International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 02, ISSN No. 2455-2143, Pages 74-80
Published Online June 2023 in IJEAST (http://www.ijeast.com)

EMOTİON RECOGNİTİON USİNG EXPLİCİT FEATURE EXTRACTİON METHOD

Govind Madhav and Faisal Rais
HMRITM (GGSIPU) Delhi, India

Abstract: **Emotions are conveyed every single second, and understanding what they typically represent is crucial in practically every industry. People's ability to accurately recognise emotions varies greatly, making technology utilized for the purpose a relatively new subject of research. We attempted to work on extracting significant features from the image datasets and feeding those characteristics to a CNN in this study to aid the emotion detection for the CKplus and JAFFE dataset. The two separate extraction techniques employed are the Gabor filter and SIFT. The extraction techniques assist in identifying the main attributes that contain a profusion of visual data.The experimental investigation demonstrates that the gabor filter paired with CNN outperforms the state-of-the-art performance using the CKplus dataset, with an overall accuracy of 98.12.**

Keywords: **Emotion detection, Gabor filter, SIFT, CNN**

I. INTRODUCTİON

The basic emotions that all people experience are universal and are consistently expressed on their faces. Automatic recognition of human emotion in pictures and videos will allow us to carry out the detection, extraction, and evaluation of these facial expressions in an efficient manner.Significantly, these emotions are consistently expressed through facial expressions. This means that, regardless of language or cultural barriers, there will always be a set of fundamental facial expressions that people assess and communicate with. Following extensive research, it is now widely accepted that humans share seven facial expressions that reflect the experience of fundamental emotions[1].

Human emotions are mental states of feelings that arise naturally rather than as a result of intentional effort coupled by physiological changes in facial muscles implying expressions [2]. In many applications of human-computer interaction, nonverbal communication techniques like gestures, eye movements, and facial expressions are used, with facial emotion being one of the most frequently used because it conveys people's emotional states and feelings. Facial emotion recognition (FER) is a problem that DL (deep learning) models can also complete due to the universality of expressions. These algorithms or models can also outperform people at analysis and problem-solving, as they can at many other crucial jobs.

Recognition of facial expressions may find use in areas where efficiency and automation are advantageous, such as entertainment industry, social media mischief, content analysis, criminal justice, defense services, and healthcare. FER is a system that analyses facial expressions in static photos and videos to disclose information about an individual's emotional state.For instance,these system can be used for the product analysis wherein the reaction of customers can be recorded to see the success of the product launched etc.

International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 02, ISSN No. 2455-2143, Pages 74-80
Published Online June 2023 in IJEAST (http://www.ijeast.com)

Fig 1 : Seven universal emotion and their characteristic expressions

Emotions are naturally recognised and comprehended by the brain, and software that can do the same is presently being developed.Emotional artificial intelligence is a type of technology that can read, mimic, understand, and respond to human facial expressions and emotions[3].Using the well-known datasets CKplus and Jaffe, this research suggests feature extraction with convolution neural networks as the FER approach. Each of the seven universal emotions is represented by an adequate number of images in the dataset.
The research and the experimentation done is outlined as: The existing FER approaches and previously tried methods are briefly reviewed in the corresponding Section 2.Section 3 showcases a quick look through the dataset description and the methodologies used in the research are explained in Section 4.The proposed FER technique's implementation is being detailed in Section 5 along with the experimentation and their underlying results.The conclusions and the scope of future work that can be implemented are drawn out in last section.

II. LİTERATURE REVİEW

A single Deep Convolutional Neural Network (DCN) model based on convolution layers and deep residual blocks was proposed by Jain et al.(2019)[4]. First, in the suggested model, all faces' image labels have been set for training. Second, a proposed DNN model is applied to the image.This model was trained using the Japanese Female Facial Expression (JAFFE) Dataset and the Extended Cohn-Kanade (CK+) Dataset. Convolutional neural networks and face emotion identification i.e FERC is a revolutionary method that Mehrabian et al.(2020)[5] discovered. Convolutional neural network (CNN) in two parts used by the FERC: The background is removed from the image in the first stage, while the second step concentrates on extracting the facial feature vectors. The five different types of typical face expression are identified by the expressional vector (EV) in the FERC model.
Chowdary et al.(2021)[6] employed transfer learning techniques to recognise emotions. The Resnet50, vgg19, Inception V3, and Mobile Net pre-trained networks are employed in this study. We swap out the pre-trained ConvNets completely connected layers for fully connected layers that are appropriate for the number of instructions in our task. The experiment used the CK+ database to run, and it was able to recognise emotions with an average accuracy of 96%.Pranav et al.(2020)[7] created a Deep Convolutional Neural Network (DCNN) model that categorizes five distinct facial expressions of human emotion. The hand

International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 02, ISSN No. 2455-2143, Pages 74-80
Published Online June 2023 in IJEAST (http://www.ijeast.com)

gathered image dataset is used to train, test, and validate the model.

Akhand et al(2021)[8] .'s proposal for a very Deep CNN (DCNN) modelling approach uses the Transfer Learning technique to adopt a pre-trained DCNN model by swapping out its dense top layer(s) suitable with FER, and the model is then adjusted with facial expression data. The pre-trained DCNN blocks are individually tuned once the dense layer(s) have been trained. The accuracy of FER has constantly improved because of this cutting-edge pipeline method. Khattak et al(2022)[9] 's model of a convolutional neural network was successful in classifying emotions from facial photographs and determining age and gender from the expressions, with an accuracy of 95.65%.

A modular system for identifying human face emotions was developed by Alreshidi et al. in 2020[10]. The framework consists of two machine learning algorithms designed for real-time applications (for detection and classification). The AdaBoost cascade classifiers are first examined in order to identify faces in the photos. A face's features are then represented by neighbourhood difference features (NDF), which are based on localised appearance data. The Local Binary Pattern (LBP) approach is used by Mukhopadhyay et al. (2021)[11] to provide textual representations of human faces and trains the CNN model using these images.

In Mohammed et al.(2021)[14] suggested a PCA technique to extract the features and employed KNN for the classification purpose.Khan et al.(2016) worked for facial expression detection using local descriptors which have fairly proven to be of a higher resistant to position changes, alignment issues and various effects of illuminations.

A unique deep network was developed by Sepas-Moghaddam et al. (2020)[12] that initially extracts spatial data using a VGG16 convolutional neural network. The Bidirectional Long Short-Term Memory (Bi-LSTM) recurrent neural network is then trained using perspective feature sequences to learn spatio-angular information while examining both forward and backward angular relationships. The framework is created to apply the hybridization of feature extraction and optimization using PCA and PSO, respectively, in order to attain a high precision rate of 94.97% in Arora et al. (2021)[13].

III. DATASET DESCRİPTİON

The Jaffe dataset, which consists of 213 tagged photos of significant human objectified emotions, was compiled from 10 distinct persons (females) in Japan. Six basic and one neutral emotional expression were required of each participant, and the images were evaluated by 60 validators who assigned overall conceptual values to each person's facial expressions.

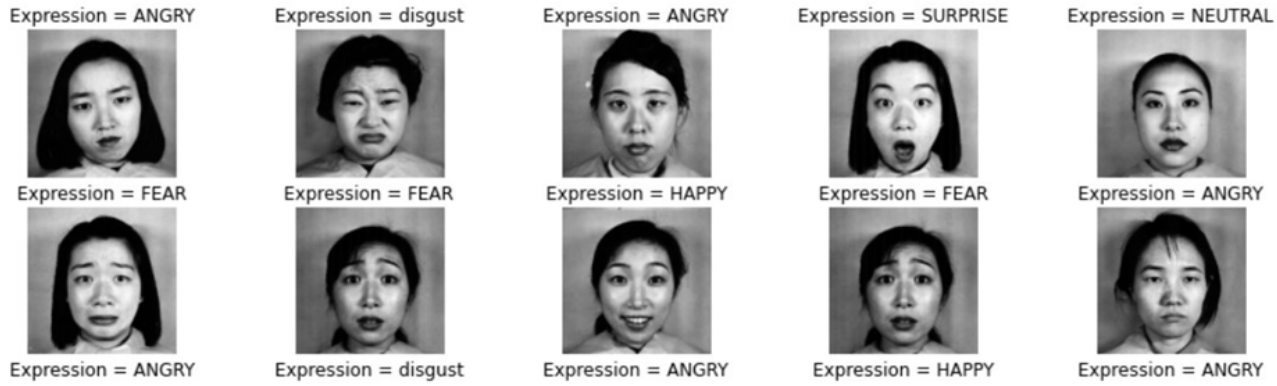

Fig 2 : JAFFE Dataset Sample

Fig 3 : CKPLUS Dataset Sample

The CKplus dataset, which has 981 grayscale pictures with a resolution of 48x48 pixels, is accessible on Kaggle. Emotional folders are used to categorize images. The collection comprises images of persons between the age restriction of 18 to 50 years, with a diverse background of genders and races. Each image has a label from one of the seven different emotion categories, which include anger, disgust, fear, happiness, sorrow, and surprise.

IV. METHODOLOGY

Automatic emotion recognition based on facial expressions is an interesting research area that has been presented and applied in a variety of domains, such as safety, health, and human-machine interactions. In order to enhance computer prediction, researchers in this field are interested in developing techniques to read, encode, and extract these traits from facial expressions. Since deep learning has been so successful, performance may now be increased using a variety of its designs. For face detection and feature extraction, image processing techniques are used by emotion recognition systems. The use of pattern recognition techniques for the classification of facial expressions comes next.Human face expressions are crucial in interpersonal communication. It has been demonstrated that the vocal component of a message, which consists of spoken words, only accounts for 7% of the whole message. The vocal component, or voice intonation, makes up 38% of the overall effect of the spoken message, while the speaker's facial expressions make up 55%.This study is majorly outlined on feature extraction and training the convolution net for the two datasets considered.

4.1 SIFT (Scale Invariant Fourier Transform)

Using the SIFT Detector, interest points on an input image are found. In applications like object recognition, pattern and gesture recognition, among others it enables the identification of localized features in images. SIFT is capable of feature detection regardless of the image's viewpoint, depth, and scale. The image data is transformed into scale-invariant coordinates to do this.
Gupta et al.(2019) used handcrafted features based on SIFT characteristics to conduct effective object detection. SIFT are especially beneficial for analysing images with varying orientation and scale. The research finds that feature extraction using SIFT works better than the modern approach. Detecting and describing local visual features can help in object recognition. SIFT characteristics are not affected by the size or rotation of the picture since they are local and dependent on how the item appears at certain interest points. Additionally, they allow for accurate item identification with little chance of mismatch and are relatively simple to extract.

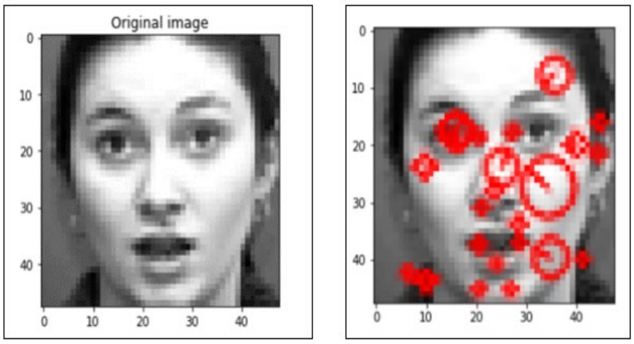

Fig 4 : SIFT Features on random CKPLUS Dataset Image

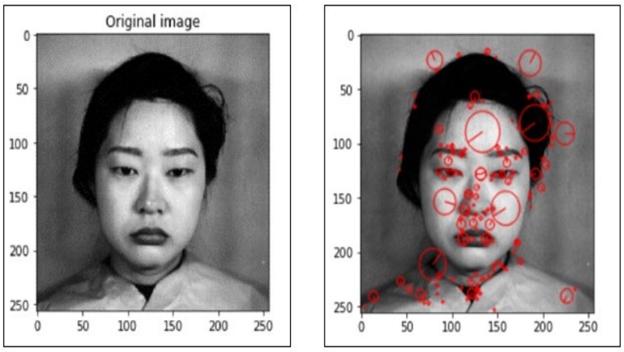

Fig 5 : SIFT Features on random JAFFE Dataset Image

4.2 Gabor Filter

A Gabor filter is a linear filter used in image processing to analyse texture. Simply said, it establishes if the image has any certain frequency content in specific directions in a small area surrounding the point or region of investigation. The frequency and orientation representations of Gabor filters, according to many contemporary vision specialists, are equivalent to those of the human visual system.
Sadeghi et al.(2019) used Gabor filters to carry out emotion recognition. Four datasets for face expression recognition are used to evaluate the suggested approach. The results of the experiments demonstrate that the suggested method performs better in recognising facial expressions in controlled and uncontrolled photos than the currently used image texture descriptors. It has been demonstrated that Gabor filters have the optimum localization characteristics in both the spatial and frequency domains.

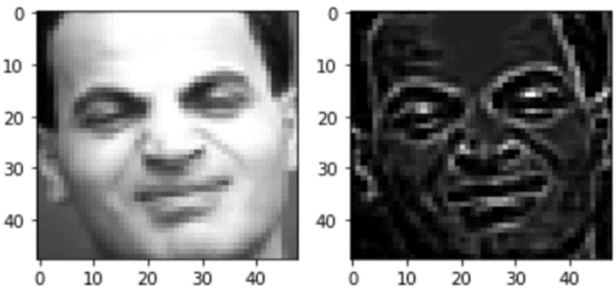

Fig 6 : Gabor Features on random CKPLUS Dataset Image

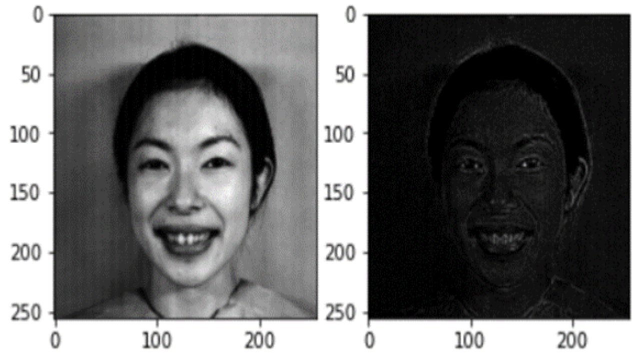

Fig 7 : Gabor Features on random JAFFE Dataset Image

4.3 Proposed Architecture
The architecture of the proposed approach can be viewed in two specific phases:

Phase-1: Feature extraction
SIFT and the Gabor filter are used to extract features. The analysis makes use of the CKplus and Jaffe dataset. In contrast to Jaffe's 231 images, the CKplus database has 981 images of trains. The seven emotions in the dataset are then further divided up into these images. The train image collection is then processed once the standard pre-processing is done. Both dataset's data are pre-processed before feature extraction techniques are used to isolate the most crucial or significant characteristics. These traits include crucial data that is relevant to the process of detecting emotions. The key traits that were extracted were details about the shape of the jawline, whether the eyes are closed or open, etc. The aforementioned factors are crucial for a more accurate examination of emotions.

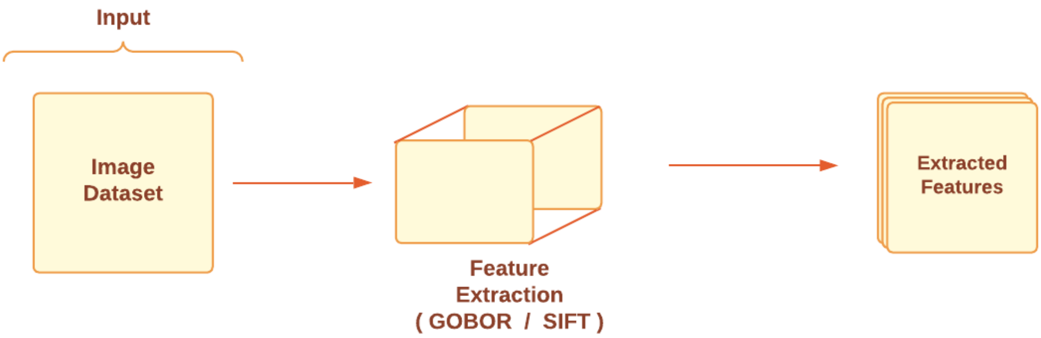

Fig 8 : Phase - 1 of Proposed Methodology

Phase-2: Convolutional neural network
A dense layer convolutional neural network is then fed with the recovered or detected features. A sequential model with three consecutive convolutional layers is constructed, and the output is then flattened. The dense layers that are fully coupled and have relu activations receive input from the flattened output. A softmax classifier layer and seven neurons make up the output dense layer, which structures the classifications for the emotion dataset.

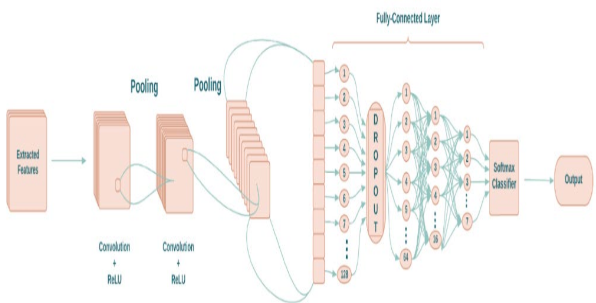

Fig 9 : Phase - 2 of ProposedMethodology

V. EXPERİMENTAL RESULTS

5.1 Implementation
There is sufficient data in both the train and test datasets to identify emotions. The datasets are loaded, and the train and test data are then divided into arrays for easy interpretation. The image labels from the train and test sets of data are reshaped and converted to categorical values using tensorflow.keras utils. The preprocessed images are then subjected to the use of the Gabor filter and SIFT feature extractions. The preprocessed images are given to the sequential model. The two core elements of the model are the convolutional network and the dense layer network. introducing the three convolution net layers, each with a separate maxpool layer, with the preprocessed images.Before being sent to the dense layers, the output of the convolution net is flattened. Dropout occurs at the end of the first dense layer. After compiling the constructed model, the model is assessed for its computational performance for the testing dataset using the evaluate function. The softmax classifier outputs categorized images.

5.2 Results

This section will show the performance of the suggested technique as well as a comparison of our approach to existing models. The experimental techniques have previously been described in the preceding section. Table 1 shows the training and testing accuracies encountered with the proposed model.Tables 2 and 3 show a comparison of the suggested model approach to current models.

Table 1.. Experimental results for Gabor-CNN and SIFT-CNN

Heading level	Training Accuracy (CKPlus)	Testing Accuracy (CKPlus)	Training Accuracy (Jaffe)	Testing Accuracy (Jaffe)
Gabor - CNN	97.96	98.12	98.00	91.43
SIFT - CNN	99.14	97.94	95.01	82.73

Table 2.. Comparison of existing and proposed model for Jaffe dataset

Approach	Accuracy
SVM [15]	87.43
Specific image processing with CNN[16]	84.48
Hybrid fully &weekly supervised[17]	89.01
Gabor-CNN(Proposed model)	91.43
SIFT-CNN(Proposed model)	82.73

Table 3.. Comparison of existing and proposed model for CKplus dataset

Approach	Accuracy
FN2EN[18]	96.80
Server handled CNN	96.60
Haar Wavelet Transform and Gabor Wavelet	97.30 & 98.00
Gabor-CNN(Proposed model)	98.12
SIFT-CNN(Proposed model)	97.94

VI. CONCLUSİON & FUTURE SCOPE

With the aid of a new and powerful framework proposed in this study, successful emotion recognition can be achieved.Data feature extraction serves as a useful approach for accuracy improvement, lowers the risk of overfitting, speeds up the training process, and enhances data processing and visualization. These are the key contributions/findings of the research conducted. In the research that follows, features are extracted using the Gabor filter and SIFT techniques, which are then used to train neural networks. The design, which combines explicit feature extraction with convolutional nets and additional deep neural layers, is what has led to this outperforming accuracy.We achieved an accuracy of 98.12% for Gabor-CNN and 97.94% for SIFT-CNN for CKPlus dataset and an accuracy of 91.43% for Gabor-CNN and 82.73% for SIFT-CNN for Jaffe dataset.This study can be expanded to include a sizable real-time dataset with images comprising various mixed diversities.

VII. REFERENCES

[1]. Pao, J. Emotion Detection Through Facial Feature Analysis.

[2]. Ali, M. F., Khatun, M., & Turzo, N. A. (2020). Facial emotion detection using neural network. the international journal of scientific and engineering research.

[3]. Danish, M., Qureshi, N., Kore, S., Indurkar, N., & Mote, N. EMOTION ANALYZER USING FACIAL RECOGNITION & RECOMMENDATIONS.

[4]. Jain, D. K., Shamsolmoali, P., & Sehdev, P. (2019). Extended deep neural network for facial emotion recognition. Pattern Recognition Letters, 120, 69-74.

[5]. Mehendale, N. (2020). Facial emotion recognition using convolutional neural networks (FERC). SN Applied Sciences, 2(3), 1-8.

[6]. Chowdary, M. K., Nguyen, T. N., & Hemanth, D. J. (2021). Deep learning-based facial emotion recognition for human–computer interaction applications. Neural Computing and Applications, 1-18.

[7]. Pranav, E., Kamal, S., Chandran, C. S., & Supriya, M. H. (2020, March). Facial emotion recognition using deep convolutional neural network. In 2020 6th International conference on advanced

computing and communication Systems (ICACCS) (pp. 317-320). IEEE.

[8]. Akhand, M. A. H., Roy, S., Siddique, N., Kamal, M. A. S., & Shimamura, T. (2021). Facial emotion recognition using transfer learning in the deep CNN. Electronics, 10(9), 1036.

[9]. Khattak, A., Asghar, M. Z., Ali, M., & Batool, U. (2022). An efficient deep learning technique for facial emotion recognition. Multimedia Tools and Applications, 81(2), 1649-1683.

[10]. Alreshidi, A., & Ullah, M. (2020, February). Facial emotion recognition using hybrid features. In Informatics (Vol. 7, No. 1, p. 6). MDPI.

[11]. Mukhopadhyay, M., Dey, A., Shaw, R. N., & Ghosh, A. (2021, September). Facial emotion recognition based on textural pattern and convolutional neural network. In 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON) (pp. 1-6). IEEE.

[12]. Sepas-Moghaddam, A., Etemad, A., Pereira, F., & Correia, P. L. (2020, May). Facial emotion recognition using light field images with deep attention-based bidirectional LSTM. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 3367-3371). IEEE.

[13]. Arora, M., & Kumar, M. (2021). AutoFER: PCA and PSO based automatic facial emotion recognition. Multimedia Tools and Applications, 80(2), 3039-3049.

[14]. Mohammed, S. B., & Abdulazeez, A. M. (2021). Deep Convolution Neural Network for Facial Expression Recognition. PalArch's Journal of Archaeology of Egypt/Egyptology, 18(4), 3578-3586.

[15]. Support Vector Machines - Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. IEEE Intelligent Systems and their applications, 13(4), 18-28.

[16]. Convolutional Neural Network - https://www.tensorflow.org/tutorials/images/cnn.

[17]. Cong, R., Qin, Q., Zhang, C., Jiang, Q., Wang, S., Zhao, Y., & Kwong, S. (2022). A weakly supervised learning framework for salient object detection via hybrid labels. IEEE Transactions on Circuits and Systems for Video Technology, 33(2), 534-548.

[18]. Ding, H., Zhou, S. K., & Chellappa, R. (2017, May). Facenet2expnet: Regularizing a deep face recognition net for expression recognition. In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017) (pp. 118-126). IEEE.