Ensemble Learning for Facial Expression Recognition

Anjali Raghav , Monika Gupta

Maharaja Agrasen Institute of Technology, INDIA
Emails : anjali.raghav.261@gmail.com; monikagupta@mait.ac.in

* Correspondence: anjali.raghav.261@gmail.com

Abstract

Facial expressions are the translation of the emotions such as anger, sadness, happiness, and disgust felt by a person. Facial expression recognition and classification of expressions which has applications in various industries such as hospitality, and medical, to name a few. There are various datasets available for facial expression recognition. We used F.E.R. 2013 dataset to build a classification algorithm. This algorithm classifies emotions into seven categories: anger, disgust, happiness, sad, fear, surprise, and neutral. The computing time is very large in a traditional convolutional neural network algorithm. Ensemble learning significantly reduced the computing time and offered a promising accuracy. Features of images were extracted using the convolutional neural network. Further, these features were implemented using XGBoost and Random Forest to build classification algorithms, with an accuracy of 77% and 74%were. This was comparable to the accuracy obtained by the traditional convolutional neural network, which was 75% also with very less computing time.

Keywords: Ensemble Learning ; Facial Expression Recognition

1. Introduction

Human-machine interaction has significantly increased in the present scenario. Facial expression recognition (F.E.R.) improves human-computer interaction and makes the experience more user-friendly. The facial expression represents 55% of human communication. Facial expressions can be captured using sensors such as a camera, eye-tracker, electrocardiogram (E.C.G.), electromyography (E.M.G.), and electroencephalograph (E.E.G.). Facial expression recognition algorithms are used to classify facial expressions into various categories. Popular datasets used for facial expression recognition as FER2013 and JAFFE. Initially, the images are pre-processed, rotated, and sharpen. This helps in better feature extraction. Even minute details are considered. These extracted features are then implemented on classification algorithms. F.E.R. can be used in building a music recommender system. Facial expression can process the mood, which will help in selecting the genre and thus suggesting the songs[1]. F.E.R. system can be installed in hospitals to alert the caretaker when a sudden discomfort or sudden change in behavior is shown by the patient. This will help in providing adequate rest to nurses and immediate care to the patient. Deep neural networks are densely connected architectures that are most commonly used in image processing, but their major drawback is large computation time[2]. This is overcome by the proposed algorithms in which CNN is used for feature extraction, and ensemble learning is used for classification. This paper is divided as Section 1 gives a brief introduction to the paper, in Section 2, the results literature survey are enlisted, in Section 3, the methodology is discussed, Section 4 discusses the result, and conclusions are mentioned in Section 5.

Doi :https://doi.org/10.54216/FPA.020104
Received: March 20, 2020 Revised: April 28, 2020 Accepted: May 30, 2020

2. Literature Survey

	Research Paper	Year of Publishing	Algorithms Used
1	Deep convolution network-based emotion analysis toward mental health care[3]	2020	CNN- AlexNet, GoogleNet, VGG-16, ResNet
2	eXnet: An Efficient Approach for Emotion Recognition in the Wild[4]	2020	eXnet
3	A Review on Facial Expression Recognition using Deep Learning[5]	2020	CNN, SVM
4	Three convolutional neural network models for facial expression recognition in the wild[6]	2019	CNN
5	Recognizing learning emotion based on convolutional neural networks and transfer learning[7]	2019	CNN, transfer learning

3. Methodology

3.1.Facial Expression Recognition Dataset

The Facial Expression Recognition 2013 dataset was published by Wolfram Research Data Repository. The dataset consisted of 35685 grayscale images of size 48x48 each. There were seven classes of emotions, namely, angry, disgust, fear, happy, sad, surprise, and neutral, in the dataset. fig. 1 shows the sample of images of the FER2013 dataset. Table I gives the classification of images based on emotions.

Figure 1: Example of Images from FER2013 dataset

Doi :https://doi.org/10.54216/FPA.020104
Received: March 20, 2020 Revised: April 28, 2020 Accepted: May 30, 2020

Table 1 : ORIGINAL CLASSIFICATION OF IMAGES

EMOTION	NUMBER OF IMAGES	LABEL
Angry	4593	0
Disgust	547	1
Fear	5121	2
Happy	8989	3
Sad	6077	4
Surprise	4002	5
Neutral	6198	6

It was observed that the number of images was distributed unequally. Therefore, the dataset was resampled to rule out the possibility of a biased classification and false result speed. Table II shows the classification of images after resampling. The data was split in a ratio of 80:20 for training and testing, and a further 20% of the training data was kept for validation purposes.

Table 2: CLASSIFICATION OF IMAGES AFTER RESAMPLING

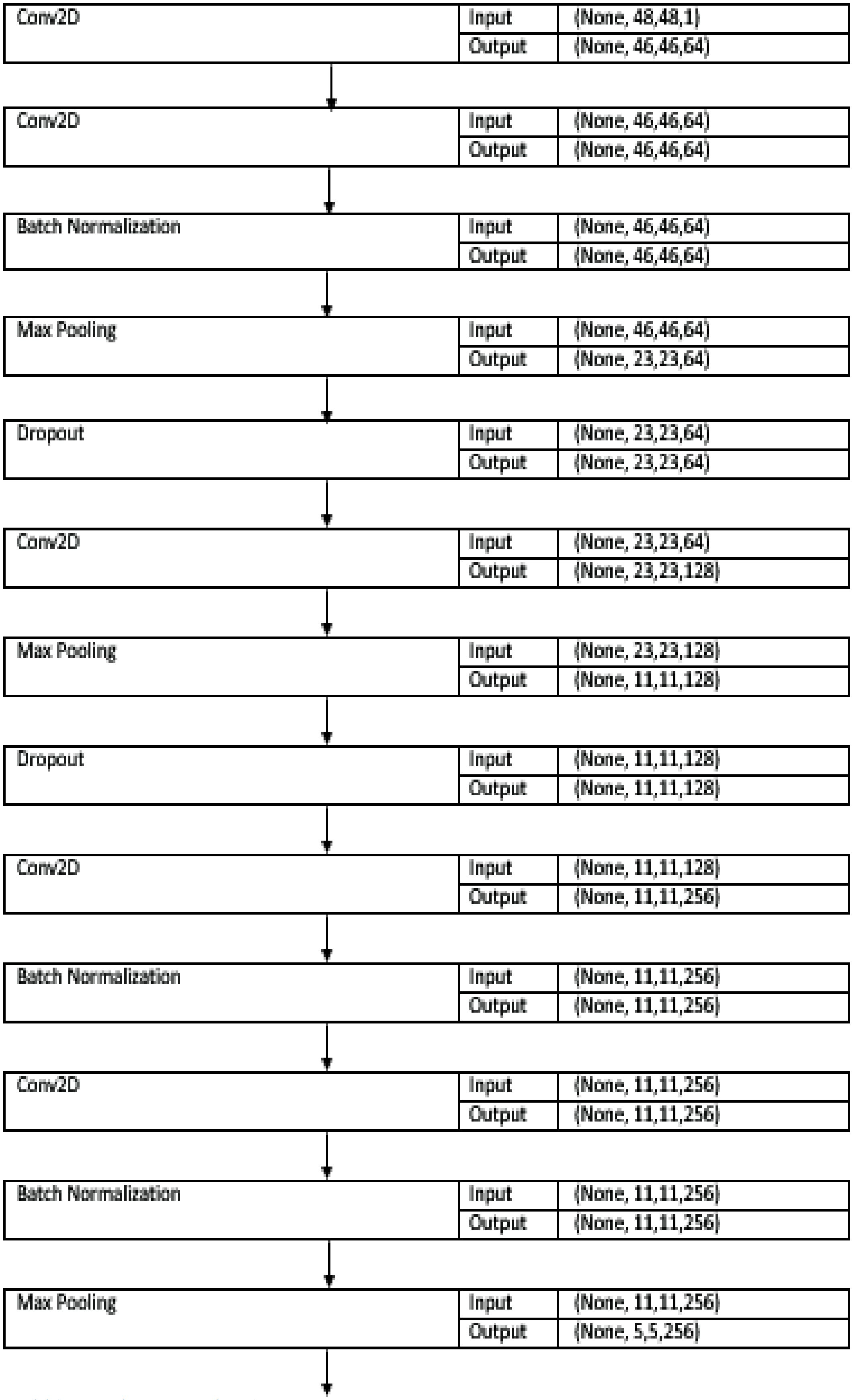
EMOTION	NUMBER OF IMAGES	LABEL
Angry	9186	0
Disgust	8192	1
Fear	10242	2
Happy	8989	3
Sad	6077	4
Surprise	4002	5
Neutral	6198	6

3.2.Algorithms

3.2.1 Convolutional Neural Network with Softmax Classifier

Convolutional Neural Network (CNN) is a deep learning network usually applied for image visualization. To reduce the processing requirements, CNN is used, which is similar to a multi-layer perceptron. CNN is inspired by human visual stimuli. Neurons are arranged similar to the frontal lobe present in humans and animals. CNN architecture comprises input layers, hidden layers, and output layers. The hidden layers comprise various layers such as convolutional layers, pooling layers, and normalization layers to get visual every piece of the image like the human eye. The architecture used in this paper is shown in fig 2.

Doi :https://doi.org/10.54216/FPA.020104
Received: March 20, 2020 Revised: April 28, 2020 Accepted: May 30, 2020



Doi :https://doi.org/10.54216/FPA.020104
Received: March 20, 2020 Revised: April 28, 2020 Accepted: May 30, 2020

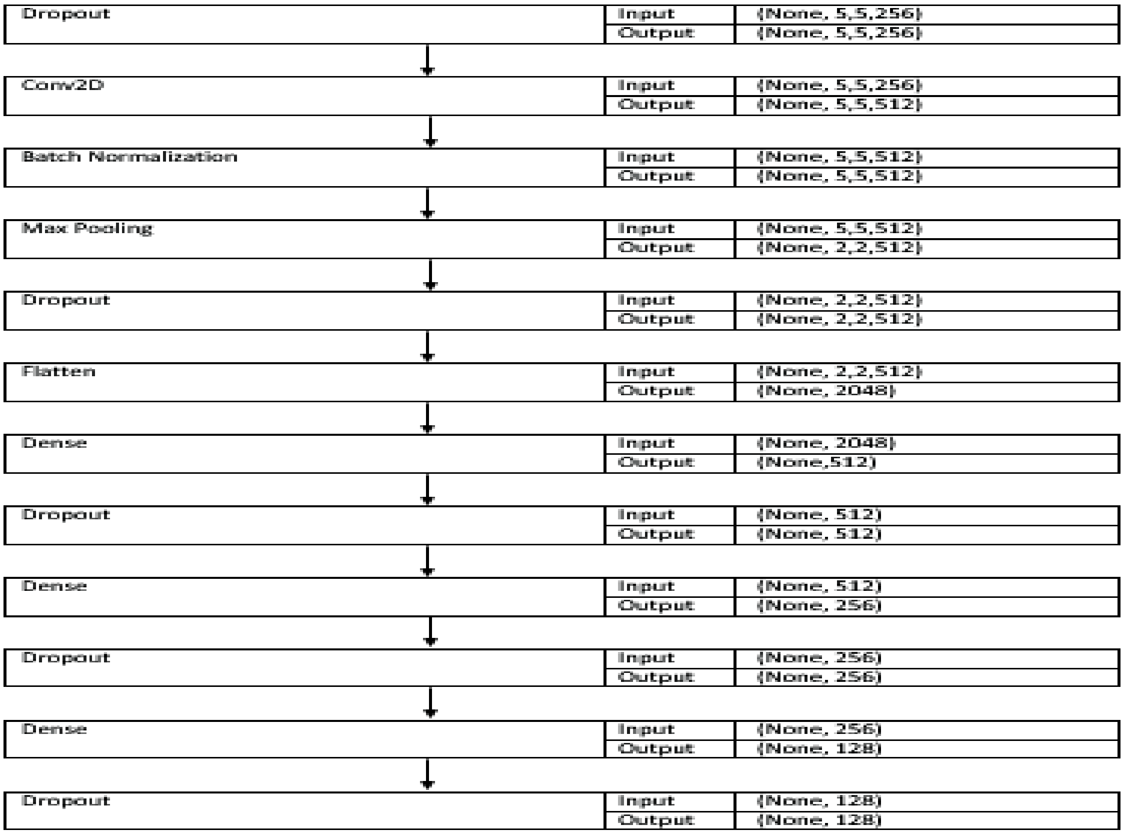


Figure 2: CNN Architecture Used

The softmax activation layer was installed at the last dense layer. Adam optimizer was used, which had a learning rate of 0.001, with loss as categorical cross-entropy, for 100 epochs.

3.2.2 XGBoost

XGBoost is an ensemble learning method. Boosting algorithms have weak base learners, but the effective combination of these weak learners results in the strong learner. Sequential building of trees by aiming to reduce error in the next tree in comparison to the previous tree happens in boosting. Hence, the subsequent tree learns from the residuals of the updated version. Gradient boosted decision tree is designed for faster processing and better performance.

We implemented this algorithm along with CNN, features are extracted from images using the architecture mentioned in 3.2.1, and further classification of these features was performed using XGBoost. The features of the classifier are as follows:

- Booster: Gbtree
- Max Depth: 10
- Learning Rate: 0.01
- No of Estimators: 200

The remaining parameters were set to the default value of the XGBoost classifier.

Doi :https://doi.org/10.54216/FPA.020104
Received: March 20, 2020 Revised: April 28, 2020 Accepted: May 30, 2020

3.2.3 Random Forest

Random forest is an ensemble learning method. This is a bagging algorithm in which multiple decision trees are made at a learning time. It prevents the overfitting of decision trees. It has low bias and very high variance. Random forest work by averaging values of different decision trees which are trained by using the same training set, fig. 3 shows the diagrammatic representation of Random Forest.

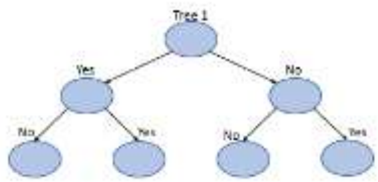


Figure 3: Random Forest

This classification algorithm was applied to the features extracted using the CNN architecture explained in 3.2.1, with the parameters listed below:

- Max Depth: 6
- No of estimators: 200
- Minimum Leaf Samples: 60
- Criterion: Gini

The remaining parameters were set to the default value of the Random Forest classifier.

4. Results

4.1.Accuracy and Training Time

Table III shows the accuracy and time taken to train the models.

Table 3: ACCURACY AND TRAINING TIME OF THE ALGORITHMS

ALGORITHM	ACCURACY		TRAINING TIME
CNN WITH SOFTMAX	TRAINING	93.11	2 hours
	TESTING	75	
CNN WITH XGBOOST	TRAINING	97.24	12 minutes
	TESTING	77	
CNN WITH RANDOM FOREST	TRAINING	93.30	16 seconds
	TESTING	74	

Doi :https://doi.org/10.54216/FPA.020104
Received: March 20, 2020 Revised: April 28, 2020 Accepted: May 30, 2020

4.2. Confusion Matrix

Tables IV, V, and VI represent the confusion matrix for CNN with Softmax, CNN with XGboost, and CNN with Random Forest, respectively.

Table 4: CONFUSION MATRIX FOR CNN WITH SOFTMAX

EMOTIONS	ANGRY(0)	DISGUST(1)	FEAR(2)	HAPPY(3)	S.A.D. (4)	SURPRISE(5)	NEUTRAL(6)
ANGRY(0)	1739	0	67	45	132	10	72
DISGUST(1)	0	1762	0	0	0	0	0
FEAR(2)	139	3	1516	34	213	40	80
HAPPY(3)	88	0	36	1402	70	35	127
SAD(4)	216	1	123	78	537	9	206
SURPRISE(5)	36	1	120	61	22	500	36
NEUTRAL(6)	170	1	72	100	238	6	691

Table 5: CONFUSION MATRIX FOR CNN WITH XGBOOST

EMOTIONS	ANGRY(0)	DISGUST(1)	FEAR(2)	HAPPY(3)	S.A.D. (4)	SURPRISE(5)	NEUTRAL(6)
ANGRY(0)	1735	0	59	40	155	15	61
DISGUST(1)	0	1762	0	0	0	0	0
FEAR(2)	62	2	1642	26	167	45	81
HAPPY(3)	58	0	41	1386	95	50	128
SAD(4)	144	0	149	68	569	17	196
SURPRISE(5)	21	0	90	53	31	551	30
NEUTRAL(6)	109	0	85	102	316	12	654

Table 6: CONFUSION MATRIX OF CNN WITH RANDOM FOREST

EMOTIONS	ANGRY(0)	DISGUST(1)	FEAR(2)	HAPPY(3)	S.A.D. (4)	SURPRISE(5)	NEUTRAL(6)

Doi :https://doi.org/10.54216/FPA.020104
Received: March 20, 2020 Revised: April 28, 2020 Accepted: May 30, 2020

ANGRY(0)	1641	0	60	31	247	17	69
DISGUST(1)	0	1762	0	0	0	0	0
FEAR(2)	87	3	1456	17	326	58	78
HAPPY(3)	56	0	22	1328	133	54	165
SAD(4)	128	1	104	57	669	12	199
SURPRISE(5)	21	0	78	42	56	548	31
NEUTRAL(6)	102	0	53	82	336	13	662

4.3. Performance Parameters

Tables VI, VII, and IX shows the precision, recall, and f1-score of all the seven classes for algorithms CNN with softmax, CNN with XGBoost, and CNN with Random Forest, respectively.

Table 7: PRECISION, RECALL AND F1-SCORE FOR ALGORITHM CNN WITH SOFTMAX

EMOTIONS	PRECISION	RECALL	F1-SCORE
ANGRY(0)	0.73	0.84	0.78
DISGUST(1)	1.00	1.00	1.00
FEAR(2)	0.78	0.75	0.77
HAPPY(3)	0.82	0.80	0.81
SAD(4)	0.44	0.46	0.45
SURPRISE(5)	0.83	0.64	0.73
NEUTRAL(6)	0.57	0.54	0.56

Table 8: PRECISION, RECALL, AND F1-SCORE FOR ALGORITHM CNN WITH XGBOOST

EMOTIONS	PRECISION	RECALL	F1-SCORE
ANGRY(0)	0.84	0.81	0.83
DISGUST(1)	1.00	1.00	1.00
FEAR(2)	0.81	0.79	0.80
HAPPY(3)	0.79	0.83	0.81
SAD(4)	0.51	0.44	0.47

Doi :https://doi.org/10.54216/FPA.020104
Received: March 20, 2020 Revised: April 28, 2020 Accepted: May 30, 2020

SURPRISE(5)	0.71	0.80	0.75
NEUTRAL(6)	0.51	0.57	0.54

Table 9: PRECISION, RECALL, AND F1-SCORE FOR ALGORITHM CNN WITH RANDOM FOREST

EMOTIONS	PRECISION	RECALL	F1-SCORE
ANGRY(0)	0.81	0.79	0.80
DISGUST(1)	1.00	1.00	1.00
FEAR(2)	0.82	0.72	0.77
HAPPY(3)	0.85	0.76	0.80
SAD(4)	0.37	0.57	0.45
SURPRISE(5)	0.78	0.71	0.74
NEUTRAL(6)	0.55	0.52	0.53

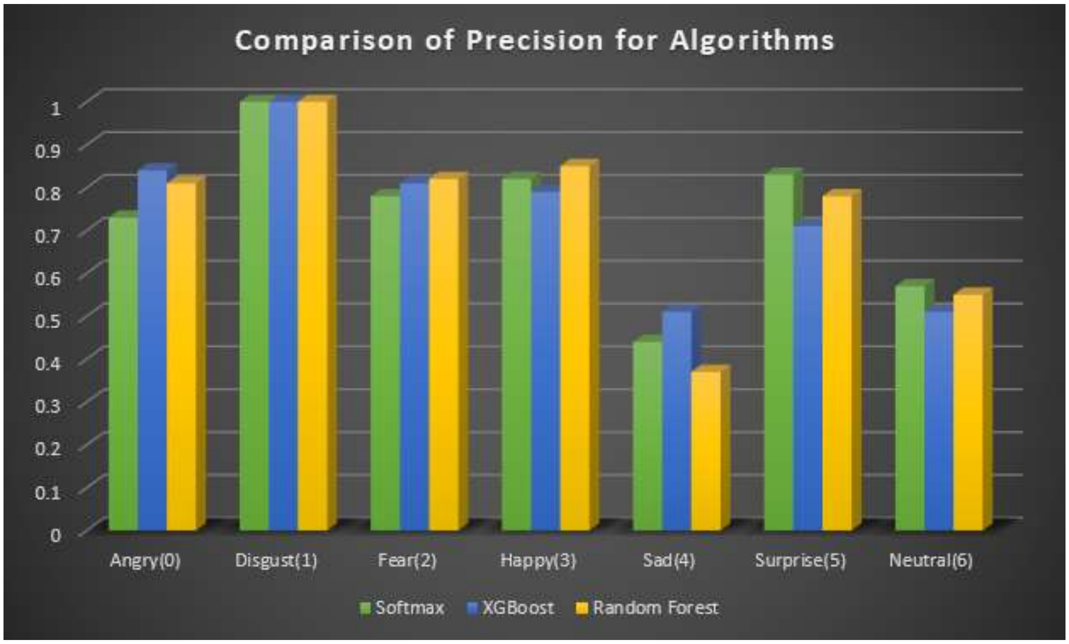


Figure. 4: Graphical Comparison of Precision for the proposed algorithms

Doi :https://doi.org/10.54216/FPA.020104
Received: March 20, 2020 Revised: April 28, 2020 Accepted: May 30, 2020

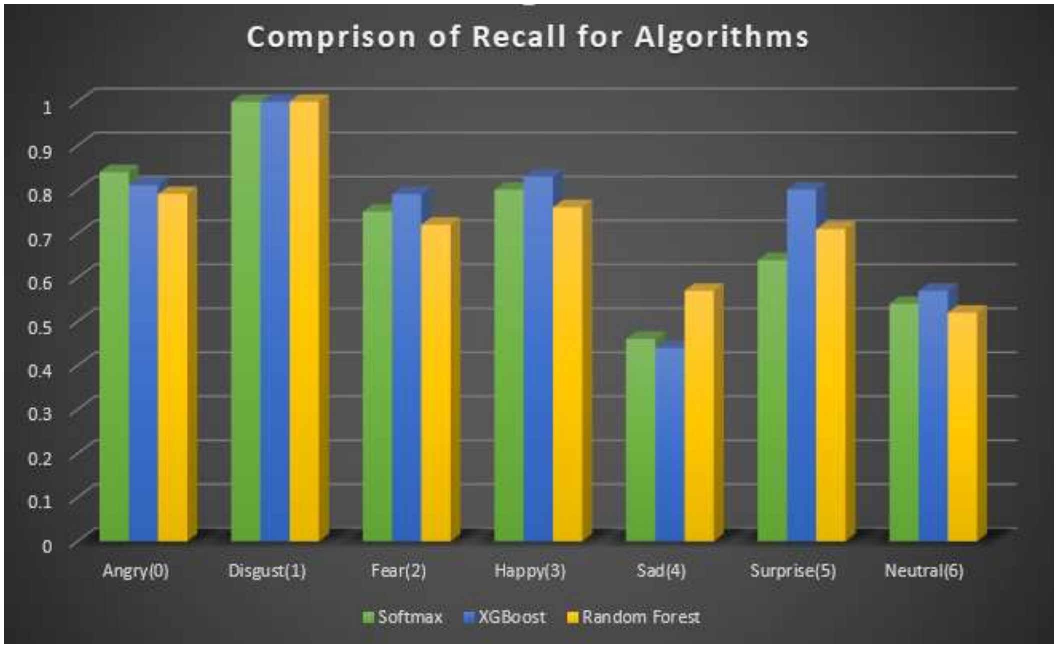


Figure 5: Graphical Comparison of Recall for the proposed algorithms

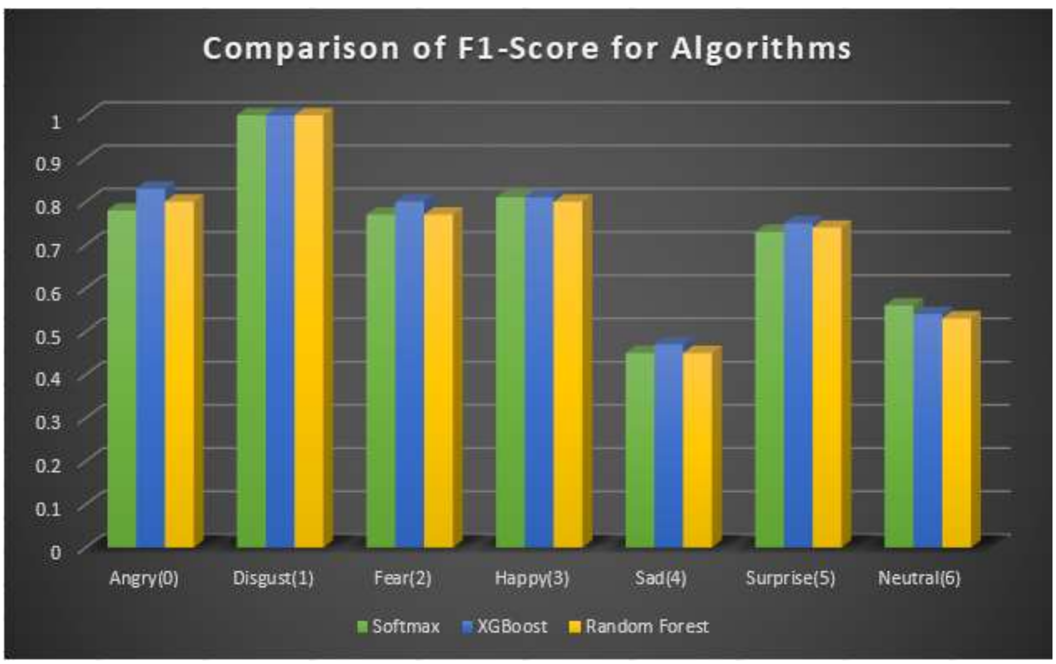


Figure 6: Graphical Comparison of F1-Score for the proposed algorithms

As per Table III under section 4.1, we found that the training and testing accuracies of the algorithms are almost similar, with a variation of 2 to 3 percent. The highest testing accuracy is 77% for the algorithm CNN with XGBoost, 75% for CNN with softmax, and 74% for CNN with Random Forest. According to this data, we believe that CNN with XGboost is the best algorithm. When training time is compared of the three algorithms, CNN with Random Forest turns out to be the fastest algorithm, followed by CNN with XGBoost and CNN with softmax, but at the same time, its accuracy in comparison to CNN with XGBoost is less by a remarkable 3%, therefore even though it has least computing time, CNN with XGBoost is the most favorable algorithm.

Doi :https://doi.org/10.54216/FPA.020104
Received: March 20, 2020 Revised: April 28, 2020 Accepted: May 30, 2020

The graph shown in fig. 4 shows a comparative analysis of precision between the different proposed algorithms for the seven emotions. Anger, disgust, fear, and sadness showed the best performance using CNN with XGBoost, but at the same time, CNN with Random Forest offers better precision in the classification of emotion as compared to CNN with Softmax. In all the mentioned emotions, CNN with Random Forest has a better precision or second-best than CNN with XGBoost. CNN with XGBoost performed best in four out of seven emotions.

The graph in fig. 5 shows a comparative analysis of recall between the different proposed algorithms for the seven emotions. Performance of CNN with XGBoost is best for the emotions fear, disgust, happy, surprise, and neutral, followed by CNN with Softmax, which is second best for most of the emotions and even performed best with fear, happy, neutral and thus performed best in only one of the seven emotions, at the same time CNN with Random Forest performed best in only one of the seven emotions.

The graph in fig. 6 shows a comparative analysis of F1-Score between the different proposed algorithms for the seven emotions. Performance of CNN with XGBoost is best for the emotions anger, fear, disgust, happiness, surprise, and sad, followed by CNN with Random Forest, which is second best for most of the emotions at the same time CNN with Softmax also showed similar performance, but CNN with XGBoost's F1-Score was highest in six out of seven emotions.

5. Conclusion

After comparing training and testing accuracy, we conclude the algorithm CNN with XGBoost performed best for the facial expression recognition dataset. Results of other performance parameters such as precision, recall, and f1-score also suggest that the proposed algorithm performed well for all the seven classes of emotions, angry, sad, disgust, happy, surprise, fear, and neutral. With this, we conclude CNN with XGBoost is the best algorithm, followed by CNN with Random Forest.

References

[1] S. Gilda, H. Zafar, C. Soni and K. Waghurdekar, "Smart music player integrating facial emotion recognition and music mood recommendation," *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai,*2017,pp.154-158, DOI: 10.1109/WiSPNET.2017.8299738.

[2] Yamashita, R., Nishio, M., Do, R.K.G., Convolutional neural networks: an overview and application in radiology. *Insights Imaging*, *Volume 9*, 2018, pp. 611–629 , DOI: https://doi.org/10.1007/s13244-018-0639-9

[3] ZixiangFei, Erfu Yang, David Day-Uei Li, StephenButler, Winifred Ijomah, XiaLi, Huiyu Zhoue, Deep convolution network based emotion analysis towards mental health care, *Neurocomputing, Volume 388,* 2020, pp. 212-227. DOI: https://doi.org/10.1016/j.neucom.2020.01.034

[4] Muhammad Naveed Riaz, Yao Shen, Muhammad Sohail, Minyi Guo, eXnet: An Efficient Approach for Emotion Recognition in the Wild, *Sensors (Basel), Volume 20,* 2020, pp. 1087-1091. DOI: 10.3390/s20041087

[5] Megha Chandran, Dr. Naveen S , A Review on Facial Expression Recognition using Deep Learning, *International Journal of Engineering Research and Technology, Volume 7*,2020, Issue 13. DOI: https://www.ijert.org/a-review-on-facial-expression-recognition-using-deep-learning

[6] Jie Shao, Yong sheng Qian, Three convolutional neural network models for facial expression recognition in the wild, *Neurocomputing, Volume 355*, 2019, pp. 82-92. DOI: https://doi.org/10.1016/j.neucom.2019.05.005

[7] Jason C.Hung, Kuan-Cheng Lin, Nian-Xiang Lai, Recognizing learning emotion based on convolutional neural networks and transfer learning, *Applied Soft Computing, Volume 84*, 2019, pp. 105724. DOI: https://doi.org/10.1016/j.asoc.2019.105724

Doi :https://doi.org/10.54216/FPA.020104
Received: March 20, 2020 Revised: April 28, 2020 Accepted: May 30, 2020