Geometric and Appearance approaches with Artificial Neural Network for Discrete Human Emotion Recognition from Static Face Images

Antonio Mpembe Franco

Faculty of Engineering, Environment & Computing

Coventry University

Coventry, England

mpembefraa@uni.coventry.ac.uk

***Abstract*—This paper focuses on fundamental emotions that theorists have researched to determine that can be deduced from facial expressions. Emotion detection is used in the therapy of autistic children and can be used for automatic sensing and flagging of depression-related behaviour. And this is done by experimenting with algorithms, that can extract features that aim to reflect the basic information of a face image by using the entire face as well as those that aim at a specific area of the face, and then use them to train and detect emotions with an artificial neural network.**

Keywords*—Class imbalance, classification, hyperparameter, convolution, cross-correlation, supervised, multilayer perceptron, distance, shape, texture.

I. Introduction

Humans' emotions are biological states associated with all nervous systems brought on by neurophysiological changes associated with various ways, thoughts, feelings, behavioural responses and a degree of pleasure or displeasure. These complex states and behaviours are triggered by an event that is either experienced or recalled.

The classification of emotion is a means of distinguishing or differentiating one emotion from another, which researchers have approached from one of two fundamental viewpoints. On the one hand, emotions are discrete and fundamentally different constructs. On the other hand, emotions can be characterized on a dimensional basis in groupings. The viewpoint for this paper is the discrete emotion theory, which states that all humans are thought to have an innate set of basic emotions that are cross-culturally recognizable. Theorists have conducted studies to determine which emotions are fundamental. A common example is a multicultural study by Paul Ekman and his colleagues in 1992, "*An Argument for Basic Emotions"* [1] in which they conclude that the six primary emotions are anger, disgust, fear, happiness, sadness, and surprise. Ekman explained that there are specific traits associated with each of these emotions, allowing them to be expressed in varying degrees. Each emotion acts as a discrete category rather than an individual emotional state [2].

The arousal of emotion is usually detected from the manifestation in our external appearances. But the manifestation may not exactly correspond to the arousal of a particular emotion. This paper focuses on emotions that can be analysed from the expressed face. Because there is a close relationship between emotion and facial expression, that was shown by Ekman, who took face images that were identified by westerners as explicitly expressing a few emotions and they also evaluated these to people in a variety of other cultures. And the fluctuation of the different facial images on emotions helps to understand how someone is feeling without nonverbal communication.

The human face is one of the most important representations of human emotions, intentions, attitudes, and moods. It is used to recognize other people, also used to judge various things such as age, gender, beauty, and others. Therefore, there is a surge of interest in the technology of emotional artificial intelligence, also known as affective computing.

Today, emotional artificial intelligence is used in the therapy of autistic children. The human face displays 10,000 different facial expressions, 7,000 of which are displayed daily. Developed people classify these expressions into maybe 15 to 20 categories. Autistic children lack this generalization ability, so for them, they see 7,00 different categories. This is the reason why they can not understand people's facial expressions. One of the problems is that when a human teaches a child these typical expressions, humans like and usually emphasize things. For an autistic child, a smile with raised eyebrows is a very different expression than just a smile.

In the future, affective computing may help to fight against depression in the case of automatic sensing and flagging behavioural cues related to depression. Depression is a big problem currently in the world, more noticeable in the western world. 25% of teenage girls, 17% of young adults aged 18 to 22 and 22% of people above 65 years suffer from this illness in the United Kingdom. [3].

II. Approaches

When a system classifies emotions, human facial expression recognition plays a significant role in the field of affective computing. But first, the extraction or detection of features is a necessary step which many computer vision algorithms do as the first step. A feature in image processing is a piece of information about the content of an image.

This paper will explore algorithms that aim to reflect the basic information of a face image by using the entire face and those that aim at a specific area in the face.

A. *Geometric-based approaches*

This approach focuses on the key points of facial regions that are used to represent a face image geometrically. By computing or extracting the distances, shapes, or textures from these points. The key points are detected in human face areas such as the mouth, jaw, eyes, brows, and so on.

In this paper, 3 types of signature-based human facial expression recognition are used together to identify human emotion recognition from distance-shape- texture-signal trio [4]. Statistical measurements from key facial points are fed into a neural network as features for training, and the results are then used to classify the six fundamental expressions and the neutral state. The key facial points can be detected from a statistical model of object shape and appearance [5].

B. *Appearance-based approaches*

Many appearance-based methods use a set of filters to extract facial features from an entire image. These techniques detect changes in the texture and shape of the face, such as the appearance and disappearance of wrinkles and other changes around key facial features.

There are three commonly extracted types of features: corners, blobs, and edges. This is referred to as geometric primitive features, and it includes those three elements as well as ridges, salient points-see, and image texture [6].

III. Related works

There are 2 main challenges to the geometric-based approach to facial expression feature extraction. Which are face recognition and key facial points de detection.

Like humans, an algorithm will have a difficult time detecting an object if its features are not clear and, depending on the conditions, this may be the case. As it is mentioned in the experiments section using one of the best python libraries, only 69% of the imagens facial keys were detected. Papers like "Face Recognition: Issues, Methods and Alternative Applications" [7] present a brief survey of issues methods and applications in the area of face recognition, such as facial recognition with artificial neural networks that takes a great number of training samples and even if with transfer learning these are still inaccurate in the same way that other statistically based methods are and others approach like Gabor wavelets and face descriptor-based methods are computationally intensive. For facial landmarks detection papers like "Facial Landmark Localization: past, present and future" [8] hat states that depending on the conditions, key facial points detection will won't as well as intended, conditions like resolution, occlusion, illumination and background clutter. And it also shows that landmark localizers trained in one database have inferior performance when tested on another database, making it harder for transfer learning to perform as well as intended. This was also shown by the paper "A Statistical Method for 2D Facial Landmarking" [9].

The geometric-based approaches have shown good results and most of the papers show this on similar databases where the conditions are good for the approach to succeed, as one of the most recent papers "Facial geometric feature extraction based emotional expression classification using machine learning algorithms" [10] hows the potential of geometric-based approaches with very high accuracies. The experiment section of this paper shows that, given different conditions, this approach struggles.

Appearance-based approaches are used in image classification in general. This approach is implemented in most state-of-the-art computer vision algorithms with the convolution kernel, such as VGG and MX Net GluonCV. The combination of geometric-based and appearance-based approaches features are mostly utilized in facial expression recognition. Some papers like "GEOMETRIC AND APPEARANCE FEATURE ANALYSIS FOR FACIAL EXPRESSION RECOGNITION" [11] explore local binary pattern variants combined with geometric-based approaches.

IV. Evaluation

The algorithms are evaluated on a laptop Xiaomi Mi Notebook Pro i5 2017.

Table 3
Computer unit specification

Processor	Storage	Memory
Intel Core i5-8250U	Samsung PM961 MZVLW256HEHP, 256 GB	7.89 GB, DDR4 2400MHz

* The computer contains other components these are the main components for the experiments.

- Confusion matrix

A confusion matrix, which is a cross-tabulation table with two dimensions displaying the actual class and the predicted class, will be presented [12].

Table 4
Confusion matrix description

		Actual class	
		Positive(P)	Negative(N)
Predicted class	Positive (P)	true positive (TP)	false positive (FP)
	Negative (N)	false negative (FN)	true negative (TN)

Note. TP eqv. with a hit, TN eqv. with correct rejection, FP eqv. with a false alarm, FN eqv. with a miss. All of this depends on which class is the focus

V. Exploratory data analysis

An image is a multidimensional matrix with width and height as well as depth depending on whether it is in a multiple colour channel. The images in this case are in a single colour channel, grayscale. Images used for facial expression recognition are typically static images or image sequences. The images in this paper are static [13].

Dataset obtained from Kaggle by a Research Prediction Competition on "*Challenges in Representation Learning: Facial Expression Recognition Challenge Learn facial expressions from an image*". Is known as Fer2013 is an open-source dataset that was created for an ongoing project by Pierre-Luc Carrier and Aaron Courville. This dataset consists of grayscale images of faces 48x48 pixels in size. The task is to categorise each face into one of seven categories based on the emotion shown in the facial expression. During the competition, 28.709 images and 3.589 images were shared with the participants as training and public test sets, respectively, while the remaining 3.589 images were kept as a private test set to determine the winner. After the competition, the dataset was made available to everyone, mainly for research purposes.

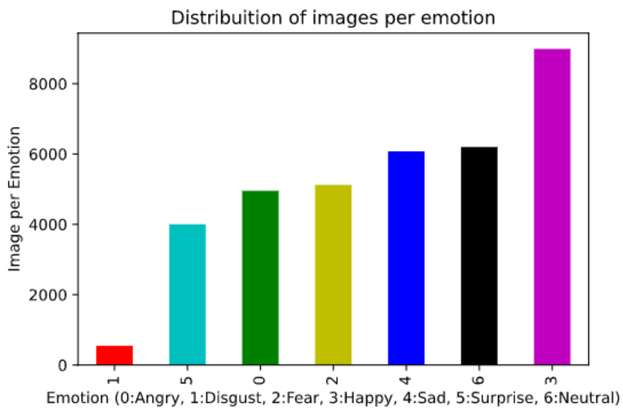

Fig. 1. Distribution of images per emotion

There are different types of emotion. The emotions are identified as happiness, sadness, disgust, fear, surprise, anger and neutral.

There are 547 images of disgust. This state is felt primarily in retaliation to the sense of taste, smell, and touch. It is normally characterised by a facial expression in which the upper lip is raised, the nose bridge is wrinkled, and the cheeks are raised.

There are 4002 images of surprise. This state is triggered by an unanticipated relevant event. It is normally characterised by a facial expression that causes someone's brows to arch, their eyes to widen, and their jaw to drop.

There are 4953 images of being angry. This is frequently associated with the fight or flight brain response to when a person feels threatened or in pain. Anger is normally characterised by a facial expression that causes someone to lower their brows, press their lips together firmly, and bulge their eyes.

There are 5121 images of fear. This is a basic survival mechanism that occurs in response to the presence of something traumatic. Fear normally causes a person to raise their brows, slightly open their mouth, and open their eyes wider than usual.

There are 6077 images related to sadness. Humans respond to sadness by becoming quiet, experiencing a lack of energy, and a desire to withdraw. Sadness is characterised by a facial expression in which the corners of the mouth are lowered, and the inner portion of the brows are raised.

There are 6198 images related to a neutral state. This is the lack of any emotional arousal. When the key facial elements are not expressive.

There are 8989 images of being happy. Happiness is associated with a mental state that reflects pleasure or joy. Happiness is characterised by a facial expression in which the corners of one's mouth are raised upwards.

A. Data Distribution

This is the general behaviour of data as it varies from instance to instance. There are numerous tools available for measuring the data's variability, including range, interquartile range, variance, standard deviation, and others [14]. The variance, for example, is the average of the difference squared between values and the mean, which is the average of the total sum of all values.

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n} \quad (1)$$

The arithmetic mean is the central value of a finite set of numbers in statistics. It is the sum of the values in a data set divided by the number of values. The population mean, also known as the expected value, is a measure of a probability distribution's central tendency. From equation (2) x_i is a data point in place i and n is the number of data points.

$$\overline{x} = \left(\frac{1}{n}\sum_{i=i}^{n} x_i\right) \quad (2)$$

The median is the value that separates the upper and lower halves of a data sample, population, or probability distribution in statistics and probability theory. It is sometimes referred to as "the middle" value in a data set. The median is crucial in robust statistics because it is the most resistant statistic. From equation (3) X is the ordered list of values in the data set and n is the number of values in the data set.

$$\mathrm{Med}(X) = \begin{cases} X\left[\frac{n}{2}\right] & if\ n\ is\ even \\ \frac{\left(X\left[\frac{n-1}{2}\right]+X\left[\frac{n+1}{2}\right]\right)}{2} & if\ n\ is\ odd \end{cases} \quad (3)$$

Table 3

The average variance, mean and median per class of emotion

Emotion	Variance	Mean	Median
Angry	3164.7	126.5	130.5
Disgust	2876.1	135.1	141.4
Fear	3187.2	135.4	141.2
Happy	2986.5	129.1	134.5
Sad	3070.3	121.0	123.3
Surprise	3217.7	146.3	154.4
Neutral	3164.8	123.9	127.2

According to Table 3, images associated with disgust have the least amount of variation on average in the intensity of the pixels, while those associated with surprise have the most variation on average in the intensity of the pixels. What can be deduced from the mean and median is that images associated with sadness have a more symmetrical distribution on average than other images, as shown in Fig. 2, Whereas red represents disgust, cyan represents surprise, green represents anger, yellow represents fear, blue represents sadness, black represents neutral, and pink represents happiness.

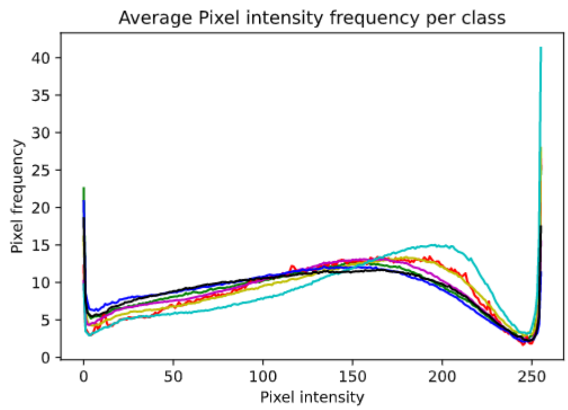

Fig. 2. Average distribution of pixel intensity frequency per emotion

B. Relation between features

It is important to understand the relationship between features. This knowledge helps to understand better the dataset, how each feature relates to the other, but in this case there is 2304 input data plus the output. This is also helpful for recognizing the most important features of the dataset. The use of covariance or correlation helps to observe the relation between the features. But in this case, having 2304 input and 7 output, it is a heavy task to compute 53408721 correlation values, not very useful.

VI. Experiments

Image classification has many challenges, such as intra-class variation, scale variation, view-point variation, occlusion, illumination, and background clutter [11]. Given the dataset used in the experiments, some of this will be a challenge, such as a scale variation, which refers to having multiple images of the same object in different sizes, zoomed in and zoomed out face images in the dataset, view-point variation, which means that an object can be oriented/rotated in multiple dimensions depending on how it is captured, and this dataset contains faces captured in many different ways, this dataset contains faces captured in a variety of orientations, occlusion, which means that a large portion of the object is hidden behind other objects; some faces in this dataset have a large portion of their face hidden, and illumination images captured in a variety of amounts and directions of light.

A. Geometric-based approaches

1) Feature extraction

The first step is to recognise the face, followed by the detection of facial key points. Face recognition libraries are used for this. When the face sites in an image are found, key points can be identified. Then there is a reduction of the key points as shown in Fig. 3. The statistical measures and the stability signature of the respective strategy are computed using these key points from these face sites: eyebrows, eyes, mouth, and nose, where the edge and middle key points are chosen for a global relation and specific ones.

Fig. 3. Landmarks of facial key point reduction

During the experiments almost 1/3, 10000 of 35000 images, faces and landmarks facial key points where are not recognised due to the occlusion.

The data set is unbalanced, but rotating, zooming, duplicating images, and other methods of image augmentation will not help in this case because the geometric-based approaches chosen for the experiment are all about the relationship between the key points, so image augmentation will not help.

A variety of features produce a promising recognition rate in facial expression. As a result, there is a need to compute statistical measures.

The range should indicate how widely distributed the values are. This is the difference between the maximum and the minimum number of the normalized signature.

The moment in statistics is a measure of something relative to the centre of the values. Moments are very important in statistics to characterize probability distributions. First, second, third and fourth moments characterize the central tendency, dispersion, asymmetry, and kurtosis, respectively [15]. Computed by the equation (4) where $v = (v1, v2,..., vn)$ is the normalised signature and $b(v(i), k)$ means raised to the power of k, where $k = 1, 2, 3, 4$.

$$l_k = \left(\sum_{i=1}^{n} b(v(i), k)\right) \times \frac{1}{n} \qquad (4)$$

Skewness is the asymmetry of the probability distribution of a normalized signature about its mean. Computed by the equation (5) where $v = (v1, v2,..., vn)$ is the normalized signature, $b(i, k)$ means raised to the power of an integer and $\bar{x}$ is the mean of the distribution.

$$u = \frac{\sum_{i=1}^{n} b(v(i) - \bar{x}, 3)}{\left(\sum_{i=1}^{n} b(v(i) - \bar{x},\ 2)\right)^{3/2}} \qquad (5)$$

Kurtosis is the measure of the thickness or heaviness of the tails of the distribution of the normalized signature. Computed by the equation (6) where $v = (v1, v2,..., vn)$ is the normalized signature, $b(i, k)$ means raised to the power of an integer and $\bar{x}$ is the mean of the distribution.

$$q = \frac{\sum_{i=1}^{n} b(v(i) - \bar{x}, 4)}{(\sum_{i=1}^{n} b(v(i) - \bar{x}, 2))^2} \tag{6}$$

Entropy is a measure of the unpredictability of a system's average infomation state. Informational value depends on the degree to which the content of the message is surprising [16]. Computed by the equation (7) where p_i is the probability of the element i that range from to the size of data point n.

$$En = -\sum_{i=1}^{n} p_i \times log \times p_i \tag{7}$$

a) Distance Signature

For this approach, every possible distance from the maximum of 25 key points is computed as shown in Fig. 4. Then the stability index and the other statical measurements from these distances are computed from the equation (4) between point (x_i, y_i) and (x_j, y_j).

$$D = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{8}$$

Fig. 4. Distance from each facial key point

b) Shape Signature

For this approach, every possible triangle is computed their signature from the maximum of 25 key points computed as shown in Fig. 5. Then the stability index and the other statical measurements from these shape signature computed from the equation (5) point h is the perimeter of the triangle *ABC* divided by 2 and *AB*, *BC* and *CA* are the distance between point *A*, *B*, and *C*.

$$S_i = \frac{\sqrt{h * (h - AB) * (h - BC) * (h - CA)}}{(h * 2)^2} \tag{9}$$

Fig. 5. Possible triangle from each facial key

c) Texture Signature

For this, a custom local binary pattern is used, which compares a pixel of an image with the neighbourhood as shown in Fig. 6 and treats the result as a binary number equation (10) where (p,r) denotes a neighbourhood of p sampling points on a circle with radius *r*, *gq* the neighbour and *gc* the mean pixel. In this case, the green cell in Table 4 is the key point pixel and the average of pixels "V" is compared to the average of "Ti". Where $T = (0,1,\dots 7)$. Where (p,r) denotes a neighbourhood of p sampling points on a circle with radius r.

$$LBPp,\mathrm{r}(\mathrm{x},\mathrm{y}) = \sum_{q=0}^{p-1} s(gq - gc)\, 2^q$$

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{10}$$

Table 4
Custom local binary pattern

T7	T6	T6	T6	T6	T6	T6	T6	T5
T0	V	V	V	V	V	V	V	T4
T0	V	V	V	V	V	V	V	T4
T0	V	V	V	V	V	V	V	T4
T0	V	V	V	V	V	V	V	T4
T0	V	V	V	V	V	V	V	T4
T0	V	V	V	V	V	V	V	T4
T0	V	V	V	V	V	V	V	T4
T1	T2	T2	T2	T2	T2	T2	T2	T3

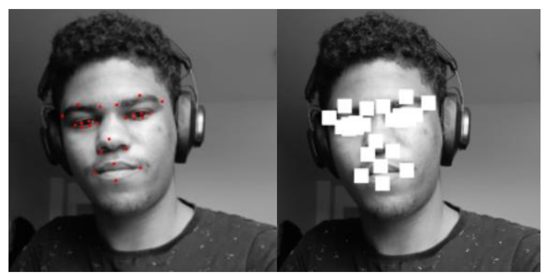

Fig. 6. Custom Local binary pattern representation

2) Bayesian optimized Multilayer perceptron

Bayesian optimization, like grid search, is another method for optimising hyperparameters. Saves time by avoiding a brute force approach. Based on the exploration and exploitation of concepts [17]. In this case, optimizing a multilayer perceptron (MLP), which is a type of artificial neural network that is feedforward, employs a supervised learning technique known as backpropagation, with at least three-node layers: an input layer, a hidden layer, and an output layer. In this case 1 input layer with 27 features, 3 hidden layers with 512, 64 and 16 nodes respectively and 1 output layer with 7 possible outcomes. Each node, except for the input nodes, is a neuron with a nonlinear activation function.

For this model techniques to avoid overfitting such as weight regularization, which penalizes the parameters, dropout, that drops a node, and 3-fold cross-validation. A 10 iteration of 25 parallel *n_points* computations Bayesian optimization for epochs ranging from 16 to 64, batches ranging from 8 to 256, weights, dropout ranging from 0 to 0.9, and a group of optimizers and activations and loss functions *"rmsprop", "adam", 'sgd', 'Ftrl' ; 'relu', 'sigmoid', 'tanh';*

'categorical_crossentropy', 'categorical_hinge', 'mean_squared_logarithmic_error'.

a) Results

The best model had *'adam', 'tanh'* and *'mean_squared_logarithmic_error'* as optimizer, activation, and loss function. 132 for batch size, 35 for epochs, 0.35000000000000003 for dropout and 0.0 for weight. 37% of accuracy for training data and 36% for testing data that can be deduced from Fig.7.

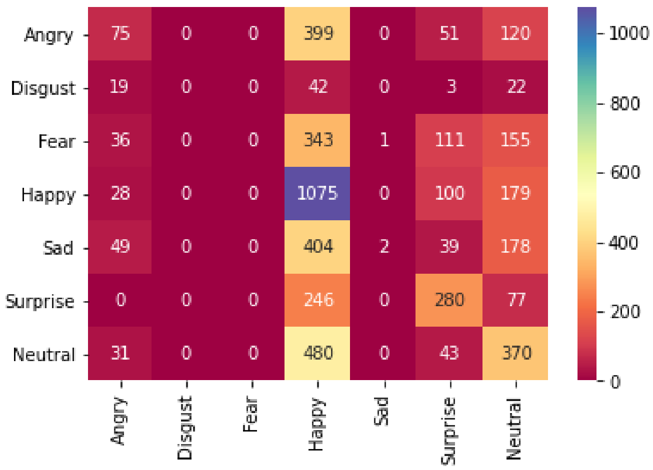

Fig. 7. MLP confusion matrix

Appearance-based approaches

3) Feature extraction

Most of these filters perform image convolution equation (12) or cross-correlation equation (11), where their input image I and two-dimensional kernel K. In this case, the multiplication of an image segment with a filter is followed by a sum. Convolution and cross-correlation are somewhat similar; the difference is that in convolution, the inputs are flipped or reversed. This is significant because an algorithm can be referred to as convolutional while employing the cross-correlation operator [18].

$$S(i,j) = (I * K)(i,j) = \sum m \sum n\, K(i - m, j - n) I(m,n) \tag{11}$$

$$S(i,j) = (I * K)(i,j) = \sum m \sum n\, K(i + m, j + n) I(m,n) \tag{12}$$

a) An example of feature estraction(Edge detection)

In computer vision, using a low-pass filter followed by a high-pass filter works well for detecting edges. This is still a subjective method of detecting edges because edges are abrupt changes in pixel intensity, and depending on the threshold, more or fewer edges will be detected. Canny edge detection is one of the more accurate edge detection algorithms, employing a variety of filtering techniques on the image. As shown in Fig. 9, a low-pass filter Gaussian blur is used to filter out noise, followed by a high-pass filter Sobel x and Sobel y to determine the strength and direction of the edges.

Fig. 8. Canny Edge detection

There is no need to investigate additional filters that can be applied to an image to extract features because the convolution kernel automatically extracts the required features in this paper.

4) Convolution neural network

Convolutional neural networks are a type of artificial neural network that uses automated learning to optimise its filters. They are made up of three layers: an input layer, hidden layers, and an output layer. In this case input layer has the size of 48 by 48, with 6 hidden layers of 64, 128, 512, 512, 256, 512 nodes respectively followed by 2 fully connected layer with 256 and 512 nodes respectively and with 1 output layer of 7 nodes. They are made up of various layers, such as the convolution layer, activation layer, pooling layer, and others. The most important layer, however, is the convolution layer, which is where convolution operations are computed.

a) Results

As shown in Fig. 9, the training data had an accuracy of 72.5 percent, the testing data had an accuracy of 66.6 percent with augmenting on the testing, and the training set had an accuracy of 80.4 percent and the testing set had an accuracy of 68.1 percent when the two features were combined as shown in Fig.10. It may be a good sign that, with augmentation, the training set's accuracy is closer to that of the testing set. However, this is simply a question of when to stop the model. At some point, the training set accuracy began to differ significantly from the testing set accuracy for both models.

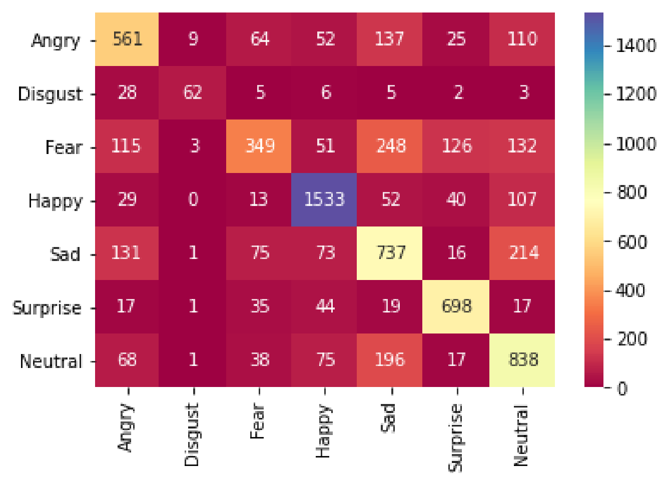

Fig. 9. CNN with augmentation confusion matrix

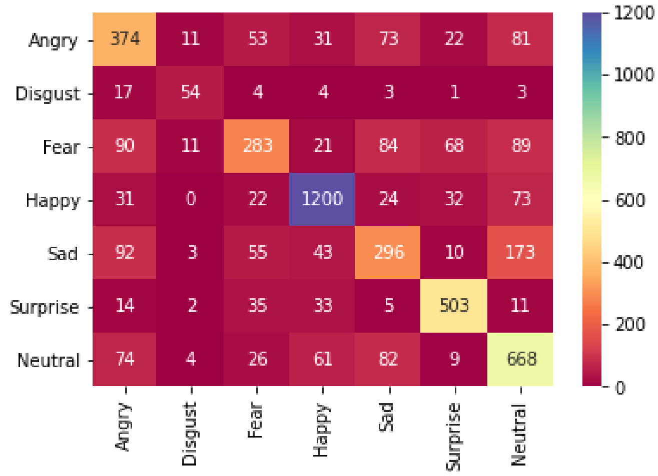

Fig. 10. Combining the two features confusion matrix

VII. Conclusion

Different approaches are stated in this paper, to extract features. As it is shown in this paper, the combination of the 2 features gets better performance for this data set. It would be naive to claim that appearance-based approaches are overall better, geometric-based approaches have very good performance depending on the dataset, as it is stated on the related works. But this demonstrates that, given these obstacles, a deep learning approach with convolution layers would work better for this dataset. The winning accuracy on Kaggle has an average of 69-70% with VGG. There are self-proclaimed papers with an accuracy of between 70-76% VGG, Resnet. Finding better features and better ways to compute them will increase the accuracy. the Exploring different feature extraction methods is critical for developing better machine learning models because if the features do not accurately represent the object, the model will fail regardless of how deep or powerful the machine is. For future works related to this subject this would something to explore.

.

.

VIII. References

[1] P. Ekman, "An Argument for Basic Emotions," in *Cognition and Emotion: Vol 6*, Lawrence Erlbaum Associates Limited, 1992, pp. 169-200.

[2] "Emotion classification," 9 March 2021. [Online]. Available: https://en.wikipedia.org/wiki/Emotion_classification. [Accessed 15 March 2021].

[3] M. Pantic, "This is why emotional artificial intelligence matters," TEDxCERN, 14 December 2018. [Online]. Available: https://www.youtube.com/watch?v=4QjZDUaDxQU. [Accessed 19 March 2021].

[4] P. D. a. A. Barman, "Human Emotion Recognition from Distance-Shape-Texture Signature Trio," in *Human Emotion Recognition from Face Images*, Singapore, Springer Nature Singapore Pte Ltd. 2020, 2020, p. 235.

[5] "Active appearance model," 17 December 2020. [Online]. Available: https://en.wikipedia.org/wiki/Active_appearance_model#:~:text=An%20active%20appearance%20model%20(AAM)%20is%20a%20computer,the%20images,%20is%20provided%20to%20the%20training%20supervisor.. [Accessed 21 MArch 2021].

[6] "Geometric feature learning," 14 May 2019. [Online]. Available: https://en.wikipedia.org/wiki/Geometric_feature_learning. [Accessed 25 March 2021].

[7] K. G. a. M. J. Waldemar Wójcik, "Face Recognition: Issues, Methods and Alternative Applications," 2016.

[8] M. O. a. B. A. Hamid OUANAN, "Facial Landmark Localization: past,future," 2016.

[9] A. A. S. a. T. G. Hamdi Dibeklioglu, "A Statistical Method for 2D Facial Landmarking," 2014.

[10] M. Murugappan, "Facial geometric feature extraction based," 2021.

[11] P. S. 2. Sonu Dhall, "GEOMETRIC AND APPEARANCE FEATURE ANALYSIS FOR FACIAL EXPRESSION RECOGNITION".

[12] "Confusion matrix," 27 September 2020. [Online]. Available: https://en.wikipedia.org/wiki/Confusion_matrix.

[13] P. D. a. A. Barman, "Human Emotion Recognition from Face Images," in Types of Emotions, Singapore , Springer Nature Singapore Pte Ltd. 2020, 2020, p. 3.

[14] J. Frost, "Statistics By Jim," 13 August 2020. [Online]. Available: https://statisticsbyjim.com/basics/variability-range-interquartile-variance-standard-deviation/.

[15] "Momento (estatística)," 7 March 2019. [Online]. Available: https://pt.wikipedia.org/wiki/Momento_(estat%C3%ADstica). [Accessed 25 March 2021].

[16] "Entropy (information_theory)," 22 March 2021. [Online]. Available: https://en.wikipedia.org/wiki/Entropy_(information_theory). [Accessed 30 April 2021].

[17] "Hyperparameter optimization," 17 November 2020. [Online]. Available:

https://en.wikipedia.org/wiki/Hyperparameter_optimization.

[18] Udacity, “Convolution vs Cross Correlation,” 26 Febuary 2015. [Online]. Available: https://www.youtube.com/watch?v=C3EEy8adxvc&t=2s. [Accessed 29 MArch 2021].

[19] S. Prabhakaran, “Mahalanobis Distance – Understanding the math with examples (python),” [Online]. Available: https://www.machinelearningplus.com/statistics/mahalanobis-distance/.

IX. Annex

Codes*: https://github.com/tonyamf/ANN*