

Facial Expression Recognition, (FER 2013) - Using Convolutional Neural Network (CNN) and Deep Learning.

ABSTRACT

As we see in the same manner, that of clinical practices and behavioural description are important for mortal-computer commerce, facial express recognition(FER) is also important. Due to the vastness of the mortal facials and dissimilarities with the reference to photographs, which are called like, making changes in the mask of facial and lightening it up a little accordingly, strengthen FER using virtual reality becomes stimulating. Because to its significant unwilling point birth and calculational efficacy, deep pedantry clones, just particularly, Convolutional Neural Networks (CNNs), have attested remarkable potential among all FER methods as of now. We take the FerNet armature on loan, so that we can precisely do adjust its tuning parameters, and workout with vibrant ameliorate approaches. By such an extent that we are aware the idiom, our model uses dispensable training data to reach up-to-the-minute single network delicacy of 88.29% on FER2013.

INDEX TERMS

Facial expression recognition, emotion recognition, face recognition, joint learning, transfer learning ,Artificial intelligence, Convolutional Neural Network (CNN), and FER-2013 Dataset of Machine learning.

INTRODUCTION

Facial expressions that pass on the basic emotions like fear, happiness, sickness, etc. and these are referred to be as facial emotion expressions. It has a crucial role in human-computer interactions and have high chances to be used in client feedback assessment, online computer gaming, digital set-frothing , and health care sectors[1]-[3]. This conspiracy has been resolved because to developments in computer vision and sciences, which have enabled highly intricate emotion placing in photographs taken in controlled settings and peaceful environments [4]. Due to substantial intra-class differentiable and minutest, littlest inter-class variation, such as changes in facial personate and small distinctions across expressions, difficulties and quite good challenges

with emotion distinguishing under natural situations still take place.

Compute creativity research is always directing to build on bracket infirmity on interconnected and affiliated issues [5]–[7]. Because to its operational efficiency and point-birth prowess , Convolutional Neural Networks (CNNs) have showed tremendous considerable potential in the image bracket [8]. Those are the fundamental models for FER[5]-[7],[9]-[11] that are utilised the most far more often.

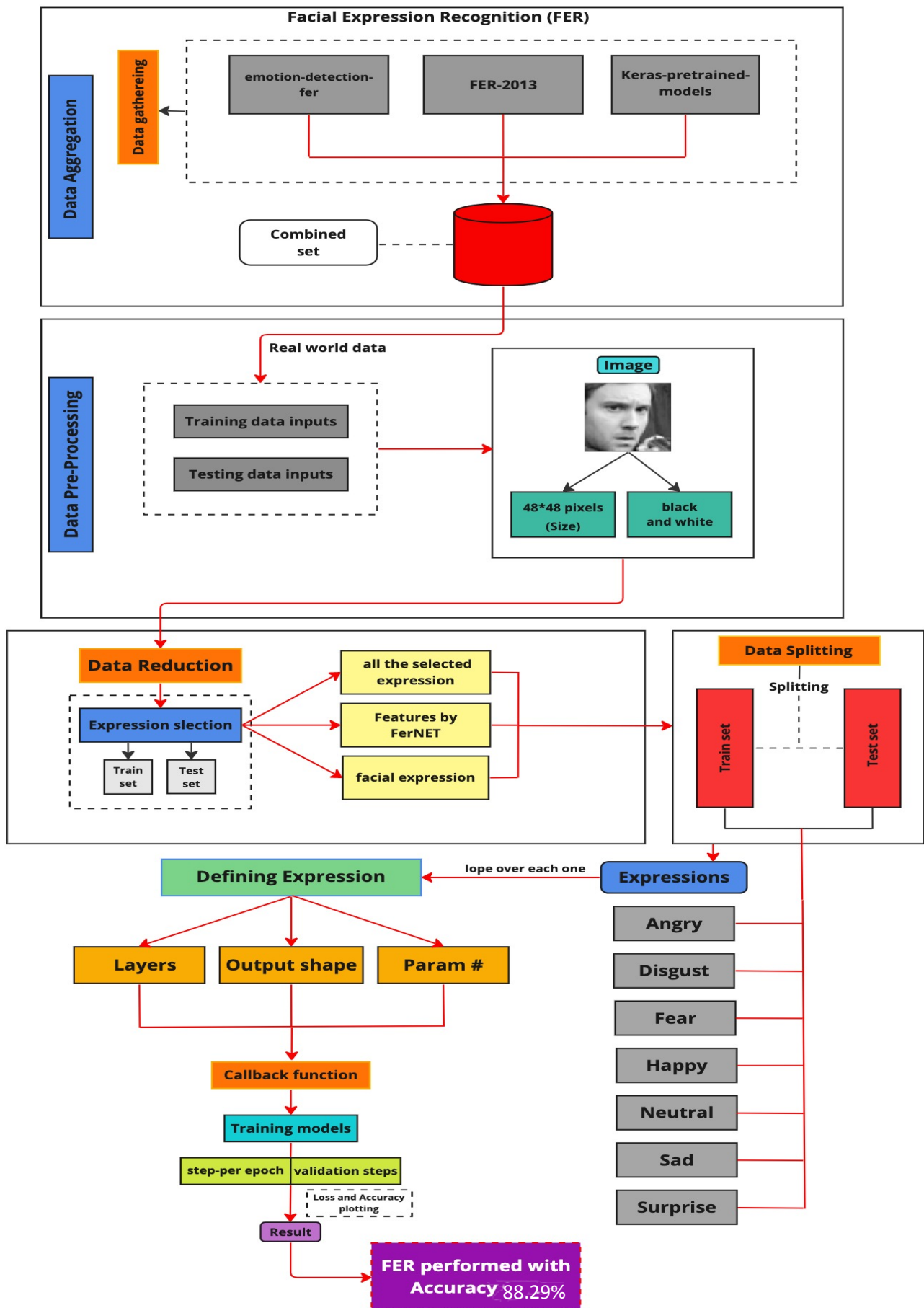
FER2013 is one specialized emotion recognition based on deep learning dataset that takes into consideration the numerous hurdles and fragile settings. It became a norm for measuring predictive accuracy in emotion assessment after being known in 2013 at the International Conference on Machine Learning (ICML). With this dataset, the highest incidence is anticipated to be about 65.5 percent [12]. We rigorously consider early study, received training and guesstimated on this dataset when comparing multiple techniques and benchmarking our research results.

The gains in related industries, such as neuroimaging, vital moving evaluation, feature detection, face masking, and authentication, enabling the inspection of bus-mantic facial expressions. dataset.

Multiple aspects, along with cognitive therapy, psychiatrist, neurophysiology, pain assessment, lying identification, intelligent ecosystems, and bidirectional earthly virtual keyboard, can benefit from natural-language facial expression analysis (HCI).

In this investigation, we adopt CNNs to strengthen the exquisite vaticination on FER2013. To analyse various simulation techniques and the learning of rate schedulers, we leverage the power of the FER Network and build colourful samples practices. With a testing specificity of 88.29%, we totally modify the model and training model parameters to produce cutting-edge outcomes. This is the largest single-network delicacy on FER2013 that has been managed to accomplish without the use

Fig.() Working diagram of proposed model.

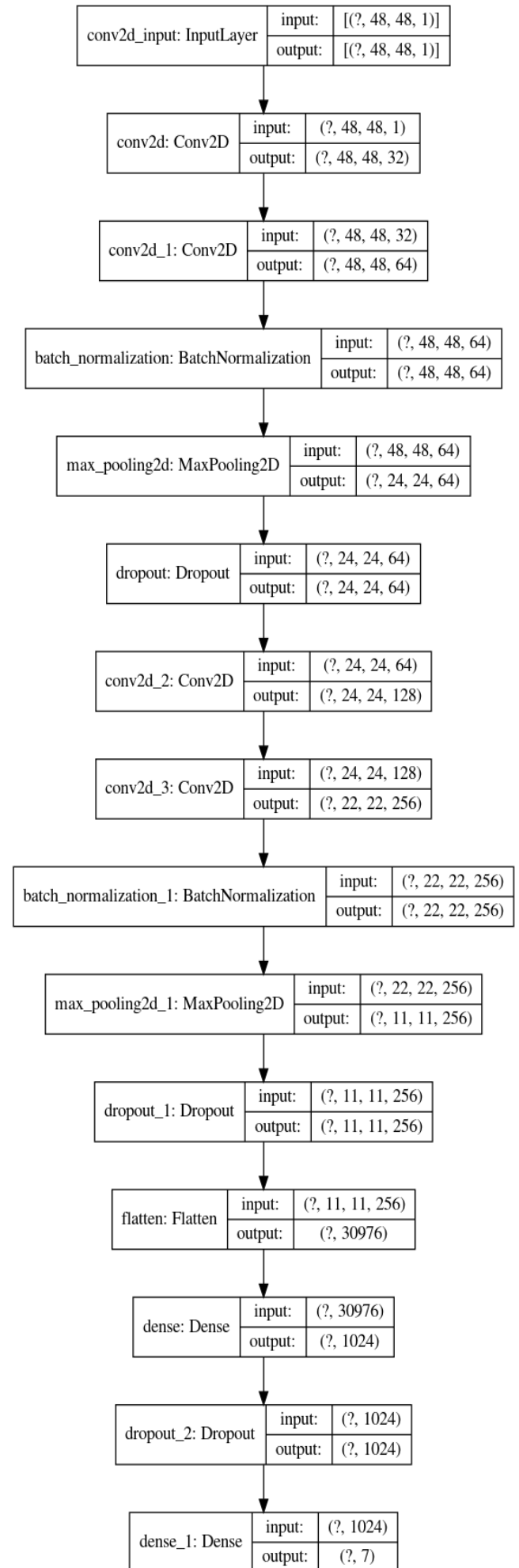


of superfluous training set, as far as we are aware. In hopes of better comprehending the network's efficiency and discretion In addition, we provide many characteristic mappings for the call action.

RELATED WORK

CNN have demonstrated promising importance in image recognition since their debut in the latter 1990s [13]. A convolutional layer, a pooling layer, and a fully connected layer are common components of a CNN subject. This makes it a n efficient for adjusting static images. Nonetheless and, the insufficiency of training sample and computational capacity at the time prevented hoe to apply CNNs. CNNs become a substantially more experimental and theoretical studies for point birth and picture anchoring after the 2010s because of breakthroughs in computational technology and the amassing of massive data sets[8].

With the aforementioned advancements and the substantial study on CNNs, they have emerged as a very convenient resource in image interpretation, machine learning based, and background subtraction applications. Once FER2013, a substantial dataset of facial expressions, was presented at ICML in 213, it promptly created a framework for gauging the efficacy of various emotion detection models. Already many CNN variations have shown provided considerable, with prediction performance running the gamut from 65% to 72.7% [26]–[34]. At being particular, to create competitive advantages, the three main CNNs underwent instruction by Liu et al. before being included. The accuracy of their single most significant network is 62.44% [26]. An accuracy of 70.02% was made possible by Minaee et al which can use an effortful convolutional network in an earlier part deep learning machine learning system [27]. A supporting vector machine had been employed in sight of the SoftMax layer in a deep neural network constructed by Tang et al, who attained a classification accuracy of 71.2% [32]. Shi et al.'s innovative Amend Representation Module (ARM) has been successfully presented to replace the layer used for pooling and achieve a testing accuracy of 71.38%. [33]. The accuracy and productivity of the three distinct constructions, VGG, Inception, and ResNet, and compared their results, are evaluated Pramerdorfer et al. However according to their findings, VGG trumps ResNet



and Inception with a high precision of 72.7%, 72.4%, and 71.6%, respectively [34]. It has been established that the functionality may be enhanced by combining several models. As an artist's expression, A renowned professor named Liu et al. together 3 CNN's and observed a certain 2.5% betterment in the performance [34]. Never the less, we first attempted to optimise the bedrock of these groups, a single network in order to enhance or emerge ensemble representation even more. Never the less, that is outside the focus of this documentation. Research paper that have been published their research papers, even they tried to improve the uniqueness and grouping in FER2013 by integrating supplementary training data.

III. Our Approach

We will describe our suggested system in detail in this whole section.

3.1 Our Network Framework Overview

In the above working we've separated the orders with Data aggression, Data pre-processing, Selection, Expressions and so on ending up with results. The following way describe the procedures of the working model :

Step 1. Data gathering is done and also real data from the combined data is transferred.

Step 2. also the real data is passed to Data-Pre-processing where data is converted from abnormal class to one class.

Step 3. After pre-processing, data reduction is done and further transferred to point selection.

Step 4. On passing through Data Split we'll initialize each model and store it by name in a wordbook.

Step 5. Finally we achieve results with delicacy.

In the above flowchart shown, we see that one image is of 48 * 48 pixels and it has the conv2d layers for input, also the other subcaste is Added Conv2D . Following with conv2d_1Conv2D . Then the batch normalization is done of the inputs we've taken. also we move to drop outting the inputs and repeating doing it, also flattening, then to get the asked thick.

Also, here as illustrated in Figure 4, the Deep Learning model's framework consists of an input segmentation subcaste, three or four added on convolutional nets (CNN) layers, and a completely integrated set of every (48*48,1) to seven classes. A standardization, pooling, and powerhouse layer with numerous supply sections comes after every one of the three CNNs.

If we only suppose, for the sake of demonstration, that the machines are developed using various picture divisions in the Training, Confirmation, and Testing sets. The accompanying figure displays 300 photos that were generated for training, evaluated on 200 more photographs, and then stored for use in the testing phase. The exam consisted of 120 photos, and the mean result was 80%. Moreover, a categorized intra- and inter loss function was used in this study. In the table below, the DCNN architecture is illustrated.

3.2 Implementation

I. Adding different libraries,

The developed model was generated in Python using the basic libraries Panda, keras, etc. in the FER-2013 Notebook.

II. Dataset,

When utilising different strategies to obtain reliable outcomes, data is thought to be the initial and most fundamental component. So, we took datasets from the FER-2013 the most famous and standard dataset.

The 2013 FER Dataset, the data consists of Grayscale pictures of (48 * 48) pixel faces. Every head typically covers the very same proportion of space for each picture and is roughly located in the centre thanks to the faces' automated registration. So every face must be categorised into one of seven categories, with 0 denoting anger, 1 denoting nausea, 2 denoting fear, 3 denoting happiness, 4 denoting sadness, 5 denoting surprise, and 6 denoting neutrality. Has two columns: "pixels" and "emotion". For each image, a string encircled in quote marks is present in the "pixels" column. The values in this string are pixel values separated by spaces and are listed in row main order. Your job is to forecast the emotion column in test.csv, which only has the "pixels" column.

Facial expression classifications ,facial recognitions system, and extraction of features, are the three steps in the categorization of the person facial expressions. The authors throughout this analysis used a framework that classifies seven is

Layer (type)	Output figure	Parameter #
Rescaling (Rescaling)	(None, 48, 48, 1)	0
Sequential (Sequential)	(None, 48, 48, 1)	0
conv2d_7 (Conv2D)	(None, 46, 46, 32)	320
conv2d_8 (Conv2D)	(None, 44, 44, 64)	18496
max_pooling2d_4 (MaximumPooling2)	(None, 22, 22, 64)	0
dropout_3 (Dropout)	(None, 22, 22, 64)	0
conv2d_9 (Conv2D)	(None, 20, 20, 64)	36928
conv2d_10 (Conv2D)	(None, 18, 18, 64)	36928
conv2d_11 (Conv2D)	(None, 18, 18, 128)	73856
max_pooling2d_5 (MaximumPooling2)	(None, 8, 8, 128)	0
conv2d_12 (Conv2D)	(None, 6, 6, 128)	147584
conv2d_13 (Conv2D)	(None, 4, 4, 256)	295168
max_pooling2d_6 (MaxPooling2)	(None, 2, 2, 256)	0
max_pooling2d_7 (MaxPooling2)	(None, 1, 1, 256)	0
dropout_4 (Dropout)	(None, 1, 1, 256)	0
flatten_1 (Flatten)	(None, 256)	0
dense_2 (Dense)	(None, 1024)	263168
dropout_5 (Dropout)	(None, 1024)	0
dense_3 (Dense)	(None, 7)	7175
In All parameters:		879,623
Adaptable parameters:		879,623
Non-Adaptable parameters:		0

to seven fundamental human expressions at the macro level:

1)Happy

In a facial expression, a smug or small smile that can convey enjoyment or like for something. To process a smug or grin smile is the cheerful expression by the upper movement of the cheek muscles part and the sides or tips of the lips.

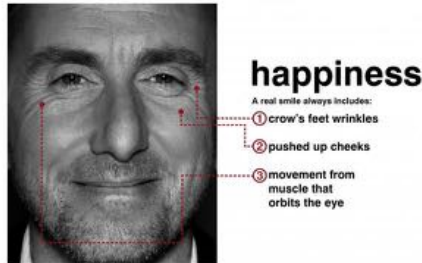


Fig. 2:Expression: Happy face (reference: [14])

2)Anger

Anger-indignated facial expression arrives when the demands or the wants and reality doesn't agree. We can showcase this expression when showing the eyes are keen when staring, also by narrowing the lips, and also when our eyebrows come closer and shrinks to each other and bend down .

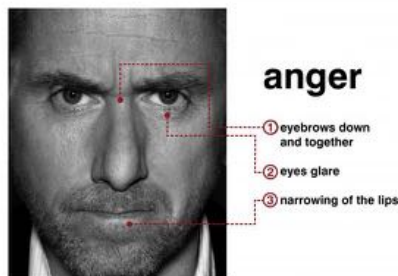


Fig. 3: Expression: Anger face (reference: [14])

3)Sadness

Here in this expression, we certainly get disappointed or a sense of missing something certain, a face and dull face will arise. Determined by the features of a sad face expression, which includes a loss of concentration to the eye, a lowering pull of the lips, and a dropping of the upper eyelid.



Fig. 4: Expression: sad Face(reference: [14])

4)Fear

Fear is a type of expression that manifests when a person finds themselves unable to handle a situation or in a frightening environment. The two eyebrows that raise simultaneously, the tightened eyelids, and the parallelly to the group, wide lips all indicate anxiety on a person's face.



Fig. 5: Expression: Fear face (reference : [14])

5)Disgust

This expression is seen when a human being sees something cringe or unusual or after hearing to something that is irrelevant. To observe this expression, humans make signs of distaste when the upper lip raises and shrink appear at the nasal bridge.

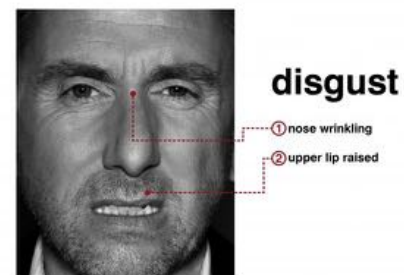


Fig. 6: Expression : Disgust face (source: [14])

6) Surprise

Here, Someone somewhere will show surprise when they learn of an abrupt, surprising, or important incident or message that they were not cognizant of beforehand. The raised brows, open eyes, and reflexive opening of the mouth signify a stunned look

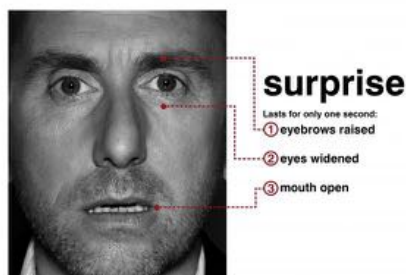


Fig. 7: Expression: surprise face (reference: [14])

7) Contempt

A man who is seen as arrogant and disrespectful typically undervalues other people based on their facial expression. The emotion can be seen by a movement that raises the oral edge.

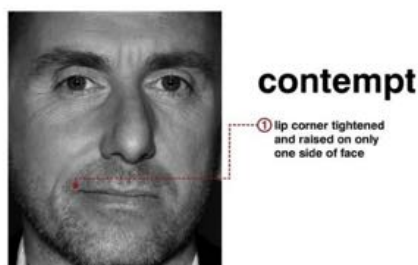
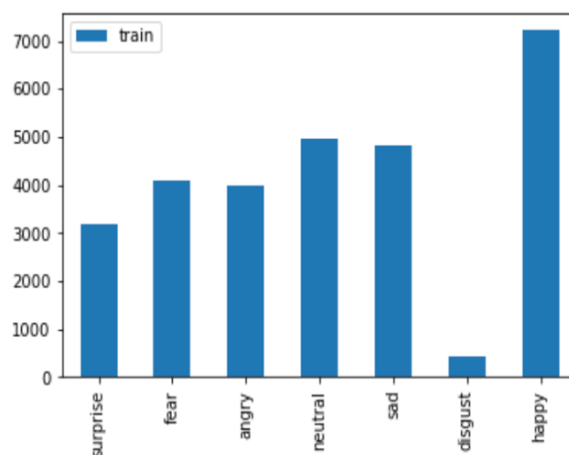


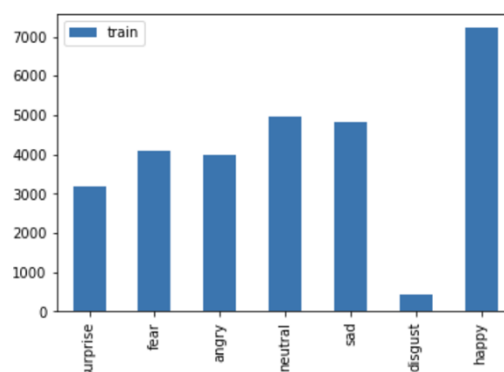
Fig. 8: Expression: contempt face (reference: [14])

Still, this public test set contains 28,709 exemplifications. The open testing dataset that was used to make the leader board has 3,589 exemplifications. In the ultimate testing dataset, which was used to select the contest winner, there are an additional 3,589 exemplifications.



n.

Pierre- Luc Carrier and Aaron Courville sat together and prepared this dataset which, as a component of a current exploration system plan. A primary interpretation of these renowned personal's dataset which courteously handed the factory organizers with to use for this contest.



With all due respect, when delivering FER2013 training, we follow the completion of authorised training, confirming, and test sets as provided by the ICML. The FER2013 collection include 35888 photos representing the following 7 emotions: anger, neutral, sickness, fear, pleasure, sorrow, and surprise. The fatal delicacy on this dataset, according to a Kaggle forum debate hosted by the competition organisers, is between 65 and 68%.

We employ a significant amount of data supplementation in training to take the variability in face emotion identification into account. With this feature, the photographs are moved horizontally and vertically by over 20 times, scaled up to 20 times their original size their original distance, and rotated up to 10 degrees.

III. Pre-processing

We employ a large amount of data augmentation in training to account for the heterogeneity in face emotion identification. They are required for this procedure in order to rectify any posture and lighting variations that may have resulted during the real-time facial detection phase. In this study, test and training photos were pre-processed using the Keras library before being fed into the Deep Convolution Neural Network (DCNN). Stretching the discovered photos and trimming the faces that were recognised are also steps in this procedure. All picture target sizes were shrunk to a dimension of 48x48x1, transformed to a grayscale colour channel, and pre-processed in a batch size of 120 for the purpose of data cleaning. Additionally, square boxes were employed for applying lighting adjustment while detecting face landmarks.

IV. Theoretical Background.

In this theoretical Framework, the emotions are classified into categories in the FER issue, including sorrow, surprise, fear, happy, rage, disgust, and neutrality [32]. Using the gathered characteristics, it develops a classification algorithm. In the older days, researchers have used classification models such as Artificial Neural Network (ANN), Support Vector Machine (SVM), Hidden Markov Model (HMM), K-Nearest Neighbour, and Support Vector Machine (SVM), (KNN), to recognise facial emotions. Following are explanations of various model sections that can make up that constitute an model for learning various degrees of abstraction and representations. These certain components include ReLU activation function, dropout, category cross-entropy loss, Adam optimizer, convolutional layers, and SoftMax activation function for seven emotion classifications, in the output layer. The modifications to these components via hyper - parameter tinkering that increased the competency and precision of the mixed system will also be addressed in this chapter.

Convolutional Layer: The Conv layer is the core element of a convolutional neural network (CNN) and is in charge of the bulk of the processing.

The convolutional layer displays convolution for a given set of inputs [32] utilising a $n * m$ sized filter f_k with a kernel and applied to an input x .

There are $n*m$ input connectivity. The equation may be used to compute the result (1).

$$C(X_{u,v}) = \sum_{i=n/2}^{n/2} \sum_{j=-n/2}^{n/2} f_k(i,j) x_{(u-i,v-j)} \quad (1)$$

Max Pooling Layer: Saravanan et al. [33] used the max function to achieve this. The input is significantly decreased as a result. Let x_i be the input and m be the filtered size. Depiction of how this layer is implemented, and the result may be estimated using the formulas in (2).

$$M(x_i) = \max (r_i+k, i+l / |k| \leq m/2, |l| < m/2, k, l \in N). \quad (2)$$

Rectified Linear Unit (ReLU) Activation Function: It determines the input to the network or neuron, the output for a certain value of p , as shown in equation (3). The fact that it does not need the calculation of an exponential distribution and does not have diminishing contour issue that is why ReLU was used in this study. Figure 2 below shows how the concatenated convolutional layers are evaluated concurrently using ReLU activation functions, upscaling accuracy and perfectly extracting facial information from images.

$$f(x) = \begin{cases} 0, & \text{for } x < 0 \\ 1, & \text{for } x \geq 0. \end{cases} \quad (3)$$

Fully Connected Layer: As it changes every neuron from the pre-recursion levels into a neuron in the current layer, this is a multilayer perceptron. The expression serves as its calculation (4).

$$f(x) = \sigma(p * x) \quad (4)$$

Where n is the count of neurons in a completely connected layer, k is the dimensions that range of x , σ is the activation function, and p is the resultant sized matrix $n * k$.

Result Layer: It is yet another hot vector that indicates the category of the input picture that has been provided [32]. It is written as an expression (5).

$$C(x) = (i|\exists v_{j \neq i} \text{ and } x_j \leq x_i) \quad (5)$$

SoftMax layer: Backpropagation of inaccuracy is its primary purpose [32]. Let P be the dimension of a vector's input. After that is completed, it may be written as a transformation function using expression(6).

$$S(x): R^P \rightarrow [0, 1]^P \quad (6)$$

The outcome is given for each component j ($1 \leq j \leq P$) using equation (7).

$$S(x)_j = \frac{e^{x_j}}{\sum_{i=1}^P e^{x_i}} \quad (7)$$

3.3 Proposed Method for Classifying Model

I. Artificial Intelligence (AI)

A subfield of computer science is artificial intelligence that examines how a computer system can do activities that would often be performed by people, and is generally denoted by or called as AI in short. Artificial intelligence (AI) is a type of computer system that advances in addition to identifying and reproduce mortal mental procedures and generate devices which stimulate human intelligence [15]. Mortals Individuals have the capacity to creatively resolve conflicts because they have experience and expertise that can be used. Computers system that have been given information and the capacity to analyse so they may behave similarly to the same extent as mankind. The three major goals of machine learning or Artificial Intelligence (AI) are to comprehend intelligence, make computers smarter, and make robots more useful. Understanding conflicting and confusing signals, responding swiftly and efficiently to novel situations, using logic to solve issues, and solving them successfully are all characteristics of Artificial intelligence (AI) [16]. There are two crucial elements that must be present in order to create an artificial Intelligence (AI) [15] application, which are as follows:

1. Knowledgebase, which includes information on connections between them, hypotheses, ideas, and observations.

1. An inference mechanism is a computer programme that can draw conclusions from the history.

II. Deep Machine Learning or Convolutional Neural Network (CNN)

The convolutional neural network (CNN) approach, which is a member of the deep neural network or learning that's hierarchical family due to its vast network deepness, functions quite well when applied to [17] image data. A Neural Networks methodology which is also said as Deep learning (DL), utilizes unique methods like the Restricted Boltzmann Machine to expedite understanding in neural networks with several layers or more than seven levels (RBM). Because the issue of missing the gradient in backpropagation is going to be less of an issue with Deep Learning, training will be quicker [18].

Hubel and Wiesel's investigations of the cat's [19] visual system served as the foundation for the original investigation that led to the CNN finding. Creatures have visually attractive cortexes that have influenced several research and led to the development of novel simulators including Neocognitron, HMAX [20], LeNet-5, and AlexNet.

The Multilayer Perceptron (MLP) approach for two-dimensional information extraction, such as analysing pictures or audio, led to the creation of the CNN method [21]. While CNN operates similarly to MLP, every neuron on CNN exists in two dimensions. Compare MLP, which only has one dimension for every neuron. Analysis of images might begin with a single feature, such as brightness or edge difficulty on features that specifically characterise objects based on surface area. Extraction Layers Features (Feature Extraction Layer) and Classifier Layer are the two distinct layer kinds in Convolutional Neural Networks (Classification Layer).

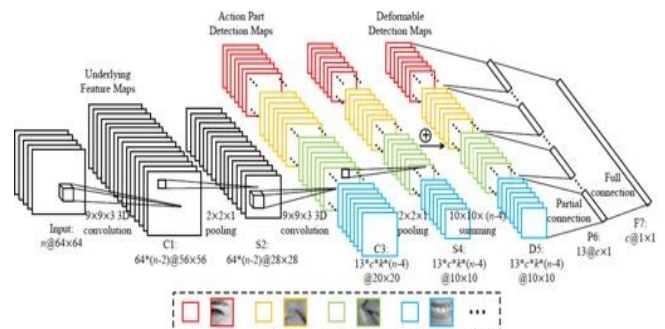


Fig. 9: Convolutional Neural Network Algorithm
(source: [22])

The neurons are arranged in three dimensions in the CNN layers (width, height, depth). Layer diameters are measured by width and height, whereas layer counts are indicated by depth. Tens to hundreds of millions of layers may make up a CNN, every single one being trained to recognise different types of pictures. In the below Figure (9) illustrates how image processing is applied to every training sample at a higher resolution, while each image's outcome being analysed and utilised as an input to the following layer.

EXPERIMENT RESULT:

A. LOSS AND ACCURACY PLOT

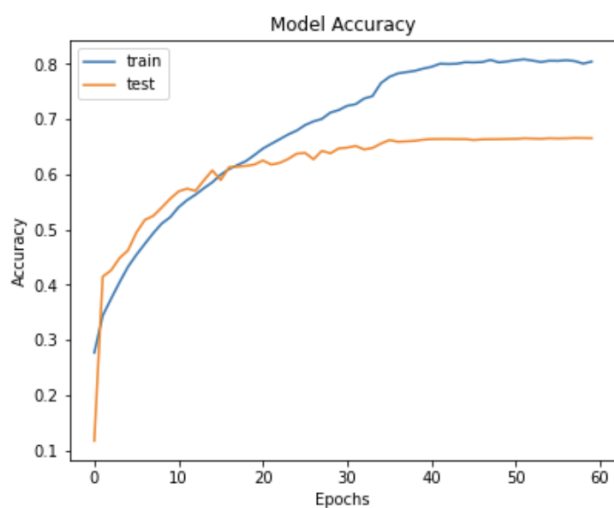


Fig (a). Plots of the training and testing Accuracy

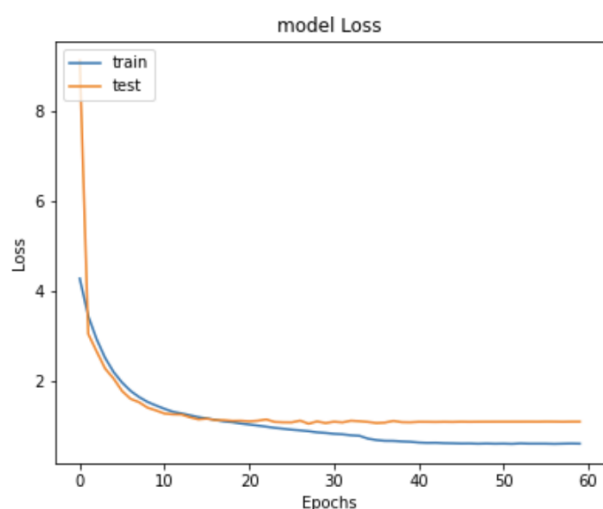


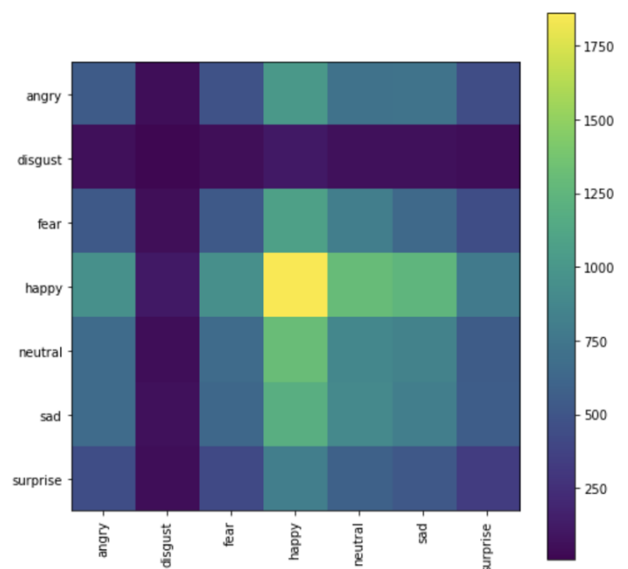
Fig. (b).Plots of the Training and testing Accuracy

The CNN model's learning curve, which charts the effectiveness of the algorithm across practise or duration, is seen in above Figures.(a) and.(b). The training dataset and a freeze-on validation dataset The dataset that are of training and freeze-on validation were used to assess the model following every upgrade. Due to their omnipresent usage for models that are in machine learning that gradually amend their internal measuring, learning curves were considered as the appropriate visual parameter for this study. It is clear from the learning curve in Figure (b) that the validation loss and training loss plots both drop to the point of consistency with a small stereotype mismatch. Also, it can be deduced from Figure (a) that there is a limited generalisation gap between the training accuracy and validation accuracy plots as training sequence and batch size grow. The model is suggested to generalise effectively and may suit the data well thematically.

B. CONFUSION MATRIX

For the final version of the confusion matrix on the FER2013 training and testing set is displayed in the table below. The model provides the most accurate classification of the emotion "disgust." On the other hand, it misclassifies "happy" first most frequently.

The following is the training set's confusion matrix followed by testing set's confusion matrix.



(Training set)

And the following is the table with accurate parameters for the above confusion matrix.

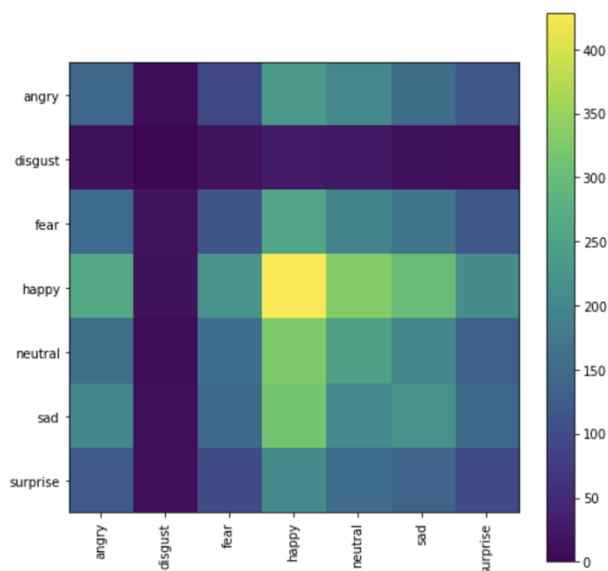
Basically, let's get introduced to the concept .A confusion matrix represents the prediction summary in matrix form. It shows how many prediction are correct and incorrect per class. It helps in understanding the classes that are being confused by model as other class.

angry	546	45	483	1017	722	728	45
disgust	61	8	56	121	73	71	40
fear	533	54	531	1073	813	647	44
happy	955	119	943	1859	1301	1251	78
neutral	659	51	664	1313	880	841	55
sad	661	66	632	1197	904	811	55
surprise	442	45	416	815	593	524	33
	angry	disgust	fear	happy	neutral	sad	surprise

angry	145	7	92	235	206	157	116
disgust	15	0	18	27	25	15	11
fear	150	16	115	256	197	169	121
happy	261	15	223	428	333	302	212
neutral	158	8	155	327	249	205	131
sad	202	9	144	316	208	220	148
surprise	123	9	99	210	155	139	96
	angry	disgust	fear	happy	neutral	sad	surprise

C. PERFORMANCE TABLE

Below table summarizes previous reported classification accuracies FER2013. The majority of described approaches outperform the expected human performance by about 65.5%. 72.7% is the largest individual accuracy that has been previously recorded [34]. We accomplish the most recent precision of 88.29% in this job.



(Testing set)

Lastly here's the following is the table with accurate parameters for the above confusion matrix.

Method	Accuracy Rate
CNN [24]	62.44 %
GoogleNet [28]	65.20 %
VGG+SVM [27]	66.31 %
Conv+Inception layer [29]	66.40 %
Bag of Words [26]	67.40 %
Attentional ConvNet [25]	70.02 %
CNN + SVM [30]	71.20 %
ARM(ResNet-18) [31]	71.38 %
Inception [32]	71.60 %
ResNet [32]	72.40 %
VGG [32]	72.70 %
CNN (this work)	88.29 %

ACKNOWLEDGEMENT

I would like to thank my professor N.G. Bhuvaneshwari for giving me such a great opportunity to work on this research project.

CONCLUSION:

The Convolutional Neural Network (CNN), here by using the approach for the forecasting of 7 (seven) Human Facial Expressions employing the dataset based FER-2013 called as Facial Emotion Recognition (FER) from the Raspberry Pi, it has altogether been successful in creating a software with a broad overview of the investigated object. The following process describes the system design in this study uses the process:

- 1) The Facial Expression Recognition 2013 Dataset, also known as FER-2013 dataset, is used in the training procedure together with the Convolutional Neural Network (CNN) approach for extraction of features and an appropriate facial algorithm is applied.
- 2) Face items discovered utilising the Haar Cascade approach, direct real-time facial expression detection, and facial expression classification using the Convolutional Neural Network (CNN) methodology.
- 3) The system's findings will be shown in the content audience while the facial expression recognition procedure is active on the emotion display board.

Using the FER2013 dataset, this article exhibits single-network state-of-the-art accuracy of classification. To create the best model possible for identifying facial expressions, we carefully adjust each hyperparameter. The greatest early test classification performance attained is 73.06%, exceeding all single-network accuracies originally recorded. Several optimizers and learning pace schedulers are investigated. In order to increase the generalization ability to 88.29%, we additionally do additional optimization on our model utilizing

Cosine Annealing and integrate the training and validation datasets. We are interested in conducting study on ensembles of several deep learning architectures on FER2013 and investigate numerous methods for image processing in hopes of further developing our presentation in facial emotion identification.

DATA AVAILABILITY STATEMENT

This study examines the Kaggle repository's Facial Emotion Recognition (FER-2013) dataset.

[<https://www.kaggle.com/datasets/msambare/fer2013>].

REFERENCE:

- [1] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan, "Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2003, vol. 5, doi: 10.1109/CVPRW.2003.10057.
- [2] F. Abdat, C. Maaoui, and A. Pruski, "Human-computer interaction using emotion recognition from facial expression," in *Proceedings - UKSim 5th European Modelling Symposium on Computer Modelling and Simulation, EMS 2011*, 2011, doi: 10.1109/EMS.2011.20.
- [3] B. Fasel and J. Luetttin, "Automatic facial expression analysis: A survey," *Pattern Recognition*, vol. 36, no. 1. 2003, doi: 10.1016/S0031-3203(02)00052-3.
- [4] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6. 2015, doi: 10.1109/TPAMI.2014.2366127.
- [5] N. Mehendale, "Facial emotion recognition using convolutional neural networks (FERC)," *SN*

Appl. Sci., vol. 2, no. 3, 2020, doi: 10.1007/s42452-020-2234-1.

[6] V. Tümen, Ö. F. Söylemez, and B. Ergen, "Facial emotion recognition on a dataset using Convolutional Neural Network," in *IDAP 2017 - International Artificial Intelligence and Data Processing Symposium*, 2017, doi: 10.1109/IDAP.2017.8090281.

[7] D. K. Jain, P. Shamsolmoali, and P. Sehdev, "Extended deep neural network for facial emotion recognition," *Pattern Recognit. Lett.*, vol. 120, 2019, doi: 10.1016/j.patrec.2019.01.008.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, 2017, doi: 10.1145/3065386.

[9] O. Gervasi, V. Franzoni, M. Riganelli, and S. Tasso, "Automating facial emotion recognition," *Web Intell.*, vol. 17, no. 1, 2019, doi: 10.3233/WEB-190397.

[10] M. M. Taghi Zadeh, M. Imani, and B. Majidi, "Fast Facial emotion recognition Using Convolutional Neural Networks and Gabor Filters," in *2019 IEEE 5th Conference on Knowledge Based Engineering and Innovation, KBEI 2019*, 2019, doi: 10.1109/KBEI.2019.8734943.

[11] E. Pranav, S. Kamal, C. Satheesh Chandran, and M. H. Supriya, "Facial Emotion Recognition Using Deep Convolutional Neural Network," in *2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS 2020*, 2020, doi: 10.1109/ICACCS48705.2020.9074302.

[12] I. J. Goodfellow *et al.*, "Challenges in representation learning: A report on three machine learning contests," *Neural Networks*, vol. 64, 2015, doi: 10.1016/j.neunet.2014.09.005.

[13] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, 1998, doi: 10.1109/5.726791.

[14] P. Ecmán, "Micro Expressions," 2017. [Online]. Available: <https://www.slideserve.com/tyrell/micro-expressions>

[15] J. Xia, F. Xie, Y. Zhang, and C. Caulfield, "Artificial intelligence and data mining: Algorithms and applications," 2013.

[16] Y. Zhang, S. Balochian, P. Agarwal, V. Bhatnagar, and O. J. Housheya, "Artificial intelligence and its applications," 2014.

[17] O. Arriaga, M. Valdenegro-Toro, and P. Ploger, "Real-time convolutional neural networks for emotion and gender classification," *arXiv preprint arXiv:1710.07557*, 2017.

[18] A. Ahmad, "Mengenai Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning," *Yayasan Cahaya Islam, Jurnal Teknologi Indonesia*, 2017.

[19] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Advances in Neural Information Processing Systems*, 2014.

[20] M. Paindavoine, J. Dubois, and P. Musa, "Neuro-inspired smart image sensor: analog Hmax implementation," *Image Sensors and Imaging Systems 2015*, vol. 9403, p. 94030J, 2015.

[21] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," in *3rd International Conference on Learning Representations, ICLR 2015 - Workshop Track Proceedings*, 2015.

[22] M. Liu, S. Li, S. Shan, R. Wang, and X. Chen, "Deeply learning deformable facial action parts model for dynamic expression analysis," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9006, 2015, pp. 143–157.

[23] K. Liu, M. Zhang, and Z. Pan, "Facial Expression Recognition with CNN Ensemble," in *Proceedings - 2016 International Conference on Cyberworlds, CW 2016*, 2016.

- [24] S. Minaee and A. Abdolrashidi, "Deep-emotion: facial expression recognition using attentional convolutional network," arXiv. 2019, doi: 10.3390/s21093046.
- [25] R. T. Ionescu, M. Popescu, and C. Grozea, "Local Learning to Improve Bag of Visual Words Model for Facial Expression Recognition," *Work. challenges Represent. Learn. ICML*, 2013.
- [26] M. I. Georgescu, R. T. Ionescu, and M. Popescu, "Local learning with deep and handcrafted features for facial expression recognition," *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2917266.
- [27] P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis, "Deep learning approaches for facial emotion recognition: A case study on FER-2013," in *Smart Innovation, Systems and Technologies*, 2018, vol. 85, doi: 10.1007/978-3-319-66790-4_1.
- [28] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*, 2016, doi: 10.1109/WACV.2016.7477450.
- [29] Y. Tang, "Deep learning using linear support vector machines," arXiv, vol. 1306.0239, 2013.
- [30] and S. Z. Shi, Jiawei, "Learning to Amend Facial Expression Representation via De-albino and Affinity," arXiv, vol. 2103.10189, 2021.
- [31] and M. K. Pramerdorfer, Christopher, "Facial expression recognition using convolutional neural networks: state of the art," arXiv, vol. 1612.02903, 2016.
- [32] A. Handa, R. Agarwal and N. Kohli, "Article: Incremental approach for multi-modal face expression recognition system using deep neural networks *Journal: International Journal of Computational Vision and Robotics (IJCVR)* 2021 Vol.11 No.1 pp.1 - 20.
- [33] A. Saravanan, G. Perichetla and K. Gayathri, "Facial Emotion Recognition using Convolutional Neural Networks," Arxiv.org, 2019. [Online]. Available: <https://arxiv.org/pdf/1910.05602.pdf>. [Accessed: 12- Oct2019].
- [34] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE, 2005, pp. 5–pp.
- [35] M. Valstar and M. Pantic, "Induced disgust, happiness and surprise: an addition to the mmi facial expression database," in *Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect*, 2010, p. 65.
- [36] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Automatic Face and Gesture Recognition*, 1998. *Proceedings. Third IEEE International Conference on*. IEEE, 1998, pp. 200–205.
- [37] J. M. Susskind, A. K. Anderson, and G. E. Hinton, "The toronto face database," *Department of Computer Science, University of Toronto, Toronto, ON, Canada, Tech. Rep*, vol. 3, 2010.
- [38] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.
- [39] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3d facial expression database for facial behavior research," in *Automatic face and gesture recognition*, 2006. *FGR 2006. 7th international conference on*. IEEE, 2006, pp. 211–216. [40] G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. PietikaInen, "Facial " expression recognition from near-infrared vid