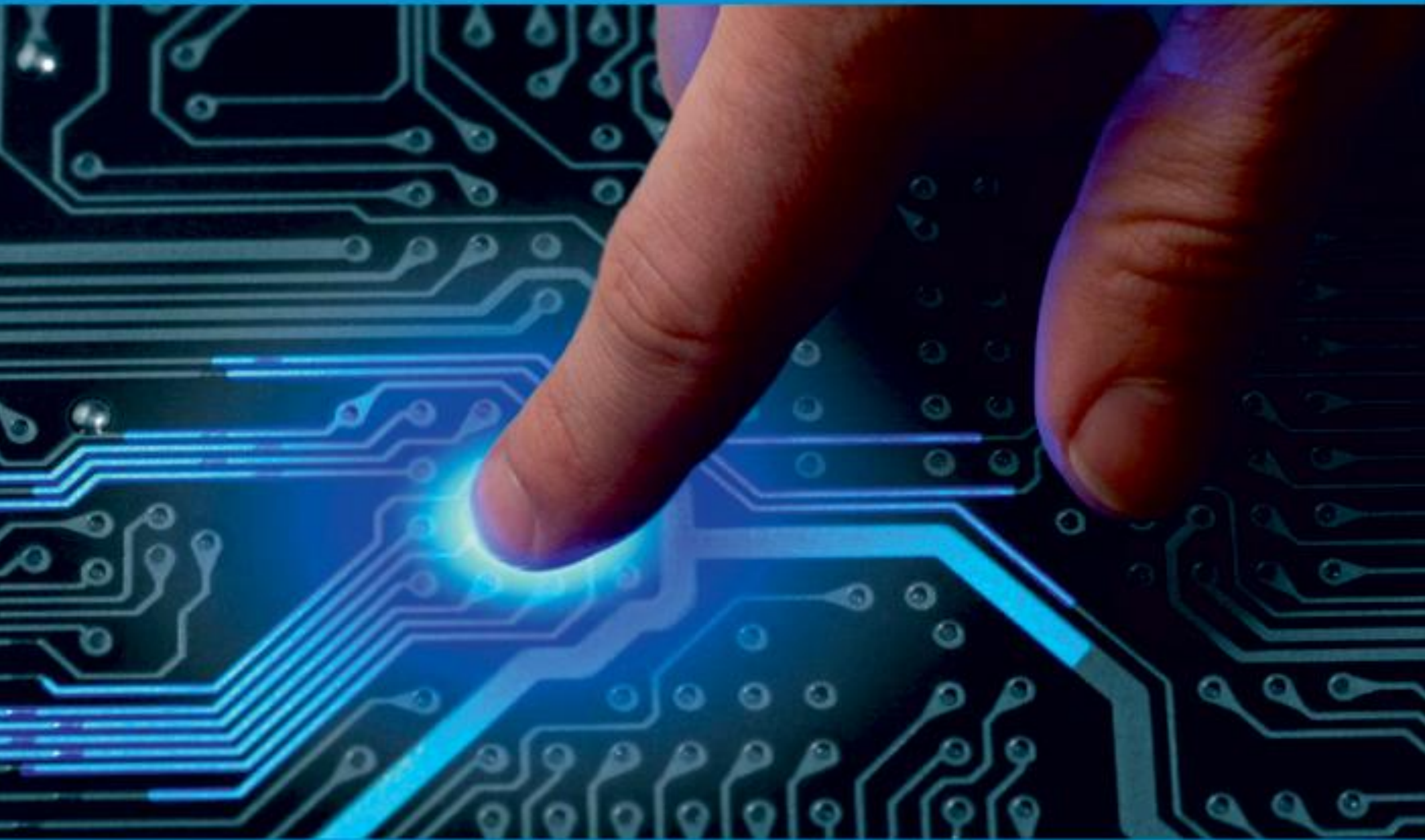
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105177|

Facial Emotion Recognition using Deep Learning

Keerthana H Magadi, Dr. Chandrsekar Shastry, Dr. Devaraj Verma C

PG Student, Department of Computer Science and Engineering, JAIN (Deemed-to-be UNIVERSITY), Bangalore, India

Professor, Department of Electronics and Communication Engineering, JAIN (Deemed-to-be UNIVERSITY), Bangalore, India

Professor, Department of Computer Science and Engineering, JAIN (Deemed-to-be UNIVERSITY), Bangalore, India

ABSTRACT: Facial expressions are the most visible way to communicate feelings. A person's facial expressions vary depending on the situation. Automated emotion identification via facial expressions is critical for realistic human-machine interactions. Although humans understand face emotions instantly, computer expression rearrangement remains a difficulty. Market research, video game testing, enhancing website customization, e-education development process improvement, software engineering, mental state, identification, security, automatic counselling systems, lie detection, automotive industry, consumer-based AI, event feedback research, cyber security, and other fields benefit from computerized facial emotion recognizers. Deep learning for FER has recently become the buzz of the town. Their popularity arises from their capacity to reliably extract features from picture data, as well as their cheap power and time consumption, greater resilience, and accuracy. This study finds that the best classification technique is SOFTMAX classifier, which has a validation accuracy of 97.58, and the best net is XCEPTION - RESNET V2, which has a validation accuracy of 90%. This project also seeks to create its own dataset of nine thousand photos.

KEYWORDS: Deep Learning, FER, Classifiers, Nets, Dataset

I. INTRODUCTION

Human-computer interaction technology is a type of technology that uses computer equipment as a medium to create human-computer interaction. With the fast growth of pattern recognition and artificial intelligence in recent years, an increasing amount of research has been undertaken in the field of human-computer interaction technology. As a key way of intelligent human-computer interaction, facial expression recognition has a wide range of applications [1]. It has been used in industries such as assisting medical, online education, interactive gaming, and public safety. Facial expression recognition uses computer image processing technology to extract data describing facial expression features from the initial input facial expression images and classify the features in accordance with human emotional expressions like surprise, happiness, neutrality, and aversion. By processing and analyzing facial picture characteristics, emotional recognition may distinguish fundamental human expressions such as surprise, grief, pleasure, disgust, wrath, and contempt. Furthermore, the possible applications of emotional recognition are numerous, including the service business, criminal investigation and interrogation, medical assistance, and so on. Most emotional recognition algorithms are broken into two stages: feature extraction and classification [6]. Feature extraction mostly analyzes facial photos to extract probable image characteristics. Because various expressions have varied facial expression properties, useful prospective features for improved categorization may be obtained. Every deep learning recognition or detection requires training algorithms and testing on appropriate data.

II. RELATED WORK

One of the main elements for precise emotion recognition using facial expressions is the dataset. WfaMellouk provides a list of available datasets along with their requirements. It also provides alternative deep learning architectures together with the percentage of accuracy with which it recognizes emotion [2]. This study fails to discuss real-time emotion identification and fails to distinguish between actual and artificial emotions. To identify facial expression, an inception net with transfer learning is utilized. NithyaRoopa.S discovered real-time photos with 35.6% accuracy. The work does not produce correct results. As a result, the model may have been more dependable with real-time images [3]. DY Liliana recognizes face emotions using the Cohn Kanade(CK+) dataset. This study uses a deep

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105177|

convolutional neural network approach to recognize the occurrence of facial action units as a component of a facial action coding system that captures human emotion. The performance achieves a 92.81% accuracy. The work fails to distinguish between genuine and fabricated emotions and may have been more accurate in real time [4]. Salem Bin SaqerAlMarri makes use of the databases FERC 2013, JAFFE, and RAVDESS. The paper employs rapid R-CNN. A high-quality video database is used to train it. In actual time, this reached 94% accuracy. The limitation of this study is that no correct findings are obtained when the resolution of the face is low, and it also fails to distinguish between actual and artificial emotions [5].

For face emotion detection, the Xuan- PhungHujuhuses database from the Chalear LAP challenge is used; a long short-term memory approach with parametric bias is employed to distinguish true and artificial emotions. The work's limitations are that it recognizes face expressions with 66.7% accuracy for real-time images [7].

III. PROPOSED ALGORITHM

This effort begins with the creation of a database, which is the first stage in developing an accurate model. There are two basic steps in developing a standard model for FER: Classification of the test image, training of the produced dataset. Every model includes the following 7 levels of training: Convolutional layer, ReLU layer, Max Pooling layer, Batch Normalization layer, dropout layer, Fully Linked Layer. These layers are classified as distinct NETS based on how many times they are repeated. Each NET's performance accuracy is checked and the results are documented. The CNN-NET trained model is used to verify the best classifier. Keeping the dataset and NET constant, the picture to be categorized [Test Image] is processed through multiple classifiers, and the detection accuracy for each classifier is recorded. The classifier that correctly recognizes the test picture will be verified as the best classifier. The outcomes of the best classifier will be recorded. By optimizing the dataset, NET, and classifier, a standard model is created.

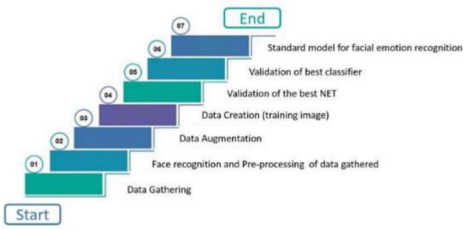


Fig. 1 Block Diagram

A. *Dataset Gathering*

To boost the dataset's robustness, a mix of existing datasets such as CK+, FER2013, Jaffe, KDEF, and a few Indian photos from Google was employed. Indian photos from Google are manually pre-processed, whereas the current collection is already pre-processed.

Fig. 2 Dataset

B. *Face Detection*

Face detection is necessary so that the model only learns the needed characteristics and the extraneous features are deleted from the photos. Several algorithms are used to recognize faces from input images, the most prominent of which being the Viola-Jones HAAR Cascade Classifier [8]. According to the Viola-Jonas method, the following formula should yield a value closer to 1 if a HAAR-like feature is present in a picture. The closer the number is to 1, the more likely it is that the HAAR feature will be detected in the image. Classifiers are used to recognize objects in images. Face detection is determined by the amount of HAAR characteristics and sliding windows (increases a number of stages increase). The Viola Jonas Method is divided into 38 phases. Face detection can occur at various stages depending on the size of the sliding windows and the location of the face, as well as the amount of characteristics.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105177|

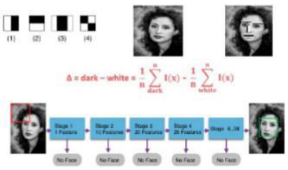


Fig. 3 Face Detection
Source: Research Gate

A. Pre-Processing of Face Images

Just the Indian facial photos from Google are pre-processed [to increase the dataset's resilience]. Photos from other datasets have previously been pre-processed. This is a crucial stage in which we eliminate noise from the image and improve the features, but before we begin, we convert our photos from RGB to gray scale. The reason for turning a rgb image to grayscale. When we convert an RGB image to a grayscale image, we only require 8 bits to keep a single pixel of the image, however when we save a single color pixel of an RGB color image, we need 8*3 = 24 bits (8 bit for each color component). As a result, storing a grayscale image will use 33% less memory than storing an RGB image [9]. Grayscale pictures are significantly easier to deal with in a range of tasks; it is easier to work with a single layered image (Grayscale image) than a three-layered image (RGB color image). In order to guarantee that each input parameter (in this case, pixel) has a consistent data distribution, data normalization is an essential step. This quickens convergence when the network is being trained. By eliminating the mean from each pixel and dividing the result by the standard deviation, data is normalized. Such data would have a distribution that resembles a zero-centered Gaussian curve. We may decide to scale the normalized data in the range [0, 1] or [0, 255] because we need positive pixel counts for picture inputs.

$$I_N = (I - \mathrm{Min})\frac{\mathrm{newMax} - \mathrm{newMin}}{\mathrm{Max} - \mathrm{Min}} + \mathrm{newMin} \qquad (1)$$

The method entails deducting 50 from each pixel's intensity to produce a range of 0 to 130, for instance, if the image's intensity range is 50 to 180 and the necessary range is 0 to 255. After multiplying each pixel's intensity by 255/130, a range of 0 to 255 is produced.

B. Data Augmentation

A huge dataset is critical for the deep learning model's success. Unfortunately, we may lack the quantity and diversity of data needed to train the model for a specific demand, resulting in poor performance. Data augmentation is a method of creating fresh training data from existing training data. This is accomplished by applying domain-specific approaches to examples from the training data, resulting in new and distinct training instances.

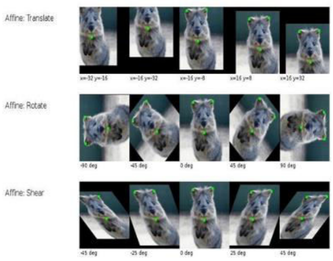


Fig. 4 Augmented Image
Source: Elsevier

C. Model

Every model has the following 7 training layers: input layer [to input photos for training, Convolutional layer, ReLU layer, Max Pooling layer, Batch Normalization layer, dropout Layer, Fully Connected Layer]. These layers are classified as distinct NETS based on how many times they are repeated.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105177|

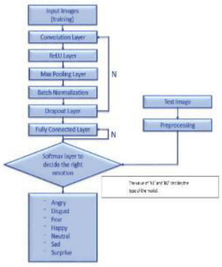


Fig. 5 Model Block Diagram

- Image input layer: An image input layer sends pictures to a network. The input data size is supplied as a row vector of integers [h w c], where w, h and c represent the width, height, and number of channels, respectively. For grayscale graphics, use a vector with c set to 1. Provide a vector with c equal to 3 for RGB pictures.

Fig. 3 Model Block Diagram
Source: Springer

- Convolutional layer: Sliding convolutional filters are used in a convolutional layer to filter the input. By moving the filters along the input both vertically and horizontally, computing the dot product of the weights and the input, and then adding a bias component, the layer convolves the input.

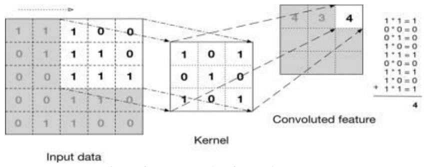


Fig. 6 Convolutional Layer
Source: Research Gate

- ReLU Layer:Each input element undergoes a threshold operation by a ReLU layer, which sets any value less than zero to zero. This procedure is equivalent to:

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (2)$$

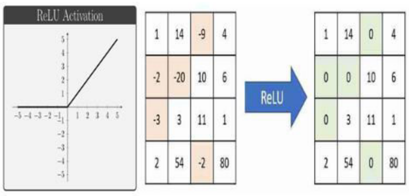


Fig. 7 ReLU Layer
Source: Research Gate

- Max Pooling Layer: When a pooling layer is introduced to a model, it decreases the dimensionality of pictures by lowering the number of pixels in the output of the preceding convolutional layer. This, in turn, decreases the computational complexity. The most significant characteristics, such as edges, are extracted by max pooling, whereas average pooling extracts features more gradually. Max pooling is preferable for obtaining extreme features. Average pooling cannot always extract excellent features since it counts everything and returns an average value, which may or may not be significant for object identification tasks.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105177|

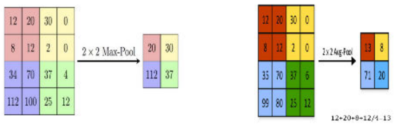


Fig. 8 Max Pooling layer
Source: Elsevier

- Batch Normalization: Batch normalization is a training strategy for very deep neural networks that standardizes the inputs to each mini-batch. This stabilizes the learning process and significantly reduces the number of training epochs necessary to create deep networks.

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1...m}\}$;
Parameters to be learned: γ, β
Output: $\{y_i = \mathrm{BN}_{\gamma,\beta}(x_i)\}$

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{// mini-batch mean}$$

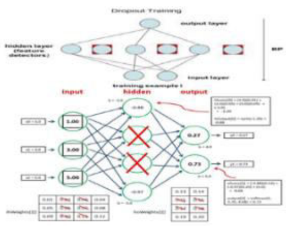
$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_{\mathcal{B}})^2 \qquad \text{// mini-batch variance}$$

$$\widehat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \qquad \text{// normalize}$$

$$y_i \leftarrow \gamma\widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \qquad \text{// scale and shift}$$

Fig. 9 Algorithm of Batch Normalizing Transform, applied to activation x over a mini batch
Source: Springer

- Drop-Out Layer: Dropout is a training method where neurons are rejected at random. At random, they are "dropped out". According to this, any weight changes are not contributed to the neuron on the backward pass and their temporally temporal influence to downstream neuron activation is erased on the forward pass [10]. The network becomes less sensitive to the precise weights of neurons as a result. This leads in a network that can generalize better and is less likely to over fit the training data.



Initialise parameters

Iterate over data:

X_i, Y_i = current mini_batch

NN = dropout Network

$\mathscr{L}(\theta)$ = compute_loss (NN, X_i, Y_i)

back_propagate(NN)

till satisfied

Fig. 10 Drop out layerFig. 10 Algorithm of Drop out layer
Source: Elsevier

- Fully Connected Layer: A feedforward network is essentially what the Fully Connected Layer is. Fully Connected Layers make up the last network levels. The final pooling or convolutional layer's output is flattened before being delivered as the input to the fully connected layer. A 3-dimensional matrix is created by the final (and any) Pooling and Convolutional Layer [11], which can be flattened by unrolling all of its values into a vector.
  $g(Wx + b)$(3)
  x — is the input vector with dimension $[p_1, 1]$
  W — Is the weight matrix with dimensions $[p_1, n_1]$ where, n_1 is the number of neurons in the current layer and p_1 is the number of neurons in the previous layer.
  b — Is the bias vector with dimension $[p_1, 1]$
  g — Is the activation function.
  This process is done for each layer.
  The first fully connected layer uses the feature analysis inputs and weights to predict the proper label. The final probabilities for each label are provided by the fully linked output layer.
  A fully connected layer's goal is to use the results of the convolution/pooling process to categorize the picture into a label (in a simple classification example).
  The output of convolution/pooling is flattened into a single vector of values, each of which expresses the probability that a particular feature is related to a given label.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105177|

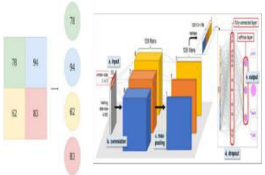


Fig. 11 Fully Connected layer
Source: Research Gate

- Classification Layer: A classification layer computes the cross entropy loss for multi-class classification problems with mutually exclusive classes. The classification layer must come after the SOFTMAX layer in the majority of classification networks. For a 1-of-K coding scheme, the train network assigns each input to one of the K mutually exclusive classes at the classification layer using the cross entropy function.

$$\text{loss} = -\sum_{i=1}^{N}\sum_{j=1}^{k} t_{ij} \ln y_{ij} (3)$$

where N is the number of samples,

t_{ij} is the indicator that the i^{th} sample belongs to the j^{th}class,?

k is the number of classes and

y_{ij} is the output for sample i for class j, which in this case, is the value from the SOFTMAX function? That is, it is the probability that the network associates the i^{th} input with class j.

layer = classificationLayer creates a classification layer

D. Network architecture

The concept allows for the usage of various network designs. A brief overview of the architectures employed in this study is provided below:

- Alexnet: The Alexnet has eight different levels of parameters that may be taught. The model consists of five layers: a max pooling layer at the top, three fully connected layers underneath it, and an output layer. All of these levels, with the exception of the output layer, use ReLU activation. This model accepts 227X227X3 photographs as input and has 8 layers with learnable parameters. RGB photos are sent into the Model. It has 5 convolution layers as well as a max-pooling layer combination. Then there are three totally connected layers. The activation function used in all stages is ReLU. It employed two Dropout layers. The activation function used in the output layer is SOFTMAX[12]. This architecture contains 62.3 million parameters in total.
- VGG 16: VGG16 (also known as OxfordNet) is an Oxford Visual Geometry Group-inspired convolutional neural network design. VGGNet-16 is a 16-layer network that is intriguing due to its very consistent architecture. It is similar to AlexNet[13] in that it has a small number of filters but a big number of convolutions. Vgg16 is composed of 138 parameters.
- ResNet: The most recent cutting-edge convolutional neural network design is ResNet 50. Its architecture is similar to that of networks such as VGG-16, but it adds the capability of identity mapping. The big breakthrough with ResNet was that it allowed us to train extraordinarily deep neural networks with 150+ layers. Prior to ResNet, training very deep neural networks was difficult due to the problem of vanishing gradients [14]. ResNet Resnet50 was the first to introduce the concept of a skip connection. Skip connections are used in this context for two reasons: they solve the problem of vanishing gradients by providing an extra shortcut channel for gradient movement, and they provide an additional shortcut channel for gradient movement. They allow the model to learn an identity function that ensures the higher layer performs at least as well as, if not better than, the lower layer.
- Xception: Xception is an architectural improvement to Inception that replaces standard Inception modules with depth-wise separable convolutions. Xception is inspired by and based on a 'extreme' interpretation of Google's Inception model. Xception is a depth-wise separable convolution layer stack with residual connections [15]. A basic and modular architecture. Xception is a deep convolutional neural network with 71 layers.
- Inception V3: On the ImageNet dataset, it has been demonstrated that the well-known image recognition model Inception v3 can achieve greater than 78.1% accuracy. The model uses dropouts, convolutions, average pooling, max pooling, concatenations, and entirely connected layers, among other symmetric and asymmetric building blocks [16]. The model frequently employs batchnorm, which is applied to activation inputs. Using SOFTMAX, loss is calculated. A deep convolutional neural network with 48 layers is called Inception-v3. Inception-v3, which has 24M parameters, is Inception-v1's successor. There are 22 layers and 5 million parameters in the Inception-v1 architecture. In this situation, the Network in Network approach is significantly employed. This is accomplished using "Inception modules."

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105177|

- XceptionResnet V2: The retrieved features from Exception and ResNet50V2 are combined, and the resulting features are then coupled to a convolutional layer that is connected to the classifier to create a concatenated neural network. Following the concatenated features is a convolutional layer with a kernel size of 1*1, 1024 filters, and no activation function. This layer was created to separate a more significant semantic feature from a spatial point's features, which are shared by all channels and are each feature maps [17]. By combining input from ResNet50V2 and Xception, this convolutional layer accelerates the learning of the network.

All of the aforementioned architectures are used in this study to assess the most accurate model for real-time facial emotion recognition.

E. *Classifiers:*

- There are different classifiers which can be used in the model. A brief description of those classifiers that are used in this work is given bellow:
- SOFTMAX: A vector of K real values is passed to the SOFTMAX function, which converts it to a vector of K real values that sum to 1. The SOFTMAX transforms input values that are positive, negative, zero, or higher than one into values between 0 and 1, making them understandable as probabilities. The SOFTMAX converts inputs into low probabilities when they are small or negative and high probabilities when they are large, but they are always between 0 and 1[18]. Multi-class logistic regression and SOFTMAX are other names for the same algorithm. This is so because the SOFTMAX, a logistic regression extension that may be used for multi-class classification, has a formula that is very close to the sigmoid function used in logistic regression and can be utilized for logistic regression. Only when the classes are mutually exclusive is it possible to utilize the SOFTMAX function in a classifier.
- Support vector machines (SVMs): Support vector machines (SVMs) are linear classifiers that use the concept of margin maximization. They reduce structural risk while increasing classifier complexity in order to achieve outstanding generalization performance. SVM is a classification and regression supervised machine learning method [8]. It modifies your data using the kernel trick and then computes a suitable boundary between the potential outputs based on these changes. SVM is a Supervised Learning technique used to solve classification and regression problems. Nonetheless, it is widely used in Machine Learning to solve classification problems. SVMs may perform both non-linear and linear classification by utilizing the Kernel approach or parameter, which implicitly turns their inputs into high-dimensional feature spaces.
- KNN: The K-nearest neighbors (KNN) approach is a type of supervised machine learning (ML) methodology that may be used for classification as well as regression prediction. Yet, it is mostly used in industry for classification and prediction tasks. The following two features describe KNN: Lazy learning algorithm KNN is a lazy learning algorithm since it does not have a separate training phase and utilizes all of the data for training while classification [20]. Non-parametric learning algorithm KNN is also a non-parametric learning algorithm since it makes no assumptions about the underlying data.
- Random forest: DT (base learner) + bagging (Row sampling with replacement) + feature bagging (column sampling) + aggregation (mean/median, majority vote) Equals random forest. Train DT to full depth length as a result [19]. We don't care about depth; we let it grow because variance diminishes with aggregation. Random-forest uses a Decision tree as a basis and does row and column sampling. Models h1, h2, h3, and h4 differ more as a result of column sampling than if just bagging was utilized.

In this work, all of the classifiers stated above were tested in order to find the most accurate model for real-time facial emotion recognition.

IV. SIMULATION RESULTS

Dataset created contains a total of 9K images with angry emotion containing 1449 images, disgust emotion containing 1013 images, fear emotion containing 1304 images, happy emotion containing 1617 images, normal emotion containing 1449 images, sad emotion containing 1482 images, surprise emotion containing 1461 images. Image pixel value is 75X75X2 which is normalized in range – 0 – 255. Link to the dataset created is: DATASET. Created dataset is kept constant in the process of validation of different models and classifiers. The table 1 shows the performance metrics for different NETS that are used for the propose of validation. Validation is done keeping the 9K dataset and SOFTMAX classifier constant. NETS are placed in the increasing of order of their training accuracy. Hence the best NET for detecting the facial emotion recognition is XCEPTION – RESNET V2 with a validation accuracy of 90%.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105177|

TABLE I. PERFORMANCE METRICS OF DIFFERENT NETS KEEPING DATASET AND CLASSIFIER CONSTANT

Sl. No.	NETS	METRICS							
		TA	VA	TL	VL	Parameters Trained	Epochs	MB	LR
1	ALEXNET	0.9486	0.79796	0.2494	0.6845	28873219	100	32	0.0005
2	INCEPTIONV3	0.9585	0.81835	0.1127	3.4558	23908135	100	32	0.0005
3	RESNET 50	0.9643	0.7813	0.1132	1.0198	23602055	100	2	0.0005
4	VGG16	0.9831	0.8054	0.0497	3.4285	33625927	100	32	0.0005
5	XCEPTION	0.9727	0.793333	0.0803	0.9082	21387823	100	32	0.0001
6	XCEPTION – RESNETV2	0.9995	0.9	0.0031	0.8348	41236958	100	32	0.0001

TABLE II. PERFORMANCE METRICS OF DIFFERENT NETS KEEPING DATASET AND NETS CONSTANT

Sl. No.	CLASSIFIERS	METRICS							
		TA	VA	TL	VL	Parameters Trained	Epochs	MB	LR
1	SOFTMAX	0.9995	0.975867	0	0.024133	70	64	0.0001	0.9995
2	SVM	0.9634	0.664765	0	0.335235	70	64	0.0001	0.9634
3	KNN [K = 5]	0.9995	0.974989	0	0.025011	70	64	0.0001	0.9995
4	RANDOM FOREST	0.9995	0.975428	0	0.024572	70	64	0.0001	0.9995
5	SOFTMAX + SVM	0.9995	0.974989	0	0.025011	70	64	0.0001	0.9995

The table 2 shows the accuracy results different classifiers that are used for the propose of validation. Validation is done keeping the 9K dataset that is created and the model XCEPTION – RESNET V2 constant. Models are placed in the increasing of order of their training accuracy. Hence the best model for detecting the facial emotion recognition is SOFTMAX with a validation accuracy of 97.58%.

The created dataset is used to validate different nets and classifiers. Out of all the existing nets XCEPTION-RESNET V2 is found more accurate. Out of all the existing classifiers SOFTMAX is found more accurate for facial emotion recognition. Hence the standard model is created using XCEPTION- RESNET V2 and SOFTMAX classifier. This standard model is having a training accuracy of 99.99% and validation accuracy of 90% for the created dataset. Fig.12 and Fig.13 shows the confusion matrix and prediction accuracy of the created standard model respectively.

Fig.14, Fig.15 shows the simulation results of two different emotions namely, happy and normal from the standard model.

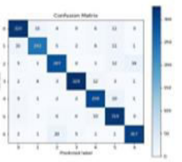


Fig. 12. Confusion Metrix

Fig. 14. Prediction of Neutral Emotion

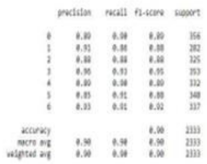


Fig. 13. prediction accuracy

Fig. 15. Prediction of Happy Emotion

V. CONCLUSION AND FUTURE WORK

The idea of emotion has a long evolutionary history that precedes the appearance of humans. This finding supports the notion that some emotions are intrinsic in people rather than socially manufactured. While most individuals find it easy to comprehend emotions, computers, on the other hand, have grappled with this idea for decades. This study covers a wide range of nationalities in the produced 9k Train data to account for these variances and so boost the model's resilience. This study verifies the best NET for face emotion recognition, XCEPTION - RESNET V2, which has 99.99% training accuracy and 90% validation accuracy [among all current models for facial emotion identification, including ALEXNET, RESNET 50, INCEPTION V3, VGG16, and XCEPTION models]. This study also supports SOFTMAX as the best classifier for face emotion identification, with 99.95% training accuracy and 97.58% validation accuracy (out of all current classifiers for facial emotion detection, which include SOFTMAX, KNN[WITH K VALUES = 5], SVM, and RANDOM FOREST).

There is room for improvement in real-time face expression identification accuracy. Numerous things contribute to an individual's emotions being conveyed. Some of these are posture, voice, facial expressions, conduct, and actions. Facial expressions are more important than the other variables described above since they are plainly discernible. When speaking with others, people can discern the emotions of other humans with a high degree of accuracy that is nearly similar to human perception. Despite the fact that several approaches to this problem have been explored, significant limits remain.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105177|

REFERENCES

1. A. Mehrabian, 'Communication without words', Psychol. Today., 1968, 2, (4), pp. 53–56
2. P. Ekman, W. V. Friesen, "Constants across cultures in the face and emotion," Journal of Personality and Social Psychology, 17(2), 124– 129, 1971, doi:10.1037/h0030377.
3. S. H. . A. G. . A. Routray., "A real time facial expression classification system using local binary patterns," in Intelligent Human Computer Interaction (IHCI), 2012 4th International Conference on., pp. 1–5, 2012.
4. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015, pages 1-9, 2015.
5. K. Trohidis, G. Tsoumakas, G. Kalliris, and I. P. Vlahavas. Multi- label classification of music into emotions. In ISMIR, volume 8, pages 325-330, 2008.
6. L. Pham , T. H. Vu, & T. A. Tran (2021). Facial Expression Recognition Using Residual Masking Network. 2020 25th International Conference on Pattern Recognition (ICPR), 4513–4519.
7. X. Liu , P. He, W. Chen, & J. Gao (2019). Improving Multi-Task Deep Neural Networks via Knowledge Distillation for Natural Language Understanding.
8. S. Minaee, Y. Boykov, F. Porikli, A. Plaza , N. Kehtarnavaz , & D. Terzopoulos (2020). Image Segmentation Using Deep Learning: A Survey.
C. Shi, C. Tan, & L. Wang (2021). A Facial Expression Recognition Method Based on a Multibranch Cross-Connection Convolutional Neural Network. IEEE Access, 9, 39255–39274.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.379

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING