

Real-Time Facial Emotion Detection for Therapeutic Psychiatric Interventions

Vishal Chauhan
Department of Computer Science and Engineering
Galgotias University
Greater Noida, India
vishal23scse1010990@gmail.com

Syed Ayan Akhtar
Department of Computer Science and Engineering
Galgotias University
Greater Noida, India
@gmail.com

Aditya Raj Shukla
Department of Computer Science and Engineering
Galgotias University
Greater Noida, India
@gmail.com

Human emotions also are critical to such behaviors as decision-making, and interpersonal communications. These emotions have hence become focal research interest in computer vision and artificial intelligence in deriving its accurate identification using facial expressions. In this paper a real-time facial emotion recognition system using ResNet50 architecture will be presented as based on transfer learning. The model is conditioned on the FER-2013 dataset, comprising of 48 x 48 gray-scale human face samples in seven emotional categories namely; Angry, Disgust, Fear, Happy, Neutral, Sad and Surprise.

OpenCV is used to support the working of the program in real-time with the use of face detection and live video-capturing, and to employ the pre-infested ResNet50 network that is fine-tuned to provide emotion recognition. Class imbalance in the dataset has been solved by using class-weighted training thus achieving a balanced sensitivity to emotions. The analysis of accuracy, recall and the confusion matrix is used in the evaluation of model performance. The suggested system has proved to have a good recognition rate and an effective inference on a regular computing device. The findings have shown that transfer learning methods using deep learning can be useful in providing scalable real-time emotion detection systems that can be used in intelligent surveillance, mental health monitoring, and human computer interaction.

I. Introduction

The human feeling is central to the perception, communication and the decision-making process that determines how individuals relate to their environment as well as the other individuals. Facial expression is among other non-verbal cues that are the most instantaneous and universally comprehended signals of emotional state of being. It is observed that since digital systems are increasingly interacting with human beings in real-time settings - including healthcare systems, smart learning systems and intelligent surveillance - automatic recognition of facial emotions has become a significant stakes in the context of computer vision and artificial intelligence research.

Facial emotion recognition (FER) allows emotion-sensitive systems to change in response to user affect to enhance usability and inclusivity. The systems can be useful in supporting those people who have a difficulty in their

communicative functions, such as patients with speech deficits, children with the autism spectrum, or those who need assisted care. This is further enhanced by the increased use of artificial intelligence and Internet of Things (IoT) technologies, which result in an increased need to have reliable and real-time emotion recognition frameworks that can work effectively with the regular hardware.

The latest development in machine learning and deep neural network has drastically changed the scope of pattern recognition, such as speech processing and object recognition as well as the behavioral recognition. Deep learning models have shown their undisputed benefits compared to conventional handcrafted feature-based methods including Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and the Support Vector Machines (SVMs). Unlike traditional methods that have preferred features that are hand engineered, deep architectures can learn hierarchical representations from data directly and therefore provide better accuracy and robustness.

With these developments, the issue of acquiring accurate real-time emotion recognition is still uncertain because of the discrepancies in the facial postures, lighting, expression magnitude, obscurement, and the complexity of a background. Convolutional Neural Networks (CNNs) and especially with transfer learning techniques have proven to be able to go a long way in solving these issues. Neural networks like VGGNet, MobileNet, and ResNet have the potential to extract discriminative feature of the face and minimize the training complexity and computation cost.

Despite the above, there are a number of available studies that use FER with respect to static images or offline video recognition, however, due to the nature of the applications, their validity cannot be applied in interactive and clinical research. FER systems that operate based on Vision have provided an efficient alternative to multimodal systems that typically implement more sensors on the physical device, as well as increased computation costs. Most existing approaches however find it hard to strike a balance between real time functionality and the real world strength.

It is a proposed real-time facial emotion detection system, which is built on the transfer learning scheme, using the ResNet50 architecture in a training scenario via the FER-

2013 database. It solves the issue of class imbalance with the help of class-weighted training, and with the aid of the OpenCV-based face detection, it can perform efficient inference in real-time. The suggested framework is set to fill the gap between scholarly studies and implementation systems through providing timely, computationally effective, and scalable emotion recognition to fit the healthcare monitoring, human-computer interaction, and intelligent systems.

II. Review of the literature.

Facial emotion recognition has turned out to be a leading field of researchers owing to its applicability in the context of human behavior and implementation of systems which can recognise emotions. Visually Early FER experiments were almost universally based on manual extraction of features, i.e. Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and geometric landmarks of the face, using traditional classifiers such as Support Vector Machines (SVMs) and k-Nearest Neighbors (k-NN). Although these methods showed decent performance in the artificial setting, the performance dropped in the natural setting that included pose variation, variation in light, and face covering.

It was a great turn that FER research took with deep learning adoption specifically convolutional Neural Networks (CNNs). CNN-based methods will automatically acquire hierarchical feature representation on raw image signals this way it avoids the manual design of features. Architectures like AlexNet, VGGNet and ResNet have continued to perform better over the traditional approaches on difficult data sets such as FER-2013. These models can be able to capture minute movements of the muscles of the face which can help differentiate between closely related emotional expressions.

Transfer learning has also developed FER by allowing deep networks to analyze reused networks who had been trained on huge datasets like ImageNet. Transfer learning enhances the classification accuracy and reduces the training time, and data used, by refining the high-level features to be used during the emotion recognition tasks. ResNet50 and MobileNet architecture are popular because of their reasonable trade-offs between the representational and computational capability, especially in real-time.

More recent studies have been concentrating on real-time FER which streams and processes live video streams. Systems are largely combined methods of face recognition with CNNs, and it is often difficult to maintain real-time behavior and robustness, particularly on commodity hardware. They are often reported to have trade-offs between the speed of inference, model complexity and the accuracy of recognition.

FER has been used in mental health monitoring, stress detection, autism assistance and affect-aware human-computer interaction in healthcare applications. In spite of the fact that multimodal systems with the use of facial cues, speech, and physiological signals may be a more efficient way of increasing the accuracy of recognition, they impose

practical limitations in terms of the availability of sensors, the complexity of implementation, and higher costs of computation. As a result, the vision based FER systems are still more practical in the field of the real-world use.

Nevertheless, FER systems show few studies in existing literature even though they have considerably advanced. The majority of research challenges view emotion recognition as a passive act of monitoring, but not the active process of supporting a decision. Moreover, it has been found that temporal dynamism and emotion persistence are frequently ignored, and most systems only make frame-level predictions. These loopholes show that real-time FER frameworks that go beyond classification are required to enable meaningful intervention, as well as, clinical decision-making.

III. Methodology

This section gives the methodology involved in the designing of the proposed real-time facial emotion recognition system including the data set selection, data pre processing, and the model configuration, training plan, and the evaluation measurement. The design has placed emphasis on strength, reusability and compatibility with real-time implementation.

A. Dataset Collection

To examine and demonstrate its functionality, the system is trained on the standard of facial emotion recognition of FER-2013, which is one of the most popular. This database journal was published as the Challenges in Representation Learning at a specific instance in ICML 2013 and has a total of 35,887 48 x 48 pixel grayscale face images. The pictures are also rounded off to seven classes of emotions which include Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise.

Others of these databases are divided into training, validation, and test sets consisting of 28,709,3,589 and 3,589 images respectively. The photos are made in the uncultivated, natural settings resulting in elevated light, pose and intensity of expression ranges. This will be forced to enhance the degree of classification but will also enhance the model generalization as well as represent the real world of operations.



preprocessing measures to facilitate the nature of learning to be more secure. The images are afterwards cut down to 224 x 224 pixel in order to fit into the input of the network. A total

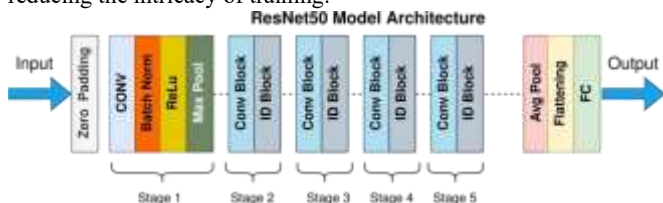
of 3 channels are replicated in order to achieve the replication of the channel in case of the ResNet50 where grayscale images are converted to RGB.

The pixel values are so as to improve numerical stability and reduce the sensitivity to variation of illumination. To address the issue of class imbalance particularly in case the category is underrepresented such as the category of disgust, class-weighted training is performed with weights equal to the inverse of the distribution of the classes.

The face locations are pulled in real-time with the help of the OpenCV and extracted face regions of the live video frames and resized, normalized and re-shaped and transmitted to the trained model.

C. Model Architecture

The proposed system will be developed under the assistance of the transfer learning strategy made on the basis of ResNet50. The feature extractors are Convolutional layers which have been trained, and encode spatially useful patterns of features in facial expression. The network is extended with various features that identify emotions, namely global average pooling, dropout, and fully connected layers that are task specific and use softmax as their activation. This format improves generalization in addition to, reducing the intricacy of training.



D. Model Training

Backpropagation Model training Backpropagation is an optimization process of the network using a loss function which is appropriate. The optimization of Adam algorithm or stochastic gradient descent strategy is applied in a number of epochs where mini-batches learn. In order to ensure that overfitting is avoided and generalization is attained, dropout and early stopping are regularization techniques. The adoption of the class-weighted loss functions also ensures that one gives a balanced learning of the groups of emotions.

E. Model Evaluation

In the determination of model performance, a separate test data is adopted. The measures, applied in their assessment, are accuracy, precision, recall, F1-score and confusion matrix analysis. In addition to predictive accuracy, computational efficiency metrics such as the inference time and frames per second (FPS) are also put to test to ensure that it is suitable to use in a real-time setting.

F. Software and Hardware

The proposed facial emotion recognition system was written using Python (version 3.8. or more) since it has an immense amount of support of machine learning and computer vision

libraries. The framework has been deployed on top of TensorFlow and Keras, which can be used as a very adaptable framework of deep learning. Other libraries were employed which are NumPy, OpenCV; libraries were targeted to complete the numbers calculations and image processing (preprocessing and detection) and video processing (in real-time).

Jupyter notebook and VS code helped in that it permitted the prototyping, debugging, and visualisation of the models when creating and testing them. In order to arrive at the models, a required environment involving a GPU was employed in order to accelerate the convergence and Google Colab was the primary environment to be applied to the models. Recognition and estimation of emotions were done in real time with a standard laptop with a web based built-in camera and a CPU and confirmed that the system will operate without the hard-to-find hardware.

Category	Specification
Programming Language	Python 3.8 or higher
Deep Learning Frameworks	TensorFlow, Keras
Supporting Libraries	NumPy, OpenCV
Development Environment	Jupyter Notebook, Visual Studio Code
Training Hardware	GPU-enabled environment (Google Colab)
Deployment Hardware	Standard laptop CPU with integrated webcam

G. Model and deployment.

Once the trainings were done the most beneficial model was saved in the.keras file as emotion_resnet50.keras with the network architecture and the trained weights saved. The method ensures that it is reproducible, and it becomes easier to do inference in a real time by using the models.

During the deployment phase, the trained model is being loaded at the required runtime, and the real-time emotion detection script begins to work which links the webcam feed and list the received video is being classified according to the emotions during this stage. The dynamism of the deployment pipeline is allowing the related systems to be supported in a very easy way like such mobile applications, systems grounded on Raspberry Pi, and cloud based application programming interfaces with only some few modifications to the underlying inference logic.

Component	Description
Model Format	.keras
Saved Model Name	emotion_resnet50.keras
Model Storage	Local/Cloud-based file system
Runtime Loading	Loaded at inference initialization
Input Source	Live webcam video stream

Component	Description
Deployment Platforms	Laptop, Mobile devices, Raspberry Pi, Cloud APIs

IV. RSA Results and Analysis.

It compared the proposed system on FER-2013 data set and implemented transfer learning architecture of ResNet50. Online evaluation included offline validation and real-time inference and it also used such measures as accuracy, F1-score along with confusion matrix.

A. Training and validation performance.

The good performance observed with the use of the model was in the high training of 98.4 percent with a high F1-score of 97.1 percent. The system had an accuracy, which was 67.8, and a macro-averaged F1-score, which was 65.3, using the validation set. The realized gap between the training and validation performance is a demonstration of the challenged set of the dataset in-the-wild, and is an indicator of controlled generalization.

Metric	Training Set Value	Validation Set Value
Loss (Categorical Cross-Entropy)	0.211	0.865
Accuracy	98.4%	67.8%
F1-Score (Macro Avg.)	97.1%	65.3%

B. Class-Specific Analysis

The confusion chart was compared and it was realized that there was intense recognition between Happy and Surprise expression and the assistance of the testing values (above 85) of remembering the words. Unconnected emotions with negative valence were more prone to getting mixed, and, specifically, the emotional state of Fear and Surprise and the emotional states of Angry, Disgust and Sad, as the related face features were similar. This was consistent with indistinct differences in expression as the Neutral one was muddled mediocely with Sad.

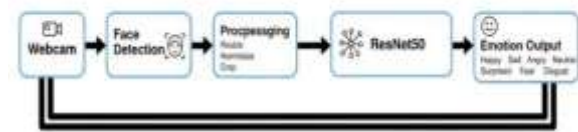
Emotion	Recall (%)
Angry	61.2
Disgust	54.7
Fear	58.9
Happy	88.3
Neutral	64.1
Sad	62.5
Surprise	86.7

C. Real-Time Performance

The system had capacity to handle live video which it could run approximately 40 FPS and median time that can be used to draw an inference of 25 ms per frame with typical GPU

equipment. This cannot be assured even by real-time requirements in compatibility to interactive and therapeutical use.

Flowchart of Emotion Recognition System



D. Implications

The results confirm the availability of the implementation of the FER systems on the basis of the deep learning in order to assist the therapeutic processes in real-time. There are categories of emotions that are hard to deal with, however the overall accuracy and efficiency factor provides scope to include it in the clinical decision support systems.

V. CONCLUSION

This paper has provided the architecture and execution of an on-the-fly facial emotion recognition model that provides the support of the therapeutic and psychiatric intervention based on the analysis with the help of deep-learning-based analysis. The proposed framework became highly dependable in detecting emotions and ensuring the ability to work in real time by using transfer learning with a fine-tuned ResNet50. The attained validation and processing rate of 67.8 and roughly 40 bpm respectively indicates that the system is capable of giving an adequate performance balance between the recognition and computation rate at standard hardware platforms.

The experimental findings suggest that combination of deep convolutional model with relevant preprocessing and class-weighted training can strongly reflect discriminative facial features even in unconstrained and real-world environments. The presence of class imbalance that was considered in the FER-2013 data helped to achieve a better sensitivity across the different categories of emotions, which also enhanced the overall reliability of the overall recognition process. The seamless inference with real time was also made possible by the incorporation of the OpenCV based facial recognition, which ensured that the system was compatible in an interactive setup.

One of the valuable contributions to this work is that it has widened facial emotion recognition beyond the traditional passive surveillance. Incorporation of Intervention Logic Unit (ILU) and the Visual Neutralization Technique (VNT) brings with it a systematic means of aiding in making clinical decisions. The system allows the psychiatrists to make objective assessments by visually controlling protracted negative emotional responses to minimize possible cognitive bias when providing therapeutic services in live. This strategy is also a basis to the emotion-conscious feedback that may be used in clinician training and supervision.

Moreover, the presented framework evidences the feasibility of operations in various sectors, which are human-computer interaction, mental health monitoring, adaptive learning systems, and intelligent surveillance. Its capability to work in

real time without the need to have special hardware also increases its practicability of wide-scale application. The modularity design offers scalability and efficiency which may be extended later without compromising the fundamental benefits of accuracy and scalability.

To sum up, the study provides a solid and implementable facial emotion recognition platform which is practical in establishing the gap between theoretical and practical advancements in the development and clinical therapeutic practice. The results highlight the promise of emotion-sensitive deep learning-based systems in improving human-computer interaction and being helpful to mental health workers in addition to playing a vital role in ensuring the development of intelligent clinical systems.

VI. FUTURE WORK

Although the proposed system presents positive results, several directions are possible where the research can be taken further. The out-of-province clinical validation of the performed work which is conducted on pilot studies in a real therapeutic setting is one of the primary directions of further practice. Such considerations would support the analysis of the feasible usefulness of the ILU and VNT systems to the reducing work load of clinician, emotional sensitivity and support of the intervention measures.

It is also possible to optimize the model that will be used in the lightweight and resource-constrained devices, so optimization of the model can be considered in the future studies. The research into the efficient deep learning structures such as MobileNetV3 or EfficientNet versions could be applied to provide real-time detection of emotions during non-GPU, mobile, and embedded clinical devices as well.

The other extension that is significant is the time dynamics and multimodal information. The time-dependent patterns of perseverance of emotions, and integration of complementary cues, e.g. of physiological or behavioral manifestation, may lead to improved differentiation between short- and long-term emotion.

Finally, there should also be provisions of ethical and privacy issues prior to carrying out a mass deployment. A future way forward would be to review the safe data handling protocol, the anonymity in data handling and compliance with health care regulatory framework to make sure that the facial emotion recognition technology will be utilized in a responsible and trust-worthy way in the clinical practice.

REFERENCES

[1] Dada EG, Bassi JS, Chiroma H, Abdulhamid SM, Adetunmbi AO, Ajibuwa OE. Machine learning for email spam filtering: review, approaches and open research problems. *Heliyon* 2019;5(6):e01802. <https://doi.org/10.1016/j.heliyon.2019.e01802>.

[2] Xie M. Development of artificial intelligence and effects on financial system. *J Phys Conf* 2019;1187:032084. <https://doi.org/10.1088/1742-6596/1187/3/032084>.

[3] Hegazy O, Soliman OS, Salam MA. A machine learning model for stock market prediction. *Int J Comput Sci Telecommun* 2014;4(12):16–23.

[4] Beckmann JS, Lew D. Reconciling evidence-based medicine and precision medicine in the era of big data: challenges and opportunities. *Genome Med* 2016;8(1):134–9.

[5] Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. *Jama* 2014;311(24):2479–80.

[6] Loconsole C, Chiaradia D, Bevilacqua V, Frisoli A. Real-time emotion recognition: an improved hybrid approach for classification performance. *Intelligent Computing Theory* 2014:320–31.

[7] Huang X, Kortelainen J, Zhao G, Li X, Moilanen A, Seppänen T, Pietikainen M. Multi-modal emotion analysis from facial expressions and electroencephalogram. *Comput Vis Image Understand* 2016;147:114–24. <https://doi.org/10.1016/j.cviu.2015.09.015>. [8] Raheel A, Majid M, Anwar SM. Facial expression recognition based on electroencephalography. In: 2019 2nd international conference on computing, mathematics and engineering technologies (iCoMET), Sukkur, Pakistan; 2019. p. 1–5.

[9] Vassilis S, Herrmann Jürgen. Where do machine learning and human-computer interaction meet?. 1997.

[10] Keltner D, Ekman P. In: Lewis M, Haviland Jones JM, editors. *Facial expression of emotion, hand book of emotions*. New York: Guilford Press; 2000. p. 236–49.

[11] Ekman P. Darwin and facial expression: a century of research in review. United State Of America: Academic Press; 2006. p. 1973.

[12] Ekman P, Friesen WV. Constants across cultures in the face and emotion. *J Pers Soc Psychol* 1971;17(2):124.

[13] Ekman P. Darwin and facial expression: a century of research in review. United State Of America: Academic Press; 2006. p. 1973.

[14] Ekman P, Friesen WV, Ancoli S. Facial signs of emotional experience. *J Pers Soc Psychol* 1980;39:1123–34.

[15] Ekman P, Friesen WV, Ancoli S. Facial signs of emotional experience. *J Pers Soc Psychol* 1980;39:1123–34.