29%
Poly	32%	14%

Table 8: Machine FER-2013 result c=1, g=0.1

As shown in table 8, experiments were conducted on different kernels. Best achieved result was by using Radial Basis Function with accuracy 99%, validation accuracy 29%.
Same results were found when using Histogram of Oriented Gradients sliding window instead.
On 3RL dataset as in Table 7 better results were achieved with same decision function, features and kernel Radial basis function:

C	Gamma	Preprocess	Acc	Test acc
1	1	ST+PCA136	92%	37%
1	0.2	ST+PCA208	99%	40%

Table 9: Machine result 3RL

The above Table 9 demonstrates results when preprocessing the features to avoid previous overfitting problem.
3RL dataset's extracted features were preprocessed with standard scalar and using principle components analysis that preformed feature reduction once with 136 components obtained 92% accuracy and 37% validation accuracy. Second with 208 components obtained 92% accuracy and 37% validation accuracy. Still not satisfied so, extra attempts were made in below Table 10 without preprocessing and got:

C	Gamma	Acc	Test acc
0.1	0.01	20%	20%
1	0.1	99%	58%
1	0.2	99%	58%

Table 10: Machine Result (without preprocessing)

Table 10 reveal the final experiments implementing SVM on HOG and Landmark features achieving 99% accuracy and 58% validation accuracy.

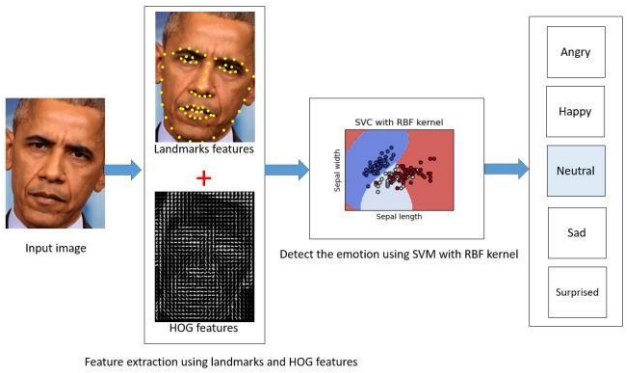

Figure 3: ML architecture

Figure 3 shows ML architecture where input image is processed by Landmark and HOG feature descriptors. Extracted features are fed to the SVM classifier with Radial basis function kernel, decision function one-vs-rest, parameter c with value 1, parameter gamma assigned 0.2.

Dataset	Train Accuracy	Test acc
FER-2013	99%	29%
CK+48	33%	32%
3RL dataset	99%	58%

Table 11: dataset ML comparative

Presenting in Table 11 dataset's accuracy with best achieved parameters for RBF, c, gamma based on Table 9 and Table 10.

2. Deep Learning

First, a similar approach as in (Correa, Jonker, Ozo, & Stolk, 2016)[19], (Chandran and Dr. Naveen S)[8], (Li and Deng)[11], (Zavarez, Berriel, & Oliveira-Santos, 2017)[20] was followed to extract features by CNN, and mainly as (Correa, Jonker, Ozo, & Stolk, 2016)[19] approach was followed and started experiments on FER2013 dataset (Table 2), four layers of convolutional neural network, the filters number in each layer was 32, 64, 128 & 128 for the fourth layer, 64 batch size and 150 epochs with changing the values of network parameters in order to obtain the best results as shown in the Table 12 below:

LR	Active	Loss FN	Dropout	Acc	Test acc
e-4	ReLU	categorical	0.25 0.25 0.5	86%	62%
3e-4	ReLU	categorical	0.25 0.5 0.5	72%	62%
3e-4	Tanh	binary	0.25 0.5 0.5	96%	92%

Table 12: DL results 150 epochs, FER-2013

As noticed, the last result is considered acceptable but the prediction results were not satisfied in all previous experiments.

Therefore, a change was made to the merged dataset in Table 4 keeping learning rate ($3e^{-4)}$ and the layers as it was and changing the drop out layers to 0.35, 0.5 and 0.5 for the last layer, then the results as followed in Table 13 were obtained:

Batch	LR	activate	Loss FN	epoch	Acc	Test acc
64	3e-4	ReLU	categorical	150	95%	66%
32	1e-4	ReLU	Binary	50	88%	68%
32	1e-4	Tanh	Binary	100	98%	90%

Table 13: DL results with Table 4 dataset

As shown in the previous Table 13, an improvement in predict was noticed in the last two experiment when heading to batch size 32, so in the next experiments, the batch size of 32 will be kept.

Another try was done using the dataset in Table 5 keeping the same other parameters as it was in the last experiment and 98.9% of training accuracy was achieved and 92.7% as a validation accuracy, the predict was good but still not enough.

Therefore, the model was trained with the three dataset adding a dataset for facial expression recognition, the last dataset was cleaned and sorted to five main classes (emotions) as in Table 6, and edited the drop out layers to 0.35, 0.5 & 0.5 for the last one, for the results were as follow in Table 14:

Batch	LR	activate	Loss FN	epoch	Acc	Test acc
32	1e-4	ReLU	categorical	50	82%	66%

Table 14: DL results table 6 dataset

For this result in Table 14 the predict was considered not bad but reaching a higher accuracy was better, so concluded that may the number of images in the dataset considered not enough for deep learning, therefore the dataset was maximized as in Table 7.

Here some results with 3RL dataset, the batch size was 32, the learning rate was $1e^{-4}$ and the drop out layers were as in the previous try.

activate	Loss FN	epoch	Acc	Test acc
tanh	Binary	200	96%	80%
sigmoid	Categorical	50	90.9%	90.4%

Table 15: DL results with 3RL dataset

In Table 11 the predict was bad so the activation function was changed to ReLU, the loss function to binary and the last layer of drop out to 0.25 keeping other parameters as it was, an improvement was noticed in predict, so more experiments were done to obtain a better predict.

Then, tying to change the number of convolutional layer's filters to be 32, 64, 128 and 512 for the last one, the learning rate was e^{-4} and the batch size was 64, Then some experiments were done as follow in Table 16:

Active	Loss FN	Dropout	Acc	Test acc
ReLU	binary	0.25 0.1 0.25	99.8%	91%
ReLU	categorical	0.01 0.1 0.25	99.5%	77%
ReLU	categorical	0.01 0.5 0.25	99%	79.5%
ReLU	categorical	0.01 0.03 0.25	99.5%	76%
tanh	categorical	0.01 0.5 0.25	97.7%	79.8%
Sigmoid	binary	0.25 0.1 0.25	99.9%	91.4%

Table 16: DL results with different parameters

As noticed in Table 16, if tangent function was used as an activation function with binary cross entropy as a loss function the result will be more satisfying than using tangent with categorical cross entropy. The best drop out layers were achieved in the first experiment at the previous table so those dropout layers will be kept.

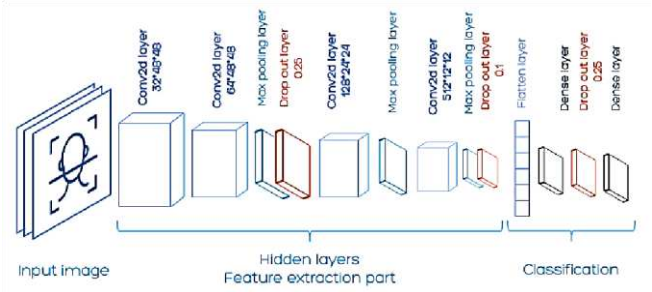

Figure 4: Proposed network architecture

The previous figure shows the architecture of proposed convolutional neural network in feature extraction and classification parts using ReLU function as an activation function in each convolution layer, Sigmoid function in first dense layer and softmax for the last dense layer.

Finally, the desired result was obtained achieving the best accuracy at 99.9% and 91.4% as a validation accuracy, 0.4 for validation loss and 0.004 for loss, the prediction was very good; the network was able to detect all five emotions correctly, the following figure 5 shows the curves of accuracy and loss in the last experiment in Table 16.

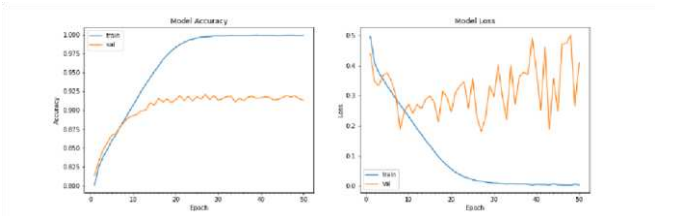

Figure 5: Accuracy and loss curves

In the following table 17: a short comparison had been done to show the benefit of 3RLdataset in prediction phase where models were tested on FER-2013, CK+48, 3RL with keeping the same parameters as in the last result in table 16, changing only the dataset.

Dataset	Acc	Test acc	Prediction
FER-2013	99.8%	89%	Recognize happy, neutral, fear and sad emotion with misclassification in other classes.
CK+48	99.9%	95.9%	Only disgust emotion was recognized correctly
3RL dataset	99.9%	91.4	Best prediction results between all experiments

Table 17: datasets DL comparative

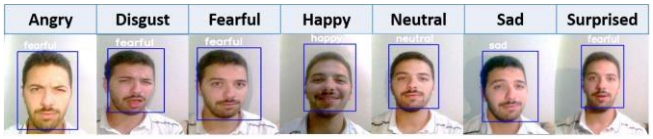

Figure 6: FER-2013 predicting results

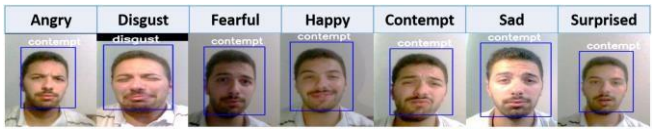

Figure 7: CK+48 predicting results

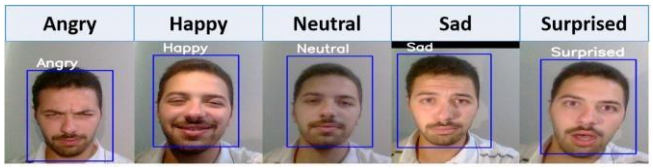

Figure 8: 3RL predicting results

So, during experiments, screenshots was taken from real time applying the best parameters as in the last try in Table 16 on the three datasets (FRE-2013, CK+48, 3RL dataset).

A complete disappearance of many emotions was noticed with some errors in prediction, and feeling fear was dominating the rest of the feelings when using the FER-2013 Dataset as shown in figure 6, with the first row of results in Table 17.

When using CK+48 as a dataset, the second row of table 17 shows the accuracy details, in Figure 7 it is noticeable that most of the emotions are classified as contempt with the absence of a complete classification of other feelings except disgust which is classified correctly, as the use of this dataset showed a significant decline in the classification compared to the FER-2013 dataset, which gave better results and a noticeable improvement in prediction than CK+48.

In 3RL dataset, the last row in Table 17, as screenshots of figure 8 no absence of any feeling from the classification was noted, but very slight and minor errors were noted, such as confusing the feelings of disgust with anger when

using the eyebrow-contract feature to express both feelings, which makes this dataset outperform its previous competitors.

6. CONCLUSION & FUTURE WORK

Emotion detection is the key to human interaction and understanding. Emotion detection is a challenging task. CNN is one of the best solutions for classifying facial emotions using large datasets such as 3RL. In this study, a new dataset named 3RL was created combining three datasets CK+48, Fer2013 and dataset for facial expression recognition plus merging classes of emotion into 5 main ones. 3RL dataset will be available at 1. Shown experiments have given high accuracies of 99%, whereas generalization was better implementing CNN as DL method.

The Final CNN architecture consists of four 2-D convolutional layers, three dropout layers (32, 64, 128, 512), three max pooling layers and two fully connected layers. The input to the network is a preprocessed gray image of 48x48 for face. The number of layers was selected to maintain a high level of accuracy while still being fast enough for real-time purposes. In addition, it is utilized max pooling and dropout more effectively in order to minimize overfitting.

In the future, increasing the number of detected emotions will allow the user to express more emotions, and expanding the dataset to include more people of different nationalities, which helps the network to better predict and apply voice along with facial expressions to predict more real feelings.

7. *REFERENCES*

1. Alshamsi, H. S., & Këpuska, V. Z. (2017). Real-Time Facial Expression Recognition App Development on Smart Phones. International Journal of Engineering Research and Applications 07(07):30-38.10.9790/9622-0707033038.

2. Patwardhan, Amol S. "Three-Dimensional, Kinematic, Human Behavioral Pattern-Based Features for Multimodal Emotion Recognition." Multimodal Technologies and Interaction, 1(3):19 (2017).

3. Youssef, Amira E , et al. "Auto-Optimized Multimodal Expression Recognition Framework Using 3D Kinect Data for ASD Therapeutic Aid." International Journal of Modeling and Optimization, 3(2), 112. (2013).

4. Zhang , Zhan , et al. "Emotion Detection Using Kinect 3D Facial Points." In 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI) (2016): 407-410. IEEE.

5. Joseph, Allen and P. Geetha. "Facial emotion detection using modified eyemap–mouthmap algorithm on an enhanced image and classification with tensorflow." The Visual Computer 36.3 (2020).

6. Lucey, Patrick , et al. "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression." IEEE Computer Society Conference on Computer Vision and Pattern recognition - workshops, 2010, pp. 94-101 (2010).

7. Duan, Kai-Bo and S. Sathiya Keerthi. "Which Is the Best Multiclass SVM Method? An Empirical Study." In International workshop on multiple classifier systems (pp. 278-285). Springer, Berlin, Heidelberg. (2005).

8. Chandran, Megha and Dr. Naveen S. "A Review on Facial Expression Recognition using Deep Learning." International Journal of Engineering Research and Technology (IJERT) (2019).

9. GLAUNER, PATRICK O. . "DEEP LEARNING FOR SMILE RECOGNITION." Uncertainty Modelling in Knowledge Engineering and Decision Making: Proceedings of the 12th International FLINS Conference (2016).

10. Ahamed, Hafiz , Ishraq Alam and Md. Manirul Islam. "HOG-CNN Based Real Time Face Recognition." 2018 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE) (2018). IEEE.

11. Li, Shan and Weihong Deng. "A Deeper Look at Facial Expression Dataset Bias." IEEE (2019).

12. Minaee, Shervin and Amirali Abdolrashidi. "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network." (2019).

13. Ghaffar, Faisal . "Facial Emotions Recognition using Convolutional Neural Net." ArXiv abs/2001.01456 (2020).

14. Liu, Xiao , Xiangyi Cheng and Kiju Lee. "GA-SVM based Facial Emotion Recognition using Facial Geometric Features." In IEEE Sensors Journal, 21(10), 11532-11542. (2020).

15. Maw, Hla Myat , Soe Myat Thu and Myat Thida Mon. ""Vision Based Facial Expression Recognition Using Eigenfaces and Multi-SVM Classifier"." International Conference on Computational Collective Intelligence (pp. 662-673). Springer, Cham (2020).

16. Facial Expression Recogntion, (FERc). (2013). Kaggle (online): https://www.kaggle.com/deadskull7/fer2013

17. Cohn-Kanade: (CK+). (2010). Retrieved from (Online):

    https://www.kaggle.com/shawon10/ck-facialexpression-detection

18. Facial expressions. Available online at:

    https://github.com/muxspace/facial_expressions

19. Correa, E., Jonker, A., Ozo, M., & Stolk, R. (2016). Emotion recognition using deep convolutional neural networks. Tech. Report IN4015.

20. Zavarez, M. V., Berriel, R. F., & Oliveira-Santos, T. (2017). Cross-Database Facial Expression Recognition Based on Fine-Tuned Deep Convolutional Network. In 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI) (pp. 405-412).