DOI:10.9756/INTJECSE/V14I5.677 ISSN: 1308-5581 Vol 14, Issue 05 2022

Mathematical Methods for Facial Expression Recognition with Applications in Challenging Situations

Naorem Tikendrajit Singh,
Research Scholar, Department of Mathematics, School of Science, MUIT, Lucknow.
Dr.Parul Saxena,
Professor, Department of Mathematics, School of Science, MUIT, Lucknow.

Abstract--- Because of the wide range of potential uses, emotion detection has become an inevitability and a difficult problem in computer science. Users get input and feelings from the system via nonverbal indicators including body language, facial expressions, and gestures. Human–Computer Interaction relies on the sensor's sensitivity and algorithmic resilience to improve recognition. Accurate detection relies heavily on sensors, which provide the system a higher level of efficiency and dependability by giving such high-quality input. It would be beneficial for the robots to learn social intelligence if they were able to recognise human emotions. An overview of several methods and strategies for recognizing emotions is provided in this work. It provides an overview of the databases that may be used as data sets for facial expression-detection systems. To observe emotion detection in Augmented Reality, Microsoft HoloLens (MHL), a mixed reality gadget, will be offered later (AR). We'll go through a quick rundown of MHL's sensors, how they could be used in emotion identification, and some early findings.

Keywords--- Convolutional Neural Network (CNN), Facial Expression, Recognition, Haar Cascade, Keras, Real-Time Model, Seven Universal Expressions, TensorFlow, Web-Interface.

I. Introduction

Facial expressions may be recognized because of the unique qualities of the human face. Facial expression change (FER) is described as a shift in facial expression produced by a person's internal emotional condition. Face image processing, facial video surveillance, and facial animation, as well as computer vision, digital image processing, and artificial intelligence, are just a few of the many HCI applications that make use of it. In recent years, several scholars have been interested in the problem of automatic face expression identification. Feature extraction is crucial in FER. According to Alek et al., facial expression accounts for 55% of overall transmission, whereas voice and spoken communication accounts for 38% and 7%, respectively.

A FER system may be designed using one of two major methods. As a first phase, some systems use a series of visuals that go from neutral to emotional. Because they have a limited amount of information, systems that employ a single picture of the face to identify associated emotions frequently perform worse than leading alternatives. A FER system may be categorized depending on the kind of characteristics used in the recognition process, with a FER system employing one or both of these feature categories. The position of the facial organs and the texture of the skin give rise to the first set of characteristics. Another sort of feature is called a "geometric feature," which contains information about different facial locations and points and is used to assess a static picture or a series of images by leveraging movement of the various positions and points within the series." Geometric characteristics may be extracted from landmarks on the face as a starting point. There are a number of facial landmarks that may be used for facial analysis. Facial landmark recognition has been the focus of several research, although they are beyond the scope of this paper. These points are located using the Python programmedlib.

Aspects of artificial intelligence are engaged in the automated identification of human emotions and psychology. In the field of psychology and artificial intelligence, researchers are trying to address the issue of how to identify emotions. These include subjects such as mood and accent, which are created in both vocal and nonverbal sensors such as the tone and audible variations, which are easily accessible and may be used for mood evaluation as well as other methods. There were 55 percent sensory (emotional and linguistic) and 7 percent unidentified physical components of information in Mehrian's research results. Facial expressions are often the first sign of an individual's emotional condition, which is why so many academics are focused on it.

In order to add additional features to an existing representation, it may be beneficial to first focus on the extraction feature space. This assumption is made by Ekman and Friesen in their study of the Facial Action Coding System (FACS) and facial action units (AUs). It was Ekman and Friesen who discovered that FACS facial movements are coded for each individual's head motions when they discovered that the FACS facial AUs used FACS facial movement.

II. Literature Review

Tanoy Debnath et. al, (2022): It is possible to recognise human emotion using a variety of means, including human–computer interfaces, human emotional processing, illogical analysis, medical diagnostics, data–driven animation and communication between humans and robots, to name just a few. A convolutional neural network (CNN)-based face emotional recognition model is presented in this study. Emotions such as anger, contempt, fear, pleasure, neutrality, sorrow, and surprise may all be detected by our suggested "ConvNet" model. Our suggested CNN model trained on the features extracted from face expression photos using the Local Binary Pattern (LBP), region-based Oriented FAST and rotating BRIEF (ORB), and a Convolutional Neural Network (CNN) (ConvNet). Our approach is fast and effective, allowing the authors to swiftly construct a real-time schema that fits the model well and can detect emotions. In addition, the behavioral parts of this examination concentrate on a guy or woman's mental or emotional things. CNN network models are trained using the FER2013 databases, and then generalization methods are used to JAFFE and CK+ datasets to assess the model's performance. Generalization on the JAFFE dataset yields an accuracy of 92.05 percent and CK+ dataset yields an accuracy of 99.13 percent; these results are among the best among available approaches. We also evaluate the system's performance by analyzing real-time facial expressions. Four convolution layers and two fully linked layers make up ConvNet. According to the experiments, 96 percent of the time, the ConvNet can outperform current models in training accuracy. However, when compared to other techniques of validation, the one proposed was more accurate than the others.

Anil Audumbar Pise et. al, (2022): There has been a lot of progress in the field of automated face expression identification in recent years (FER). To improve human-machine interactions, FER has been applied in a variety of contexts, including human center computing and the emerging field of emotional artificial intelligence (EAI). Human facial expressions and behaviour may be predicted and analysed more accurately by computers with the help of researchers in the EAI sector. Due to the rapid evolution of neural networks and their ability to tackle more challenging problems, deep learning has had the largest impact on this sector. In this post, we'll look at some of the most recent developments in deep learning models for automated emotion identification. Both deep learning-based FERs and architecture-related models, such databases, may interact well to give extremely accurate results.

Chirag Dalvi et. al, (2021): Human thoughts and sentiments are reflected in the expressions on our faces. The spectator receives a slew of social indications from it, such as the subject's aim, focus, drive, and mood. For quiet communication, it is considered a powerful instrument. The study of these facial expressions reveals a great deal more about human nature. Researchers are increasingly turning to AI-based Facial Expression Detection (FER), which has applications in dynamic analysis, pattern recognition and many other areas. As a result of the Covid-19 epidemic, there has been an increased need for novel FER analytic frameworks that take use of the growing amount of visual data created by films and images. FER studies must also take into account the differences in facial expressions between children, adults, and the elderly. This is an area that has seen a great deal of investigation. However, there is a dearth of a complete literature review that highlights the previous work done and gives the aligned future paths of the field. There is a thorough overview of AI-based FER methodology, including datasets, feature extraction techniques, algorithms and the newest advances with their applications in facial expression detection, offered here by the authors of this work. Only one other review publication has stated all the characteristics of FER for different age groups, and this one is expected to greatly affect the research community in the next years, according to the author.

Francesca Nonis et. al, (2019): Human-computer interaction is one of the many domains where facial expression analysis and recognition (FER) has recently emerged as an active research issue. Problems relating to the nature of the data mean that solutions based on 2D models are not totally suitable for real-world applications. It is now possible to increase FER systems' accuracy by using 3D face data captured in both still photos and video. There are still significant shortcomings to 3D algorithms, making them useful only for a limited range of applications; in order to overcome these limits, multimodal 2D+3D analysis may be used. In this study, we examine the limitations and merits of standard and deep-learning FER methodologies, in order to provide the research community an overview of the outcomes achieved in the near future. The most often used databases to study facial expressions and emotions are also described in depth, showcasing the conclusions of the different writers. In this study, the various methods are evaluated, and some conclusions are formed about the best recognition rates attained.

Dhwani Mehta et. al, (2018): Because of the wide range of potential uses, emotion detection has become an inevitability and a difficult problem in computer science. Users get input and feelings from the system via nonverbal indicators including body language, facial expressions, and gestures. Human–Computer Interaction relies on the sensor's sensitivity and algorithmic resilience to improve recognition. Accurate detection relies heavily on sensors, which provide the system a higher level of efficiency and dependability by giving such high-quality input. It would be beneficial for the robots to learn social intelligence if they were able to recognise human emotions. An overview of several methods and strategies for recognizing emotions is provided in this work. It provides an overview of the

DOI:10.9756/INTJECSE/V14I5.677 ISSN: 1308-5581 Vol 14, Issue 05 2022

databases that may be used as data sets for facial expression-detection systems. To observe emotion detection in Augmented Reality, Microsoft HoloLens (MHL), a mixed reality gadget, will be offered later (AR). We'll go through a quick rundown of MHL's sensors, how they could be used in emotion identification, and some early findings. The research finishes by comparing the MHL's findings to those of a standard camera in terms of emotion recognition.

III. Face Detection Methods

FER's face detection phase is crucial. In order to automatically recognise the facial area in a static or video picture, a competent automated system may be constructed. Using facial traits such as edge, skin colour, texture, and face muscle action, a face area is identified in the picture series These characteristics make it easy to distinguish a face from the rest of the image. A face area and a non-face region are separated at this stage of image processing from the overall picture. Face identification approaches like eigenspace, adaptive skin colour, and Viola–Jones use Haar classifiers, Adaboost, and contour points in their algorithms. The study examines the accuracy with which people can recognise one another's faces, as well as how well they do in both restricted and unrestricted conditions.

Eigenspace Method

The eigenspace method was developed by Pentland et al. to find a face in a variety of positions. Modular Eigenspace Descriptor (MED) face recognition is also used. With the use of the eigenspace approach developed by Essa and Pentland, it was possible to find a face in any random picture sequence (PCA). An image subspace known as "facial space" is defined by the eigenforms. The projection coefficients and signal energy were used to measure the distance between the observed picture and the face space. For this purpose, a technology known as spatio-temporal filtering was applied. In order to analyze ‘motion blobs' over a period of time, the thresholding idea was used to the picture that was filtered. For each motion blob, the face is represented by a human head.

Using a library of 7562 photos of people of both sexes with occluded facial features like as hair, glasses, and other accessories, Pentland et al. developed a real-time technique that was put to the test.

Adaptive Skin Colour Method

In order to identify a person's face, the colour of the skin is an important factor. Any one of the colour systems is chosen based on colour dependence. RGB, CMY, YIQ, YUV, and YCbCr are the most prevalent colour schemes. For colour intelligence, YIQ and YUV colour systems, which eliminate the brightness impact during the processing step, are the most often utilized options. I refers to hue and Q refers to saturation in the YIQ colour model, and the formula to compute I is as follows:

$$I = 0.596 \times R - 0.274 \times G - 0.322 \times B$$

As a result of high mistakes caused by lighting and position variations, the majority of studies employ skin colour for face identification based on a set threshold method. The geometric pattern of the face is satisfied by an iterative thresholding technique. However, because of its high processing cost, it cannot be used for real-time applications. By using a linear discriminant function and adaptive threshold settings, we're able to distinguish between a complicated backdrop and an individual's skin colour. The lighting and posture variation may be influenced using the gamma correction approach. We have developed an adaptive skin colour and structure model for multiple photos in a complex backdrop, which greatly increases accuracy and effectively removes the influence of light intensity.

Haar Classifier Method

regarded as a reliable real-time face detection approach. A person's facial borders, lines, gestures, and skin colour may all be detected using a person's Haar characteristics. The Haar features, displayed in Fig. 1, are a black-and-white rectangular box used to extract features. In order to investigate the placements of Haar features, pixel intensities may be increased or decreased in various sections of a picture. The difference between the total of pixels in the black and white areas inside the rectangular box represents the value of the located feature. During the learning phase, the Haar classifier identifies the traits that cause issues with face identification. Detection accuracy increases when the computational burden and complexity of the messaging phase are reduced.

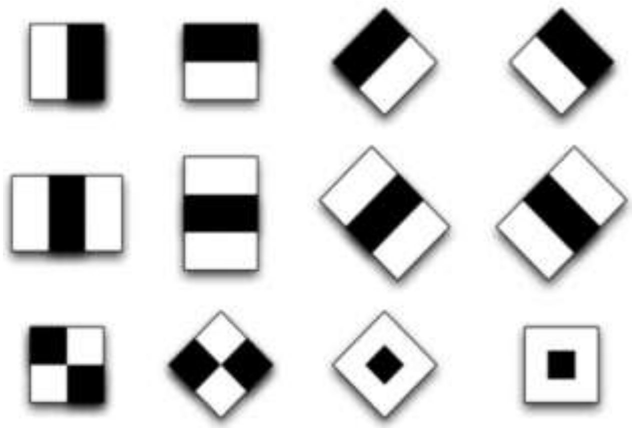

Figure 1: Haar Features

AdaBoost Method

Accuracy and computational complexity are the main advantages of the AdaBoost technique for face identification. A popular face detection algorithm with a low false positive rate is used. Adaboost is sensitive to noisy data and outliers, which is a big drawback. Classifiers in cascade utilizing Adaboost are used to train a collection of image characteristics in order to exclude the negative examples. Fig. 2 shows a cascade structure in which the first classifier's output is utilized as input for the second classifier, which is used to accurately determine the facial area. As a result, a powerful classifier was constructed that reduces the number of features and so improves the detection rate. A robust and hybrid face recognition method for colour and complicated pictures was suggested by Kheirkhah and Tabatabaie. The Adaboost-based face identification is combined with skin colour information in the hybrid technique. It is more accurate and takes less time to execute than other methods.

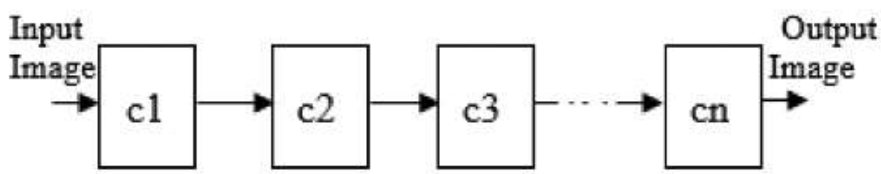

Figure 2: Structure of Cascade Classifier

Contours

Better accuracy is achieved when using contour points for face detection. When scanning an image series, the first pixel of the first frame is deemed to be the first contour point of the head, based on skin colour. Similarly, the remaining contour points are calculated inside a frame. Seed point refers to the pixel under consideration. It is possible to detect in a clockwise or counterclockwise direction by using a seed point to set the contour point's initialization route in the desired direction. To identify face motion, we look for a change in two consecutive frames and an increase or decrease in contour points over a certain threshold. Contour-based facial recognition was used by the researchers in this study. In order to get the desired facial shape, logical operations and Gaussian filters are used. In order to recognise and track a face from an image sequence, the scalar and vector distance of a rectangle window created from four corner points of two successive frames is determined.

IV. Feature Extraction Techniques

FER's next stage is feature extraction after face detection. Extracting facial features without losing face information is the primary goal of facial feature extractions. The face motion and distortion of facial features are used to classify two types of feature extraction techniques: geometric and appearance-based.

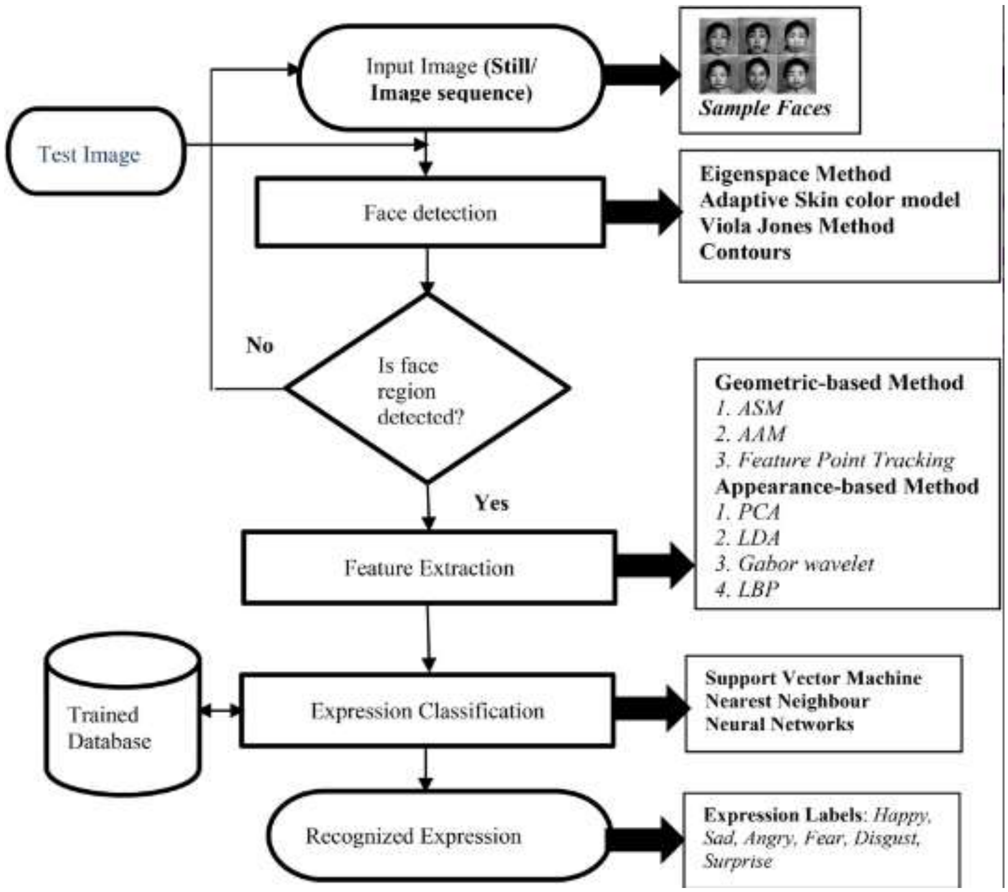

Figure 3: Flowchart for FER

Comparative Analysis and Discussion on FERWe made it clear in this paper that there is a strong demandfor expanded research on the topic of FER far beyondshallow learning techniques. We full form of the automaticFER goes through four main data processing, a few proposedarchitectures, and finally getting to the main model'semotion recognition. Many preprocessing techniques werementioned in this review, such as the cropping and resizingof images to speed up training and normalization, as well asthe overfitting of data. Lopes et al. [120] have presented allthese techniques well, according to Lopes and his colleagues.Figure 9 illustrates deep learning-based FER models. Thevarious methods and contributions presented in this reviewachieved high levels of accuracy. Moselhi et al. [121] demonstrated the key significance of the use of neural networksand connectivity expansion layers in neural network architectures. Like many before them, the authors Mohammadpour et al. [122] prefer to extract AU from the face,rather than do face-to-recognition first, first. is study isbeing conducted to determine whether occlusion imagesexist or not, as well as to try to gain greater insight into thenetwork. Pise et al. [8] have examined the incorporation ofthe leftover blocks. While text images allow only for largereyes and smaller faces, the addition of the iconized face to thenetwork improves accuracy when using small images, as isdemonstrated by Yolu and Ayiv [108]. Two-concept CNNarchitecture expansion was added after a long and thoroughanalysis of the recognition rate by offering two more featurearticles to know the impact of CNN parameters. Favorableresults have been observed in most of the methodsattempted, which means more than 90% of these projectsachieved some level of success. Researchers who studyspatial and temporal features first provided several combinations: the combination of CNN-L and 3D-CNN is typically applied to give a boost to spatial features but booststemporal features too. As can be demonstrated according tothe work of Yu and his colleagues, the methods proposed byKim et al. [118] and Liang et al. [111] provide a better level ofprecision than the one that was performed by the Kim group[118]. at comes out to an effective volume expansionfactor of about 99%.In CNN applications, both temporal and spatial networks have demonstrated their accuracy. At is why theresearchers chose LSTM, which is effective for sequentialdata in general, but especially for time-dependent data inorder to achieve high accuracy in FER. To date, CNNparametric modeling and the most difficult algorithms usedby CNN researchers are softmax and Adam optimization. Tovalidate the neural network architecture, we also tested themodel across multiple databases, and our findings indicatethat there are no significant differences in results.

Table 1 summarizes the arguments made previously,with particular emphasis on the architecture, database, andrecognition rate discussed in the linked articles.comparative Analysis and Discussion on FERWe made it clear in this paper that there is a strong demandfor expanded research on the topic of FER far beyondshallow learning techniques. We full form of the automaticFER goes through four main data processing, a few proposedarchitectures, and finally getting to the main model'semotion recognition. Many preprocessing techniques werementioned in this review, such as the cropping and resizingof images to speed up training and normalization, as well asthe overfitting of data. Lopes et al. [120] have presented allthese techniques well, according to Lopes and his colleagues.Figure 9 illustrates deep learning-based FER models. Wevarious methods and contributions presented in this reviewachieved high levels of accuracy. Moselhi et al. [121] demonstrated the key significance of the use of neural networksand connectivity expansion layers in neural network architectures. Like many before them, the authors Mohammadpour et al. [122] prefer to extract AU from the face,rather than do face-to-recognition first, first. Is study isbeing conducted to determine whether occlusion imagesexist or not, as well as to try to gain greater insight into thenetwork. Pise et al. [8] have examined the incorporation ofthe leftover blocks. While text images allow only for largereyes and smaller faces, the addition of the iconized face to thenetwork improves accuracy when using small images, as isdemonstrated by Yolu and Ayiv [108]. Two-concept CNNarchitecture expansion was added after a long and thoroughanalysis of the recognition rate by offering two more featurearticles to know the impact of CNN parameters. Favorableresults have been observed in most of the methodsattempted, which means more than 90% of these projectsachieved some level of success. Researchers who studyspatial and temporal features first provided several combinations: the combination of CNN-L and 3D-CNN is typically applied to give a boost to spatial features but booststemporal features too. As can be demonstrated according tothe work of Yu and his colleagues, the methods proposed byKim et al. [118] and Liang et al. [111] provide a better level ofprecision than the one that was performed by the Kim group[118]. At comes out to an effective volume expansionfactor of about 99%.In CNN applications, both temporal and spatial networks have demonstrated their accuracy. At is why theresearchers chose LSTM, which is effective for sequentialdata in general, but especially for time-dependent data inorder to achieve high accuracy in FER. To date, CNNparametric modeling and the most difficult algorithms usedby CNN researchers are softmax and Adam optimization. Tovalidate the neural network architecture, we also tested themodel across multiple databases, and our findings indicatethat there are no significant differences in results.Table 1 summarizes the arguments made previously,with particular emphasis on the architecture, database, andrecognition rate discussed in the linked articles.Comparative Analysis and Discussion on FER.

We made it clear in this paper that there is a strong demandfor expanded research on the topic of FER far beyondshallow learning techniques. We full form of the automaticFER goes through four main data processing, a few proposedarchitectures, and finally getting to the main model'semotion recognition. Many preprocessing techniques werementioned in this review, such as the cropping and resizingof images to speed up training and normalization, as well asthe overfitting of data. Lopes et al. [120] have presented allthese techniques well, according to Lopes and his colleagues.Figure 9 illustrates deep learning-based FER models. Wevarious methods and contributions presented in this reviewachieved high levels of accuracy. Moselhi et al. [121] demonstrated the key significance of the use of neural networksand connectivity expansion layers in neural network architectures. Like many before them, the authors Mohammadpour et al. [122] prefer to extract AU from the face,rather than do face-to-recognition first, first. is study isbeing conducted to determine whether occlusion imagesexist or not, as well as to try to gain greater insight into thenetwork. Pise et al. [8] have examined the incorporation ofthe leftover blocks. While text images allow only for largereyes and smaller faces, the addition of the iconized face to thenetwork improves accuracy when using small images, as isdemonstrated by Yolu and Ayiv [108]. Two-concept CNNarchitecture expansion was added after a long and thoroughanalysis of the recognition rate by offering two more featurearticles to know the impact of CNN parameters. Favorable

results have been observed in most of the methodsattempted, which means more than 90% of these projectsachieved some level of success. Researchers who studyspatial and temporal features first provided several combinations: the combination of CNN-L and 3D-CNN is typically applied to give a boost to spatial features but booststemporal features too. As can be demonstrated according tothe work of Yu and his colleagues, the methods proposed byKim et al. [118] and Liang et al. [111] provide a better level ofprecision than the one that was performed by the Kim group[118]. At comes out to an effective volume expansionfactor of about 99%.In CNN applications, both temporal and spatial networks have demonstrated their accuracy. At is why theresearchers chose LSTM, which is effective for sequentialdata in general, but especially for time-dependent data inorder to achieve high accuracy in FER. To date, CNNparametric modeling and the most difficult algorithms usedby CNN researchers are softmax and Adam optimization. Tovalidate the neural network architecture, we also tested themodel across multiple databases, and our findings indicatethat there are no significant differences in results.Table 1 summarizes the arguments made

DOI:10.9756/INTJECSE/V14I5.677 ISSN: 1308-5581 Vol 14, Issue 05 2022

previously,with particular emphasis on the architecture, database, andrecognition rate discussed in the linked articles.

V. Comparative Analysis and Discussion on Fer

There is a tremendous need for more study on the issue of FER that goes beyond shallow learning approaches, as we made apparent in this article Four primary data processing, many potential architectures and eventually to the main model of emotion recognition are all necessary to get to the complete form of the automated FER Preprocessing strategies such as cropping and shrinking photos to speed up training and normalization, and overfitting data were discussed extensively in this paper. In the opinion of Lopes and his colleagues, Lopes et al. have well-presented all of these strategies. Accuracy was obtained using a variety of methodologies and contributions mentioned in this study. In neural network designs, Moselhi et al. proved the critical importance of using neural networks and connection expansion layers. The writers Mohammad Zadeh et al. choose to extract AU from the face, rather than perform face-to-recognition first, in order to assess whether or not occlusion pictures exist and to acquire a better understanding of the network. An investigation of the incorporation of remaining blocks has been conducted by Pise and colleagues. Yolu and Ayiv show how adding an iconized face to the network enhances accuracy when tiny pictures are used, while text images only allow for bigger eyes and smaller faces. Expanding the CNN architecture to include two concepts came about as a result of an extensive study of the recognition rate in which two more feature articles were published to better understand the influence of CNN parameters. Many strategies have had positive outcomes and more than 90 percent of these initiatives have had some degree of success. For starters, researchers that focus on spatial and temporal aspects have suggested the following possible pairings. It is common to use CNN-L and 3D-CNN together to enhance spatial characteristics, but it also works to enhance temporal ones as well. When it comes to determining the volume expansion factor, Yu and his colleagues have found that the techniques suggested by Kim and Liang offer a superior degree of accuracy than the one achieved by the Kim group. To attain high accuracy in FER, the researchers turned to LSTM, which works well with sequential data in general but particularly well with time-dependent data. This is why the researcher's used LSTM in CNN applications. SoftMax and Adam optimization have been the most demanding algorithms utilized by CNN researchers to date when it comes to CNN parametric modelling. We also evaluated the neural network design across several datasets, and our findings show no major changes in outcomes.

VI. Conclusion

A Comparative Study and Discussion of the FER There is a tremendous need for more study on the issue of FER that goes beyond shallow learning approaches, as we made apparent in this article Four primary data processing, many potential architectures and eventually to the main model of emotion recognition are all necessary to get to the complete form of the automated FER Preprocessing strategies such as cropping and shrinking photos to speed up training and normalisation, and overfitting data were discussed extensively in this paper. According to Lopes and his colleagues' numerous methodologies and contributions provided in this review, great levels of accuracy have been attained using the strategies presented by Lopes et al. Key to the success of neural network designs is the employment of neural networks and connection expansion layers. Although Mohammadpour et a research's is focused on finding out whether or not there are pictures that are obscured by occlusions, it is also trying to obtain a better understanding of how the network works. An investigation of the incorporation of remaining blocks has been conducted by Pise and colleagues. Yolu and Ayiv show how adding an iconized face to the network enhances accuracy when tiny pictures are used, while text images only allow for bigger eyes and smaller faces. Expanding the CNN architecture to include two concepts came about as a result of an extensive study of the recognition rate in which two more feature articles were published to better understand the influence of CNN parameters. Many strategies have had positive outcomes and more than 90 percent of these initiatives have had some degree of success. The combination of CNN-L and 3D-CNN is often utilized to offer a boost to spatial characteristics but enhances temporal features as well, according to researchers that investigate spatial and temporal features. These newer approaches, which were developed in part by Yu and his colleagues, had more accuracy than the Kim group's, which had an effective volume expansion factor that was roughly 95%. This was proved by Yu and his colleagues in their study. To attain high accuracy in FER, the researchers turned to LSTM, which works well with sequential data in general but particularly well with time-dependent data. This is why the researcher's used LSTM in CNN applications. SoftMax and Adam optimization have been the most demanding algorithms utilized by CNN researchers to date when it comes to CNN parametric modeling. We also evaluated the neural network design across several datasets, and our findings show no major changes in outcomes.

References

[1] Debnath, T., Reza, M.M., Rahman, A. et al. Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity. Sci Rep., 12, 6991 (2022). https://doi.org/10.1038/s41598-022-11173-0

DOI:10.9756/INTJECSE/V14I5.677 ISSN: 1308-5581 Vol 14, Issue 05 2022

[2] Anil Audumbar Pise, Mejdal A. Alqahtani, Priti Verma, Purushothama K, Dimitrios A. Karras, Prathibha S, Awal Halifa, "Methods for Facial Expression Recognition with Applications in Challenging Situations", Computational Intelligence and Neuroscience, vol. 2022, Article ID 9261438, 17 pages, 2022. https://doi.org/10.1155/2022/9261438

[3] Chirag Dalvi et. al, (2021)," A Survey of AI-Based Facial Emotion Recognition: Features, ML & DL Techniques, Age-Wise Datasets and Future Directions" Digital Object Identifier. 10.1109/ACCESS.2021.3131733

[4] Francesca Nonis et. al, (2019)," 3D Approaches and Challenges in Facial Expression Recognition Algorithms—A Literature Review" doi:10.3390/app9183904

[5] Mehta, Dhwani et al. "Facial Emotion Recognition: A Survey and Real-World User Experiences in Mixed Reality." Sensors (Basel, Switzerland) vol. 18,2416. 1 Feb. 2018, doi:10.3390/s18020416

[6] Anil J., Suresh L.P. Literature survey on face and face expression recognition; Proceedings of the 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT); Nagercoil, India. 18–19 March 2016; pp. 1–6. [Google Scholar]

[7] Kumari J., Rajesh R., Pooja K. Facial expression recognition: A survey. Procedia Comput. Sci., 2015; 58:486–491. Doi: 10.1016/j.procs.2015.08.011.

[8] Cohn J.F., Zlochower A.J., Lien J.J., Kanade T. Feature-point tracking by optical flow discriminates subtle differences in facial expression; Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition; Nara, Japan, 14–16 April 1998; pp. 396–401.

[9] Li C., Soares A. Automatic Facial Expression Recognition Using 3D Faces. Int. J. Eng. Res. Innova., 2011; 3:30–34.

[10] Dhall A., Goecke R., Joshi J., Hoey J., Gedeon T. Emotiw 2016: Video and group-level emotion recognition challenges; Proceedings of the 18th ACM International Conference on Multimodal Interaction; Tokyo, Japan. 12–16 November 2016; New York, NY, USA: ACM; 2016, pp. 427–432.

[11] Mohammed A.A., Minhas R., Wu Q.J., Sid-Ahmed M.A. Human face recognition based on multidimensional PCA and extreme learning machine. Pattern Recognin., 2011; 44:2588–2597. doi: 10.1016/j.patcog.2011.03.013.

[12] Rivera A.R., Castillo J.R., Chae O.O. Local directional number pattern for face analysis: Face and expression recognition. IEEE Trans. Image Process., 2013; 22:1740–1752. doi: 10.1109/TIP.2012.2235848.

[13] De Marsico M., Nappi M., Riccio D. FARO: Face recognition against occlusions and expression variations. IEEE Trans. Syst. Man Cybern. Part A Syst. Hum., 2010; 40:121–132. doi: 10.1109/TSMCA.2009.2033031.

[14] Alex A.T., Asari V.K., Mathew A. Gradient feature matching for expression invariant face recognition using single reference image; Proceedings of the 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC); Seoul, Korea. 14–17 October 2012; Piscataway, NJ, USA: IEEE; 2012. pp. 851–856.

[15] Kahou S.E., Pal C., Bouthillier X., Froumenty P., Gülçehre Ç., Memisevic R., Vincent P., Courville A., Bengio Y., Ferrari R.C., et al. Combining modality specific deep neural networks for emotion recognition in video; Proceedings of the 15th ACM on International Conference on Multimodal Interaction; Sydney, Australia. 9–13 December 2013; New York, NY, USA: ACM; 2013, pp. 543–550.

[16] Yu Z., Zhang C. Image based static facial expression recognition with multiple deep network learning; Proceedings of the 2015 ACM on International Conference on Multimodal Interaction; Seattle, WA, USA. 9–13 November 2015; New York, NY, USA: ACM; 2015, pp. 435–442.