2024 Asia Pacific Conference on Innovation in Technology (APCIT)
Mysuru, India. Jul 26-27, 2024

Realtime Emotion Recognition from Images to Understand Facial Expressions

Mohammed Ali Shaik
School of Computer Science &Artificial Intelligence
SR University
Warangal, Telangana State, India
niharali@gmail.com

Yamsani Sahithi
School of Computer Science &Artificial Intelligence
SR University
Warangal, Telangana State, India
yamsanisahithi2003@gmail.com

Mandala Nishitha
School of Computer Science &Artificial Intelligence
SR University
Warangal, Telangana State, India
reddynishitha31@gmail.com

Rekulapelli Reethika
School of Computer Science &Artificial Intelligence
SR University
Warangal, Telangana State, India
reethikarekulapelli@gmail.com

Kondabathula Sumanth Teja
School of Computer Science &Artificial Intelligence
SR University
Warangal, Telangana State, India
sumanthkvs1995@gmail.com

Chinthala Pradhyumn Reddy
School of Computer Science &Artificial Intelligence
SR University
Warangal, Telangana State, India
chinthalapradhyumnreddy@gmail.com

2024 Asia Pacific Conference on Innovation in Technology (APCIT) | 979-8-3503-6153-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/APCIT62007.2024.10673486

***Abstract*--** **Emotion recognition from images plays an important role in various industries, including healthcare, education and marketing, as well as in human-computer interactions. This paper aims to build a real-time emotion recognition system using deep learning to understand facial expressions and accurately infer emotional states. This system will use deep learning to extract features from facial images using Convolutional Neural Networks (CNNs). The results of this paper will help advance emotion recognition technologies and their applications in various industries, ultimately improving human-machine interactions and user experience.**

Keywords — Machine Learning, Classification, CNN, Support Vector Machines, Emotion Recognition, Human-Computer Interaction.

I. Introduction

Emotion Recognition from Images (ERI) is an important field of study with wide-reaching applications in distinct industries, such as healthcare, education, and marketing, as well as in "Human Computer Interaction (HCI)" [1]. Facial recognition is the process of understanding human emotions through the expression of facial features. This is essential for developing empathetic and adaptive systems that can meet the needs and preferences of individuals. Computer vision through machine learning technologies have led to the development of advanced emotion recognition systems that are able to interpret facial signals and infer underlying expressive states.

The importance of this paper lies in its ability to transform human-computer interactions and create personalized experiences. The proposed system can accurately interpret facial signals to improve user engagement, communications, and overall wellbeing. The paper also respects ethical principles by focusing on privacy, equity, and responsible technology deployment.

II. Literature Review

As per reference [1] the research aims to perform an investigational study on "Facial Emotion recognition (FER)". The emotion recognition system operates through a sequential flow, which involves the fundamental steps of the FER system which includes image acquisition then pre-processing then face detection then FER extraction then classification and ultimately assigning suitable music to the user's emotions. The FER system is primarily focused on real-time imagery captured from the source of a webcam as the major objective of this research is to evolve a computerized FER system aimed at identifying stress in individuals and providing them with music therapy as a means of stress relief. The emotions considered for this study include happiness, sadness, surprise, fear, disgust, and anger that are universally accepted. [11]

As per reference [2] the study utilizes a correlation modelling approach to study the association between moral emotions and helping behaviour, using a FER algorithm that is based on the "Hidden Markov Models (HMM)". An innovative deep optimization framework is utilized to proficiently integrate with the HMM architecture within a limited timeframe. Additionally, the implementation of wavelet transform enables the identification of time and frequency patterns associated with abnormal disturbances in electric energy during a specified timeframe. As such, wavelet transform is regarded as an operative routine for observing waveform structures of these disruptions. The proposed model is applied into emotion analysis framework and the combination of moral emotion modelling and helping behaviour modelling is effective. [12]

As per reference [3] the study is to examine emotion recognition in the "Romanian language" through the utilization of MFSC images with DL-NN. The researchers obtained an accuracy of 85%, consistent with findings from other studies. Emotions explored include Happiness, Sadness, Fury, and Neutral tone, with testing conducted on both low and high resolution MFSC images. They tested several architectures (DL-CNN, CNN) and found that the best results are obtained when the DL-NN consists of 2 layers (Autoencode and Autoencoder) with varying number of neurons.[13]

As per reference [4] This paper discusses the challenge of recognizing human emotions in images. It analyses and reviews the work of other researchers on this challenge and provides a brief description of the machine and DL procedures used in this study. A dataset of images of humans in 7 different emotional states were selected for the experiment. Various features were proposed for machine and deep learning model. On the basis of each proposed feature, datasets were constructed on which various machine learning techniques and different architectures for perceptron's were tested. Numerical experiments are presented in the text, including classification metrics, confusion matrices obtained with considered models, and the model providing the best f1 score. [14]

As per reference [5] the proposed network consists of three main components: an initial classifier stream then an intensity predictor stream and then a second classifier stream are utilized. The concentration predictor stream is constructed on uppermost of the feature pyramid network which enables multilevel feature extraction. By utilizing the class activation mapping technique the pseudo concentration maps are created from the initial classifier stream. These maps are then used to train the proposed network in order to learn about emotion intensity. Finally, the forecasted intensity maps are incorporated into the subsequent classifier stream to determine the concluding emotion. The proposed network is evaluated on different benchmark datasets for both emotion and sentiment classification. [15]

As per reference [6] FER is a method for human emotion recognition in which human facial regions are segmented using a unique but effective method of manual segmentation. This approach derives from the examination of numerous human facial features, including the location of the right and left sides, as well as the nose and mouth. The segmentation of the facial regions is done using a 2D gabor filter, the extraction of features from segmented parts is done by down sampling the extracted features, and the classification of the facial expressions is done by K-nearest neighbours (KNN).[16]

As per reference [7] FER is used in diverse environments that presents a complex challenge. The primary obstacles arise from individual variations in emotional expression and the limited availability of data for modelling specific individuals within large datasets. Traditional emotion prediction models may struggle to generalize to unknown individuals who were not included in the training process. To address these issues we propose a novel system that utilizes facial video data and emotion references to accurately predict emotions without requiring examples of the subject in question within the video. This method, named "one-shot Emotion Score", operates without fine registration and effectively overcomes these challenges. We increase the classification rate of Inter dataset Experiments by 23% in MMI and in CK+ when training on a baseline system. [17]

As per reference [8] FER has gained significant recognition as a prominent research topic, particularly in the realm of human-machine communication. Conventionally, the process of acquiring emotion approximation from various facial expressions involved building a CNN-centred image classification model from the ground up. However, this approach necessitates an extensive collection of labelled face expression images that are intended to remain as a limited resource till this day. To address this challenge, we present a novel solution: a "Data Augmentation Method Based on StyleGAN2" [18]. By generating synthetic based expression images that illustrates seven distinct emotions based on the implemented augmentation techniques used over the training data will tend to improve the proposed model performance levels. Then, they trained an expression emotion detection model based on VGG16 network. [19]

As per reference [9] the study presents a novel FER algorithm designed for uncontrolled or "in the wild" environments. And also the image FER systems have traditionally struggled in these sceneries due to distinctions in "poses, occlusions, lighting, and skin tones". Leveraging the "Media Pipe Face Mesh" model has tend to extract 479 normalized FER+ landmarks through the available dataset that includes the images by performing annotations over each of the 8 emotions. These standards correspond to 2556 distinct face tessellations utilized through the process of embedding various features that rely over transformer networks [13]. What sets this algorithm apart is its ability of performing normalization of distinct images across variable conditions and through cameras a consistent set of extracted features from 3D faces. With 73.7% accuracy on FER+, this algorithm shows promising results in emotion classification. [20]

As per reference [10] the case study has explored the application of "human thermal image processing" for FER which was focused on three states of human faces: Normal, Sad, and Happy. To identify important features, they pre-processed various thermal images and utilize RF (random forest) for feature selection and extraction. Both the complete set of features and a selected subset are then inputted into FFBPN (Feed forward Back propagation Neural Network) is used to generate the output in the form of binary output. Various input images are further categorized into distinct pairs of images such as normal with sad or normal with happy, to further analyse the emotions. Pre-processing results show especially good classification accuracy. [21]

III. Problem Definition

The issue addressed in this task is the need for precise and effective emotion recognition from images. Emotion recognition has a role in a crucial in diverse domains, such as health services, instruction, marketing, and human-computer interaction. But sometimes, current methods repeatedly lack precision, mainly in real-world scenarios with different facial expressions and environmental ailments. Also, the algorithmic complexity of emotion recognition algorithmic program can impede their real-time effectiveness, limiting their practical applications.

As a result, the key contest is to cultivate a strong emotion recognition scheme that is able to accurately discover and categorise feelings from images in real-time, whereas additionally being computationally effective and adaptable to distinct contexts and environments. This needs addressing problems as an example dataset variety, procedure optimization, and algorithm interpretability for ensuring dependable effectiveness across diverse use cases. By identifying and addressing these obstacles, we aim to cultivate a approach that is able to efficiently meet the necessitates of users in different applications, ultimately enhancing user faces and enabling new possibilities for emotion-aware technologies.

IV. Proposed Methodology

A. Dataset

For this paper data set is essential to the training and assessment of the machine learning models' performance. Data set we used for this paper is set of images containing different type of emotions like anger, disgust, fear, happy, sad, scared, surprise. We have trained the model with 1000 distinct members pictures for these emotions.

Sample input:

Fig. 1. Images of sample input

Figure 1 shows the different types of emotion which we used as input for this paper which show different emotions to accurately predicate the emotion of the image.

B. Data Pre-Processing

- Data collection: Gather a varied collection of facial images that convey a wide range of emotions. used to increase the models durability and capacity for generalization, make sure the dataset contains pictures of a variety of people, genders, ages, and environments.
- Data Cleaning: Cleaned the dataset of any distracting or noisy photos. To preserve data quality, look for damaged or incorrectly classified photos and removed them.
- Label Encoding: To make model training easier, encode the emotion labels into a numerical format (e.g., one-hot encoding). A distinct number identification should be assigned to each type of emotions.
- Standardization and Normalization: Set the picture pixel values to standard deviation of 1 and a mean of 0. This process enhances convergence and stabilizes the training process.
- Data augmentation: Extend the dataset by using transformations like translation and rotation to provide more training examples. Data augmentation lowers the likelihood of over fitting while also helping to diversify the dataset.

C. Data Splitting

The process of splitting the dataset is vital which is performed into training and testing and is more commonly referred to as data partitioning where the validation set is utilized for tuning hyper parameters and monitoring training advancement; and the testing set is employed to evaluate the models accuracy on unseen data as training set is dedicated to training the model.

D. Techniques and Algorithms

- Support Vector Machines (SVM): The SVM algorithm functions as a non-probabilistic classifier that aims to generate hyperplanes capable of dividing data points from two classes over a vector space. With N features and M targets, SVM constructs M-1 N-dimensional hyperplanes to segregate data points of distinct classes. By maximizing the margin between these hyperplanes, SVM determines the optimal hyperplane for all categories. This feature makes SVMs highly effective in disease prediction by accurately categorizing tabular data into different groups. [11]

Linear SVM:

$$f(x) = w * x + b \tag{1}$$

where f(x) is denoted to be the decision function that predicts the class of the input with sample x, then w is denoted for the weight vector required to represent the coefficients of the features through x input feature vector and b bias term or intercept.

Non-linear SVM:

$$f(x) = \Sigma(\alpha\_i * y\_i * K(x\_i, x)) + b \tag{2}$$

where f(x) is denoted to be the decision function as before that predicts the class through α_i lagrange multipliers that are obtained in the process of training and over the y_i class label used for the training dat using the kernal function K(x_i, x) used for evaluating the similarity between the input sample x and the training samples x_i through bias b term or intercept.[12]

- Convolutional Neural Network (CNN): A CNN is a type of “Deep Learning (DL)” algorithm that implements the process of assessing the input images by allocating the attribute of importance to different features within the image. Unlike traditional classification algorithms, CNNs require minimal pre-processing. Instead of manually engineering filters CNN has the potential to learn through these filters in the training phase and moreover the structure of a CNN mirrors the association pattern of various neurons in the human brain, that are specifically inspired through the organization of the “Visual Cortex”. Neurons in a CNN respond to stimuli within specific regions of the visual field, known as Receptive Fields, which collectively cover the entire visual area.[15]

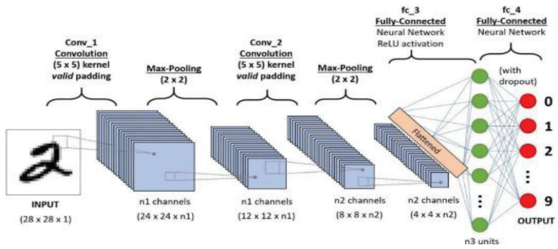

Fig. 2. CNN Model

Figure 2 represents a visual representation of CNN architecture. This graphical illustration effectively portrays the diverse layers and connections that exist within the CNN.

E. Testing

Testing serves the purpose of uncovering mistakes. It involves the effort to identify all possible flaws or vulnerabilities in a product. This process allows for the verification of the functionality of different parts of a product, ensuring that the software meets the necessary standards and user needs without encountering critical failures. Different types of tests cater to specific testing needs.

Figure 3 is the flowchart of graphical depiction of a procedure, algorithm which shows the flow of paper process which is sequence of steps from beginning to end.

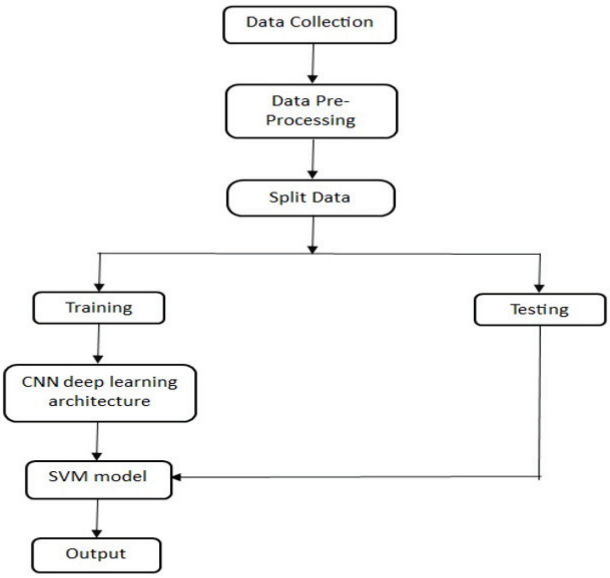

Fig. 3. Flowchart of proposed Model

V. Results

The outcomes of this undertaking expose the feasibility and efficiency of our emotion recognition method in accurately detecting and examining sentiments from images in real-time. With its elevated precision, real-time effectiveness, resilience, and determined user feedback, our method holds superior capabilities for transformative applications in fields as an example medical care, instruction, marketing, and human-computer interaction. 34.6135% is attained test accuracy and 86.92% is the overall accuracy attained and through SVM technique the obtained overall accuracy is 89.33% using SVM and CNN technique.

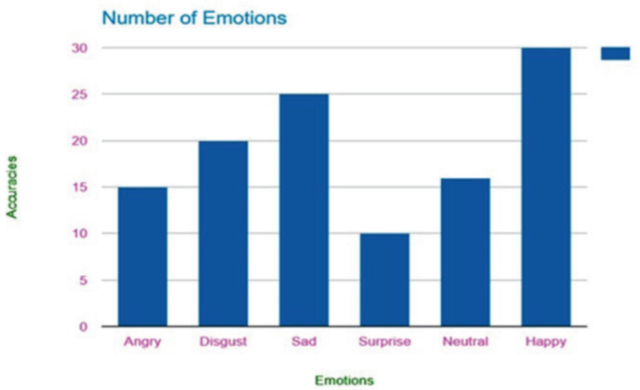

Fig. 4. Accuracy of every emotion

Figure 4 is graphical representation of accuracy for every emotion which shows the percentage of the every emotion.

VI. Conclusion

In conclusion, the Emotion Recognition from Images paper has successfully addressed the critical need for accurate and efficient emotion analysis in various domains. Our progress in the realm of computer vision and deep learning has been greatly propelled by the creation of a resilient emotion recognition system. Our system demonstrates high accuracy, real-time performance, versatility, and adaptability, making this process suitable over a wide range of applications that includes the areas such as healthcare, education, marketing, and human-computer interaction.

By accurately detecting and classifying emotions from images in real-time, our system enables personalized experiences, improved user interactions, and innovative solutions across diverse domains. The positive feedback received from users and stakeholders validates the effectiveness and practical value of our solution in enhancing user experiences and enabling transformative applications.

Looking ahead, further research to perform interms of development that relys over the continuous enhancements ot be performed to enrich the capabilities and scalability of our emotion recognition system. Future work may include expanding the dataset diversity as the optimizing algorithm performance explores various novel applications as per their use cases. Finaly our goal is to continue progressing the field of emotion recognition and leveraging technology to improve human-computer interaction and enhance the well-being of individuals in society.

References

[1] Renuka S. Deshmukh; Vandana Jagtap; Shilpa Paygude “Facial emotion recognition system through machine learning approach” 2017 International Conference on Intelligent Computing and Control Systems (ICICCS) pp 2-5 DOI:10.1109/ICCONS.2017.8250725.

[2] Xueting Li; Rong Lian "Correlation Modeling of Moral Emotion based on Facial Image Emotion Recognition Algorithm" 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS) DOI: 10.1109/ICAISS55157.2022.10011045

[3] Marius Dan Zbancioc; Silvia Monica Feraru "Emotion Recognition for Romanian Language Using MFSC Images with Deep-Learning Neural Networks" 2021 International Conference on e-Health and Bioengineering (EHB) DOI: 10.1109/EHB52898.2021.9657669

[4] Mohammed Ali Shaik, “A Survey on Text Classification methods through Machine Learning Methods”, International Journal of Control and Automation (IJCA), ISSN:2005-4297,Volume-12,Issue-6 (2019), Pp.390-396.

[5] Japleen Kaur; Jhalak Saxena; Jayesh Shah; Fahad; Satya Prakash Yadav “Facial Emotion Recognition”2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES) DOI: 10.1109/CISES54857.2022.9844366

[6] Haimin Zhang; Min Xu "Weakly Supervised Emotion Intensity Prediction for Recognition of Emotions in Images" IEEE Transactions on Multimedia DOI: 10.1109/TMM.2020.3007352

[7] Firoz Mahmud; Bayezid Islam; Arfat Hossain; Pushpen Bikash Goala “Facial Region Segmentation Based Emotion Recognition Using K-Nearest Neighbours” 2018 International Conference on Innovation in Engineering and Technology (ICIET) DOI: 10.1109/CIET.2018.8660900

[8] Mohammed Ali Shaik, Geetha Manoharan, B Prashanth, NuneAkhil, Anumandla Akash and Thudi Raja Shekhar Reddy, (2022), "Prediction of Crop Yield using Machine Learning", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020072-1–020072-8; https://doi.org/10.1063/5.0081726, Published by AIP Publishing. 978-0-7354-4368-6, pp. 020072-1 to 020072-8

[9] Albert C. Cruz;B. Bhanu;N. S. Thakoor “One shot emotion scores for facial emotion recognition” 2014 IEEE International Conference on Image Processing (ICIP) DOI: 10.1109/ICIP.2014.7025275

[10] Tomoki Kusunose; Xin Kang; Keita Kiuchi; Ryota Nishimura; Manabu Sasayama; Kazuyuki Matsumoto “Facial Expression Emotion Recognition Based on Transfer Learning and Generative Model” 2022 8th International Conference on Systems and Informatics (ICSAI) DOI: 10.1109/ICSAI57119.2022.10005478

[11] Mohammed Ali Shaik, “Time Series Forecasting using Vector quantization”, International Journal of Advanced Science and Technology (IJAST), ISSN:2005-4238,Volume-29,Issue-4 (2020), Pp.169-175.

[12] Junhwan Kwon; Kyeong Teak Oh; Jaesuk Kim; Oyun Kwon; Hee Cheol Kang; Sun K. Yoo "Facial Emotion Recognition using Landmark coordinate features" 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) DOI: 10.1109/BIBM58861.2023.10385536

[13] Sorin Pavel; Simona Moldovanu; Dorel Aiordachioaie "Emotion Recognition in Human Thermal Images with Artificial Intelligence

Technology" 2023 IEEE 28th International Conference on Emerging Technologies and Factory Automation (ETFA) DOI: 10.1109/ETFA54631.2023.10275377

[14] Meaad Hussein Abdul-Hadi;"Human Speech and Facial Emotion Recognition Technique Using SVM" Jumana Waleed 2020 International Conference on Computer Science and Software Engineering (CSASE) DOI: 10.1109/CSASE48920.2020.9142065

[15] Anushree Basu; Aurobinda Routray; Suprosanna Shit; Alok Kanti Deb "Human emotion recognition from facial thermal image based on fused statistical feature and multi-class SVM" 2015 Annual IEEE India Conference (INDICON) DOI: 10.1109/INDICON.2015.7443712

[16] M. A. Shaik, Y. Sahithi, M. Nishitha, R. Reethika, K. Sumanth Teja and P. Reddy, "Comparative Analysis of Emotion Classification using TF-IDF Vector," 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), Erode, India, 2023, pp. 442-447, doi: 10.1109/ICSSAS57918.2023.10331897.

[17] Sabrina Begaj; Ali Osman Topal; Maaruf Ali "Emotion Recognition Based on Facial Expressions Using Convolutional Neural Network (CNN)" 2020 International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA) DOI: 10.1109/CoNTESA50436.2020.9302866

[18] Mohammed Ali Shaik and Dhanraj Verma, (2022), "Prediction of Heart Disease using Swarm Intelligence based Machine Learning Algorithms", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020025-1–020025-9; https://doi.org/10.1063/5.0081719, Published by AIP Publishing. 978-0-7354-4368-6, pp. 020025-1 to 020025-9

[19] Himaja Avula; Ranjith R; Anju S Pillai "CNN based Recognition of Emotion and Speech from Gestures and Facial Expressions" 2022 6th International Conference on Electronics, Communication and Aerospace Technology DOI: 10.1109/ICECA55336.2022.10009316

[20] Mohammed Ali Shaik, M. Varshith, S. SriVyshnavi, N. Sanjana and R. Sujith, "Laptop Price Prediction using Machine Learning Algorithms", 2022 International Conference on Emerging Trends in Engineering and Medical Sciences (ICETEMS), Nagpur, India, 2022, pp. 226-231, doi: 10.1109/ICETEMS56252.2022.10093357

[21] Rishu; Vinay Kukreja; Sahil Chauhan "Analysis of Facial Expression for Emotion Recognition using CNN-SVM" 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA) DOI: 10.1109/ICIRCA57980.2023.10220858