IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304222 |

The Educational J44:Analyzing Facial Gestures with the Help of Machine Learning

DR.T.ARAVIND, ETHAMUKKALA LIKHITHA, REDDY YOGESWARI, KAMMA PENCHALA BABU,

Assistant Professor, Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

ABSTRACT: In this paper, a detailed analysis and comparison are presented on FER approaches. We categorized these approaches into two major groups: (1) conventional ML-based approaches and (2) DL-based approaches. The convention ML approach consists of face detection, feature extraction from detected faces and emotion classification based on extracted features. Several classification schemes are used in conventional ML for FER, consisting of random forest, AdaBoost, KNN and SVM. In contrast with DL-based FER methods, the dependency on face physics-based models is highly reduced. In addition, they reduce the preprocessing time to enable "end-to-end" learning in the input images. However, these methods consume more time in both the training and testing phases. Although a hybrid architecture demonstrates better performance, micro-expressions remain difficult tasks to solve due to other possible movements of the face that occur unwillingly.

KEYWORDS: facial expressions; facial emotion recognition (FER); technological development; healthcare;

I.INTRODUCTION

It is also conducive to detecting human attention, such as behavior, mental state, personality, crime tendency, lies, etc. Regardless of gender, nationality, culture and race, most people can recognize facial emotions easily. However, a challenging task is the automation of facial emotion detection and classification. The research community uses a few basic feelings, such as fear, aggression, upset and pleasure. However, differentiating between many feelings is very challenging for machines. In addition, machines have to be trained well enough to understand the surrounding environment—specifically, an individual's intentions. When machines are mentioned, this term includes robots and computers. A difference is that robots involve communication abilities to a more innovative extent since their design consists of some degree of autonomy. The main problem is classifying people's emotions is variations in gender, age, race, ethnicity and image quality or videos. It is necessary to provide a system capable of recognizing facial emotions with similar knowledge as possessed by humans. Recently, FER has become an emerging field of research, particularly for the last few decades. Computer vision techniques, AI, image processing and ML are widely used to expand effective automated facial recognition systems for security and healthcare applications.

As described above, FER models must be trained on a set of labeled images before they can be used to classify new input images. Training for these applications requires large datasets of facial imagery with each displaying a discrete emotion — the more labeled images, the better. Making a decision on which dataset to train the network on is no easy task, particularly because high-quality FER datasets are hard to find. Few such datasets are publicly available, and those that are, come with idiosyncrasies which must be understood and taken into account. The most crucial points to consider when making a dataset selection are the *size* and *quality* of the set. The size is arguably the most important, and also the simplest to explain. Ideally a dataset should contain thousands, or preferably millions of images.

Many publicly-available FER datasets come with only hundreds, or sometimes thousands of images. By contrast, Affectiva's emotion database of over 5 million faces is used for commercial emotion classification products. However, this data is not publicly available. The quality of a dataset can be considered in various ways. One key consideration is the level of representation of given emotion classifications. It is common for certain emotions to be disproportionately under or over-

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304222 |

represented by comparison to others in many datasets. Datasets generally contain far fewer examples of disgust, for example, than happiness or anger. Another quality consideration is how the emotions expressed in datasets were stimulated. Were actors told to 'play out' emotions? Or, were people subjected to authentic emotional experiences, which were then recorded? These two approaches can lead to very different results. A further consideration is that of head pose.

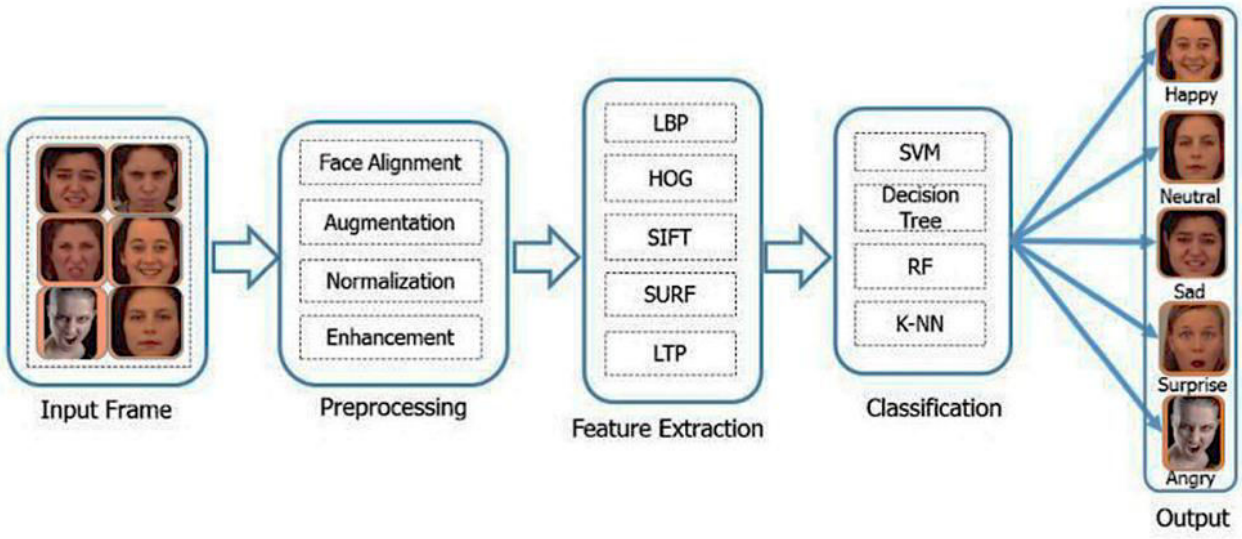

Fig 1: Facial Emotion Recognition

What is the range of variability in terms of the illumination of the subject in each image, and how does that relate to your target environment? How were images labeled when the dataset was created and by whom? Is there any underlying bias in the labels that will increase the bias of your model? All of these are questions of dataset quality which should be taken into account when selecting a dataset. The selection of a dataset must be conducted with an eye to the set of target emotions for classification. As mentioned previously, some emotional expressions resemble each other. Also, subtle expressions, such as contempt, can be extremely hard to pick up on. Therefore, some datasets will outperform others for certain emotional sets.

Neural networks trained on a limited set of emotions generally tend to result in higher rates of accurate classifications. To give the most extreme example, training on a dataset containing only examples of happiness and anger tends to produce very high accuracy. Two such distinct target classifications mean that the overlap in the facial expressive range is minimized. When neural networks are trained on datasets containing under-represented emotions, such as disgust, they tend to misclassify those emotions. This occurs because multiple emotions can share similar facial features and the dataset does not contain large enough example sets of one emotion to find definitive distinctive patterns.

II.RELATED WORK

A convolution neural network (CNN) with a focus function (ACNN) was suggested by Li et al. [that could interpret the occlusion regions of the face and concentrate on more discriminatory, non-occluded regions. A CNN is an end-to-end learning system. First, the different depictions of facial regions of interest (ROIs) are merged. Then, each representation is weighted by a proposed gate unit that calculates an adaptive weight according to the area's significance. Two versions of ACNN have been developed in separate ROIs: patch-based CNN (pACNN) and global-local-based ACNN (gACNN). Lopes et al. classified human faces into several emotion groups. They used three different architectures for classification: (1) a CNN with 5 C layers, (2) a baseline with one C layer and (3) a deeper CNN with several C layers. Breuer and Kimmel presented multi-level mechanisms for detecting facial emotions by combining temporal and spatial features. They used a CNN architecture for spatial feature extraction and LSTMs to model the temporal dependencies. Finally, they fused the output of both LSTMs and CNNs to provide a per-frame prediction of twelve facial emotions. Hasani and Mahoor presented the 3D Inception-ResNet model and LSTM unit, which were fused to extract temporal and spatial features from the input frames of

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304222 |

video sequences. (Zhang et al., and Jain et al. [61] suggested a multi-angle optimal pattern-dependent DL (MAOP-DL) system to address the problem of abrupt shifts in lighting and achieved proper alignment of the feature set by utilizing optimal arrangements centered on multi-angles. Their approach first subtracts the backdrop, isolates the subject from the face images and later removes the facial points' texture patterns and the related main features. The related features are extracted and fed into an LSTM-CNN for facial expression prediction.

Al-Shabi et al. qualified and collected a minimal sample of data for a model combination of CNN and SIFT features for facial expression research. A hybrid methodology was used to construct an efficient classification model, integrating CNN and SIFT functionality. Jung et al. [63] proposed a system in which two CNN models with different characteristics were used. Firstly, presence features were extracted from images, and secondly, temporal geometry features were extracted from the temporal facial landmark points. These models were fused into a novel integration scheme to increase FER efficiency. Yu and Zhang [64] used a hybrid CNN to execute FER and, in 2015, obtained state-of-the-art outcomes in FER.

III.METHODS

The *image preprocessing* stage can include image transformations such as scaling, cropping, or filtering images. It is often used to accentuate relevant image information, like cropping an image to remove a background. It can also be used to augment a dataset, for example to generate multiple versions from an original image with varying cropping or transformations applied. The *feature extraction* stage goes further in finding the more descriptive parts of an image. Often this means finding information which can be most indicative of a particular class, such as the edges, textures, or colors. The training stage takes place according to the defined *training architecture*, which determines the combinations of layers which feed into each other in the neural network. Architectures must be designed for training with the composition of the feature extraction and image preprocessing stages in mind. This is necessary because some architectural components work better with others when applied separately or together.

Fig 2: Analysis of Machine Learning Algorithms

For example, certain types of feature extraction are not useful in conjunction with deep learning algorithms. They both find relevant features in images, such as edges, and therefore it is redundant to use the two together. Applying feature extraction prior to a deep learning algorithm is not only unnecessary, but can even negatively impact the performance of the architecture.

IV.RESULT ANALYSIS

A comparison of training algorithms

Once any feature extraction or image preprocessing stages are complete, the training algorithm produces a trained prediction model. A number of options exist for training FER models, each of which has strengths and weaknesses making them more or less suitable for particular situations.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304222 |

In this article we will compare some of the most common algorithms used in FER:

- Multiclass Support Vector Machines (SVM)
- Convolutional Neural Networks (CNN)
- Recurrent Neural Networks (RNN)
- Convolutional Long Short-Term Memory (ConvLSTM)

Multiclass *Support Vector Machines (SVM)* are supervised learning algorithms that analyze and classify data, and they perform well when classifying human facial expressions. However, they only do so when the images are created in a controlled lab setting with consistent head poses and illumination.

SVMs perform less well when classifying images captured "in the wild," or in spontaneous, uncontrolled settings. Therefore, the latest training architectures being explored are all deep neural networks which perform better under those circumstances. *Convolutional Neural Networks (CNN)* are currently considered the go-to neural networks for image classification, because they pick up on patterns in small parts of an image, such as the curve of an eyebrow.

CNNs apply kernels, which are matrices smaller than the image, to chunks of the input image. By applying kernels to inputs, new activation matrices, sometimes referred to as feature maps, are generated and passed as inputs to the next layer of the network. In this way, CNNs process more granular elements within an image, making them better at distinguishing between two similar emotion classifications.

Fig 3: Result analysis of facial expression recognition

Alternatively, *Recurrent Neural Networks (RNN)* use dynamic temporal behavior when classifying an image. This means that when an RNN processes an input example, it doesn't just look at the data from that example — it also looks at the data from previous inputs, which are used to provide further context. In FER, the context could be previous image frames of a video clip.

The idea of this approach is to capture the *transitions between* facial patterns over time, allowing these changes to become additional data points supporting classification. For example, it is possible to capture the changes in the edges of the lips as an expression goes from neutral to happy by smiling, rather than just the edges of a smile from an individual image frame.

V.CONCLUSION

Additionally, different datasets related to FER are elaborated for the new researchers in this area. For example, human facial emotions have been examined in a traditional database with 2D video sequences or 2D images. However, facial emotion recognition based on 2D data is unable to handle large variations in pose and subtle facial behaviors. Therefore, recently, 3D facial emotion datasets have been considered to provide better results. Moreover, different FER approaches and standard evaluation metrics have been used for comparison purposes, e.g., accuracy, precision, recall, etc. FER performance has increased due to the combination of DL approaches. In this modern age, the production of sensible machines is very

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304222 |

significant, recognizing the facial emotions of different individuals and performing actions accordingly. It has been suggested that emotion-oriented DL approaches can be designed and fused with IoT sensors. In this case, it is predicted that this will increase FER's performance to the same level as human beings, which will be very helpful in healthcare, investigation, security and surveillance.

REFERENCES

1. Jabeen, S.; Mehmood, Z.; Mahmood, T.; Saba, T.; Rehman, A.; Mahmood, M.T. An effective content-based image retrieval technique for image visuals representation based on the bag-of-visual-words model. *PLoS ONE* **2018**, *13*, e0194526. [Google Scholar] [CrossRef] [PubMed] [Green Version]
2. Moret-Tatay, C.; Wester, A.G.; Gamermann, D. To Google or not: Differences on how online searches predict names and faces. *Mathematics* **2020**, *8*, 1964. [Google Scholar] [CrossRef]
3. Ubaid, M.T.; Khalil, M.; Khan, M.U.G.; Saba, T.; Rehman, A. Beard and Hair Detection, Segmentation and Changing Color Using Mask R-CNN. In Proceedings of the International Conference on Information Technology and Applications, Dubai, United Arab Emirates, 13–14 November 2021; Springer: Singapore, 2022; pp. 63–73. [Google Scholar]
4. Meethongjan, K.; Dzulkifli MRehman, A.; Altameem, A.; Saba, T. An intelligent fused approach for face recognition. *J. Intell. Syst.* **2013**, *22*, 197–212. [Google Scholar] [CrossRef]
5. Elarbi-Boudihir, M.; Rehman, A.; Saba, T. Video motion perception using optimized Gabor filter. *Int. J. Phys. Sci.* **2011**, *6*, 2799–2806. [Google Scholar]
6. Joudaki, S.; Rehman, A. Dynamic hand gesture recognition of sign language using geometric features learning. *Int. J. Comput. Vis. Robot.* **2022**, *12*, 1–16. [Google Scholar] [CrossRef]
7. Abunadi, I.; Albraikan, A.A.; Alzahrani, J.S.; Eltahir, M.M.; Hilal, A.M.; Eldesouki, M.I.; Motwakel, A.; Yaseen, I. An Automated Glowworm Swarm Optimization with an Inception-Based Deep Convolutional Neural Network for COVID-19 Diagnosis and Classification. *Healthcare* **2022**, *10*, 697. [Google Scholar] [CrossRef]
8. Yar, H.; Hussain, T.; Khan, Z.A.; Koundal, D.; Lee, M.Y.; Baik, S.W. Vision sensor-based real-time fire detection in resource-constrained IoT environments. *Comput. Intell. Neurosci.* **2021**, *2021*, 5195508. [Google Scholar] [CrossRef]
9. Yasin, M.; Cheema, A.R.; Kausar, F. Analysis of Internet Download Manager for collection of digital forensic artefacts. *Digit. Investig.* **2010**, *7*, 90–94. [Google Scholar] [CrossRef]
10. Rehman, A.; Alqahtani, S.; Altameem, A.; Saba, T. Virtual machine security challenges: Case studies. *Int. J. Mach. Learn. Cybern.* **2014**, *5*, 729–742. [Google Scholar] [CrossRef]
11. Afza, F.; Khan, M.A.; Sharif, M.; Kadry, S.; Manogaran, G.; Saba, T.; Ashraf, I.; Damaševičius, R. A framework of human action recognition using length control features fusion and weighted entropy-variances based feature selection. *Image Vis. Comput.* **2021**, *106*, 104090. [Google Scholar] [CrossRef]

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details