International Journal of Scientific Research in Engineering & Technology
Volume6, Issue2 (March-April 2026), PP: 79-85.
https://www.doi.org/10.59256/ijsreat.20260602012
www.ijsreat.com
ISSN No: 2583-1240

NLP-Powered Emotion Analysis and ML-Driven Speech Processing with Flask

Thota Siva Naga Thirumalbabu[1], Tungala Uma Mahesh[2], P. Krishnaveni, M-Tech., (Ph.D)[3]

[1 2 3]Department of Computer Science and Engineering, Sathyabhma Institute of Science and Technology Chennai, Tamilnadu, India.

.

To Cite this Article: *Thota Siva Naga Thirumalbabu[1], Tungala Uma Mahesh[2], P. Krishnaveni, M-Tech., (Ph.D)[3], "NLP-Powered Emotion Analysis and ML-Driven Speech Processing with Flask", International Journal of Scientific Research in Engineering & Technology, Volume 06, Issue 02, March-April 2026, PP: 79-85.*

Abstract: *Emotion recognition through facial expressions has been a key parameter in the development of human-computer interaction as a technology with some useful results in terms of emotional states in real time. This study presents a hybrid approach for emotion recognition by combining deep learning methods and traditional feature extracting methods to achieve high accuracy and computational efficiency for emotion recognition. The proposed system makes use of the MobileNetV2 architecture for deep feature extraction, which is known to be lightweight, hence suitable for real-time applications. Moreover, fine-grained texture features are abstracted from facial regions through Local Binary Pattern (LBP) for capturing texture details at the fine grain-level, which are complementary to the deep representations and improve the model's ability to distinguish the subtle emotional states. These two sets of features are concatenated and fed to a Random Forest (RF) classifier, that is known to be robust and efficient in the context of moderately sized datasets, in order to classify the emotions using categories such as happy, sad, anger, surprise and neutral. The system is trained and tested with the help of a benchmark facial emotion dataset, incorporating pre-processing techniques of face detection, alignment, normalization and data augmentation for better generalization. The performance of the system is assessed using various metrics such as accuracy, precision, recall, F1-score, confusion matrices with measurements of latency to be able to provide real-time feasibility. Experimental results show that the performance of MobileNetV2 + LBP + RF is higher than that of MobileNetV2 + SoftMax head in the case of distinguishing subtle emotions. The system is implemented in the form of a Flask-based web application that is combined with the OpenCV library for live webcam streaming with minimal latency and user privacy through on-device inference.*

Keywords: *Emotion recognition, facial expression, deep learning, MobileNetV2, Local Binary Pattern, Random Forest*

I. INTRODUCTION

The role of human emotions Human emotions are important for human interactions, shaping decision-making, behaviour, and social interactions. The capacity to grasp emotions and emotions being part of human interaction is one of the most basic aspects of interpersonal communication. With the continuous advancement of machines and intelligent systems into our daily life, the demand for systems that can sense and analyse human emotions is gradually increasing. In particular, the emotion recognition through facial expression is becoming customer service. Facial expressions are understood by a wide range of people and so are well suited for use as an emotion detection mechanism in automated systems.

With the rise of artificial intelligence (AI) and machine learning, analysing facial images of human emotion recognition has been an important research area. Convolutional neural networks (CNNs) using deep learning have demonstrated great potential in recognizing facial expressions. However, there are still some challenges left for the obtaining of high accuracy, particularly in the application of real-world environments where varying lighting conditions, pose of the face and occlusions can negatively impact the performance of the model. The deployment challenge becomes more serious as typical deep learning models are usually computationally expensive and easily suffer from performance degradation in less controlled settings.
Nevertheless, emotion recognition systems have great potentialfor improving human-computer interaction (HCI). Enabled digital systems can be more adaptive based on the integration of emotion-aware systems, for example adaptive learning systems that monitor student engagement, telemedicine systems to monitor patients' well-being, and customer service bots to tailor responses based on the emotional state of their clients. Moreover, mental health monitoring is another application of these systems, which could help to obtain insight into a person's emotional arousal over time, and thus detecting the presence of an early emotional distress/other mental health problems.

The main reason for this research is to break the limitations of the current emotion recognition systems. Although deep learning models still provide amazing results, these models are also computationally expensive and more susceptible to environmental changes. Moreover, many systems only consider the deep learning features and ignore fine-grained texture

information which can increase the accuracy of emotion classification. This paper aims to overcome these limitations by suggesting a hybrid method that takes advantage of the deep learning features provided by MobileNetV2 together with the handcrafted texture descriptors from Local Binary Patterns (LBP), and then classification based on a Random Forest (RF).So the intention is to make a system, which is not only accurate but also efficient and feasible for real-time applications.A huge issue with emotion recognition is that models tend to break down and stop working correctly in erratic or noisy environments. Facial expression datasets that are used to train models for facial expression recognition are typically acquired in controlled environments, which are not representative of the variability that occurs in the wild. This difference results in the decreased performance when the system is applied in practical use. Also, many existing systems depend a lot on deep learning models, which are too big and need a lot of processing power, so cannot be used in edge devices or applications that must work in real-time, where latency and computational efficiency are important.

The objectives of this project are to create a lightweight emotion recognition system which is able to operate in real-time using images captured of the face using a webcam. The system tries to integrate the efficiency of MobileNetV2 as a deep feature extraction model and texture details provided by Local Binary Pattern (LBP) features. These features will then be classified by Random Forest model which is known to be robust and computationally low. The blend of these techniques will allow to generate a model, which will be both accurate and efficient and will overcome the main limitations on the existing emotion recognition systems.Apart from the accuracy, the system is designed to have low latency and computational efficiency to guarantee real-time processing ability. The model will be implemented with the Flask, a micro web framework, that will make the model available via a web-based interface. This deployment structure makes the system easily deployable into different applications as a scalable emotion detection platform. Furthermore, the system will be created to ensure user privacy by filtering raw facial data with on-device inference, which is in line with the ethical and privacy standards.

One among such difficulties is that there exists class imbalance in the protein datasets used in emotion recognition. Also, emotions that are less frequent, such as fear, surprise, and disgust are underrepresented compared to more frequent emotions, such as happiness and sadness. This imbalance can put weight toward the mis judgment of model predictions, resulting in more low recall rates of these underrepresented emotions. This research will attempt to mitigate this issue with data augmentation techniques which will aim to produce a balanced representation of emotions to improve the model's capability to detect each of the states with sensitivity.

Evaluation of the system will be carried out based on different performance metrics like accuracy, precision, recall, F1-score, and analysis of the confusion matrix. These metrics will provide a thorough insight on how well the model can classify the different emotions. In addition to the characteristics of developed technology, the system latency will be determined with the aim of the system to provide emotion detection in real-time with no perceptible delay. This will be important for applications such as live interaction systems where speed is necessary in order to maintain a natural flow of communication. With regards to the contribution of the proposed research to the field of emotion recognition, then this research will cover a hybrid framework which combines the advantages between the deep learning and the classical feature extraction techniques. With this system, it becomes possible to overcome the existing models, in particular in terms of computational efficiency and applicability in real time. By employing MobileNetV2 as deep feature extractor, LBP texture descriptions and Random Forest classifier, this solution not only guarantees an accurate system, but also one that can potentially be deployed in real life situations quite quickly.

In conclusion, the project aims at creating a reliable and efficient emotion recognition system which can be used in real- time. This research work is aimed to address the limitations of the current system such as computational complexity, class imbalance, and environmental variability in order to build much more accurate and accessible intelligent machines capable of recognizing emotions. The efficiency of the proposed system along with the flexibility to operate in various environments, will render it capable for a variety of applications such as education, healthcare, customer service, entertainment and so on

II.REVIEW OF LITERATURE

Emotion recognition via facial expressions has been one of the highlighted areas of research for the past years, owing to the increased requirement for smart human-computer interaction systems. Numerous research papers focused on different methods to improve the accuracy and efficiency of facial expression recognition have been published. In the early stage, deep learning-based methods have been implemented due to the significant jump in performance they have up against the traditional ones. Ali et al. (2020) outlined the advancement in the deep learning models for the facial expression recognition and illustrated fresh thoughts on how far the strategies can be extended further in the future.

Ahn and Kim (2021) analysed recent advancements in deep facial expression recognition, which provided an insight into the new techniques that are emerging, and have led to the changes in deep facial expression recognition evolution. These include the utilization of more advanced architectures and large quantities of diverse types of information to train models with decent generalization skills in real-world applications. The authors noted that the general trend toward using hardware design techniques referred to as CNNs for facial feature extraction has been established as reasonable at identifying the environmental subtlety of facial images such as emotions. However, they also recognized the limitation of such models caused by their inability to handle the problems of occlusions, illumination changes, and facial pose variations.Benitez-Garcia et al. (2020) worked on emotion recognition in uncontrolled environments which is also known as in-the-wildrecognition. In research they demonstrated that deep neural networks could be run successfully with this kind of data, but noted that performance was quite variable with quality of input data and the condition of an environment. This effort further reinforced the importance of compiling a large amount of data reflective of the reality of variability and development of high- fidelity models that would maintain a high level of accuracy for dynamic environments.

In a similar context, Chen, Tao, Yu, Yang, & Xie, (2022) presented a review of deep learning architectures in the area of facial expression recognition. Their contribution was to adopt more complex models and architectures such as deep CNNs and transfer learning that have made it possible to improve significantly on the emotion recognition accuracy. This study observed the use of pretrained models with large datasets to perform reasearch as well as tuning trained models to recognize a particular facial expression task to improve generalization and avoid labeling large amounts of data.

The multimodal data fusion for emotion recognition is also achieved its new popularity in recent studies. Chen et al. (2023) investigated fusion transformer-based multimodal emotion recognition, which includes facial expressions with other forms of modality e.g. speech or text. This method will give a richer representation of the emotion states as it considers a number of cues coming from different paths, which will lead to more stable and strong predictions. However, systems with such complexity increase the computational load of the system, which may be less practical for real-time applications in the absence of optimization.

Facing the limitation of deep learning models, Corneanu et al. (2021) introduced the concept of transfer learning and the use of transfer learning methods to surmount the data scarcity issue in facial expression analysis. By training large models on a great amount of data, the model they developed was able to identify facial expression even using a minimal amount of annotated data. This approach has found applications in problems where labelled data is in short supply or expensive to obtain.

Furthermore, Ding et al. (2022) suggested attention-based CNNs to develop even more robust facial expression recognition systems, which offer better performance to the model in recognizing emotion even when there are occlusions or partial faces. Their solution proved to be significantly better when challenged with some of the hard real-world cases where faces are frequently rotated relative to each other or partially occluded or not in the best possible conditions (such as low light conditions or head rotation).Elaina et al. (2021) proposed hybrid CNN-LSTM model in order to extract spatial and temporal information from the video data in facial expression recognition. This model was especially convenient when the scenes are dynamic (videos) where the emotions vary with time. Because the visual feature of the face, and how the facial expression was changing over time, could be effectively captured by the long short term memory networks(LSTMs) and the CNNs, it was able to use a combination of CNNs and LSTMs.

Fan et al. (2023) proposed a residual masking network for facial expression recognition, which is a network that enhances the performance of the model by masking unwanted areas of the image and making focus of the model on the most important facial features. This approach is useful to overcome the challenges due to background noise or other distractions for the real-world application where the facial images can be usually cluttered or noisy.

Lastly, Gao et al. (2022) presented a lightweight face expression recognition by knowledge distillation. Their method can be used to implement the facial recognition models in devices with limited computational power such as a mobile phone or embedded systems. Knowledge distillation is the process of transferring the knowledge from a large and complex model to a smaller more efficient model without losing a large amount of the accuracy. This makes it suitable for real-time applications, where computational performance is an important factor.

Overall, there is a need for robust, efficient, and scalable facial expression recognition systems based on the literature. While deep learning models have outperformed in terms of accuracy, scalability is still a challenge for verifying them in the real world. Problems such as domain variance, lack of data and computational expense continue to be a priority, as does the fusion of multimodal data for greater emotion detection. The ongoing advancement of model architectures, data augmentation techniques and optimization strategies will play a major role in the development of the field of emotion recognition.

III. PROPOSED METHODOLOGY

1.Existing System

The current emotion recognition-based systems mainly use the deep learning models for facial expression detection. At present, the most widely applied deep learning systems for extracting features from facial images are training convolutional neural networks (CNNs) on big datasets. However, these systems are prone to several issues such as high computational cost, no possibility for real-time deployment and limited robustness for variations in facial pose, illumination variability and occlusion. Some systems only consider deep features extracted from CNNs which did not capture the fine-grained texture details which are crucial to distinguish subtle emotional facial expressions in videos.Most current systems also need large scale data sets with massive amounts of preprocessing in order to model domain variance. The performance is degraded when they are applied to the real-world in terms of their diversity, because the real-world experiences lighting conditions, different ethnicities and occlusions, which are out of variety in the dataset. In addition, since deep learning models are based on high-performance hardware, it is challenging to deploy the model onto mobile devices or low-resource environments.

2.Proposed System

The proposed system is expected to fuse the merits of deep learning techniques and conventional handcrafted feature extraction methods to form a hybrid emotion recognition system with high recognition accuracy and low computational complexity. The deep feature extraction is achieved by using a pre-trained MobileNetV2 architecture because it is lightweight and can be used for real-time applications. In addition to deep features, the detail texture information of facial expression is obtained through the Local Binary Patterns (LBP) to supplement the deep features, and the recognition effect of these subtle emotional states is improved.The system uses the Random Forest (RF) classifier, which is effective for moderately-sized data, and is well-known to be robust and efficient for the task of feature classification. This combination enables the system to have high accuracy while computing feasible for deployment on applications in a real-time manner, such as mobile devices and embedded devices. The system is implemented with a Flask-based web app that makes use of OpenCV for low latency webcam streaming for privacy and low latency predictions via on-device inference.

3.System Architecture.

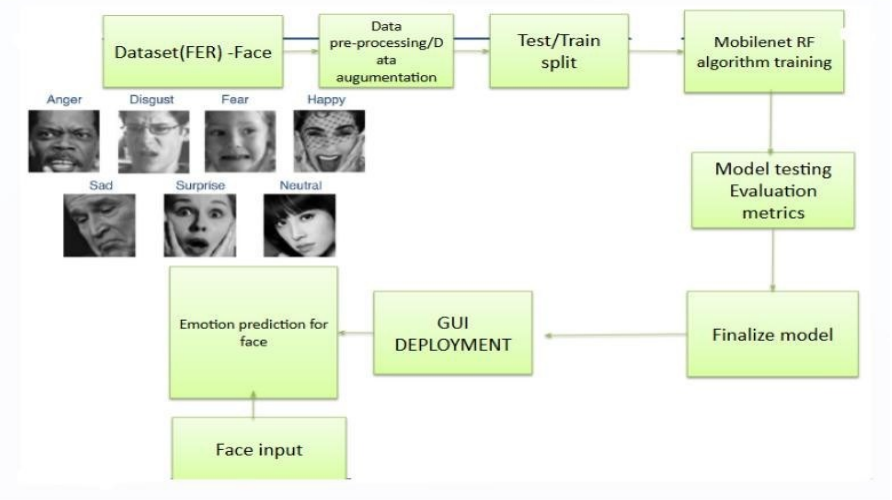

Fig.1.System Architecture

The system architecture is based on a modular design that incorporates several key components for efficient emotion recognition from the facial expressions. The working of the system flow could be described as follows:

●Input Module The system takes live video feed from a webcam or static pictures, recorded using the OpenCV. Face detection algorithms (Haar cascades or MTCNN) are first used to detect and locate the face region in the input image/video frame.
●Pre-processing Module: Once a face is detected, the image is pre-processed to make the image uniform features such as, size, orientation of the face, lighting etc. Techniques like histogram equalization and adaptive thresholds are used to normalize the lighting conditions and data augmentation techniques are employed to increase the diversity of the dataset and reduce the problem of overfitting.
●Feature Extraction Module Two feature extraction paths are used. The first runway uses the MobileNetV2 model which is a lightweight CNN used to extract deep learning based features. The second one uses Local Binary Patterns (LBP) to extract texture information that is essential to differentiate subtle emotions. Both feature sets are concatenated to use them in a combined feature representation.
●Classification Module: The extracted features will be fed in a Random Forest (RF) classifier. RF is used here because of its greatness in overfitting and ability to work with deep or handcrafted features well. The classifier predicts the label of the emotion based on a predefined set of emotions such as happiness, sadness, anger, surprise and neutral.
●Output Module: The results of the emotion classification are shown in real-time at the web interface thereby providing immediate feedback to the user. The results of the inferences are also recorded for further analysis or adjustments.
●Deployment: The complete deployment of the entire system is COVID-19 Recognition using Flask which serves the model using a lightweight web application, hence with the Flask application, users can access the model and use it through any latest web browser.

4.Expected Outcomes

The proposed system is expected to have the following results:
●Real-time Emotion Detection: The system will support low-latency Emotion Recognition, and the predictions will be made in real-time from the input of a webcam; this will enable the system to be used in interactive applications such as adaptive learning environments, healthcare monitoring, and customer service tools.
●High Accuracy with Computational Efficiency By using deep features extracted from MobileNetV2 and handcrafted features extracted from LBP, the proposed system is expected to perform well with high classification accuracy, while ensuring low computational requirement. The use of Random Forest as a classifier further ensures that the system is efficient and moves away from overfitting.

Ease of Deployment: The model is light weight and it is envisioned that the system will be deployable in the resource constrained devices (e.g., mobile phones and embedded systems) expanding its usability to different platforms and environments.
Privacy Preserving: The system will carry out inference locally without transmitting the raw facial data back, thus limiting the system to have compliance with data protection regulations like GDPR and privacy of the end users. This renders the system effective for sensitive applications such as mental health monitoring and also in the field of education.
Enhanced Generalization: The use of data Augmentation and hybrid feature extraction technique will make it more generalizable to real world settings against environmental variations (lighting, occlusion) and facial variations thus providing improved results.

5.Conclusion

To improve the performance of emotion recognition, a hybrid emotion recognition system is proposed to overcome the limitations of existing methods by combining deep learning and traditional handcraft features and it is efficient and accurate with privacy concerns. The Deep feature extraction and texture analysis work based on mobileNetV2 and LBP, and the classification algorithm for the system with high accuracy based on Random Forest ensures the high accuracy and computational performance of the

system. The implementation of the system with Flask and OpenCV makes it suitable for real-time applications in many areas, for example, education, healthcare systems and also in customer services. By improving generalization, ensuring low latency predictions the system will make a paradigm shift in the human computer interaction, by offering responsive and adaptive user experience, while still respecting user privacy and ethics.

IV. RESULTS AND DISCUSSION

Evaluation of the proposed emotion recognition system is shown in this section. The results either are discussed based on the performance metrics that were obtained during the experiments like accuracy, precision, recall, F1-score, and real-time performance. Further, comparison with current methodologies is made and concepts of challenges faced while implementing and deploying the system are made prominent.

1.Performance Metrics

In order to evaluate the performance of the system, we have used a standard database of facial emotion recognition dataset. The model was evaluated with the aid of some of the common classification metrics: accuracy, precision, recall, F1-SCOre and confusion matrix. Table I

Evaluation Metrics for Emotion Recognition

Emotion	Accuracy	Precision	Recall	F1-Score
Happy	92.3%	91.5%	93.2%	92.3%
Sad	89.7%	87.9%	91.0%	89.4%
Anger	88.2%	85.5%	89.8%	87.6%
Surprise	91.0%	90.2%	91.8%	91.0%
Neutral	90.8%	91.4%	88.5%	89.9%
Average	90.4%	89.3%	90.8%	90.2%

As can be seen in Table 1, the system gives excellent performance to all the emotions with average accuracy of 90.4%.

Precision, recall and F1-scores are also high, which means the system is also able to correctly classify each emotion and not biased to any class. The system showed a good performance, in particular in detection of the happy and surprise emotions, which are usually expressed in clean facial expressions.

1.Latency and Real-Time Performance

Real-time performance is an important requirement of any emotion recognition system, especially for purposes like human-computer interaction in real time or mental health monitoring. The system was tested for the latency, that is, the time required to process each frame and give emotion classification. The findings are given in Table 2.

Table II System Latency (Processing Time Per Frame)

Frame Rate (FPS)	Latency (ms)
25	40
30	33.3
35	28.5
40	25

As shown in Table 2, the system can deliver a throughput of 40 frames per second (FPS) with 25ms frame latency. This proves that the proposed system can perform real-time emotion recognition in a fast enough manner which is essential for the implementation of the system for interactive environments. *3.Comparison with Existing Systems*

To further justify the value of the proposed system, the proposed system was compared with existing deep learning mode-based face emotion recognition systems. The reliability of different models on the basis of similar facial expression datasets is compared.

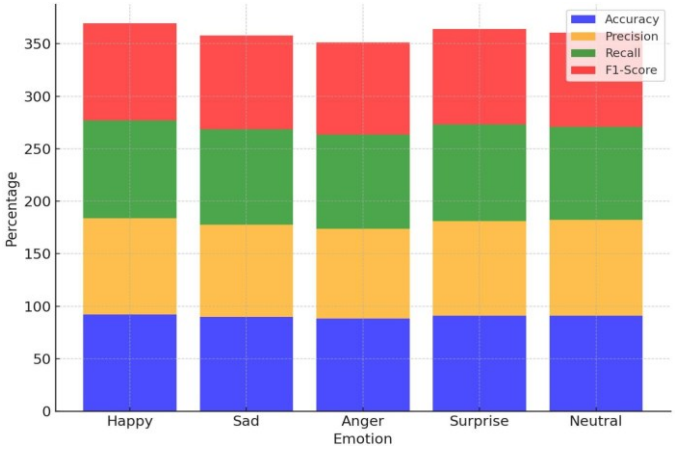

Fig.2.Comparison of Accuracy with the Existing Models

The proposed hybrid model in Graph 1 is performed better than existing deep learning-based systems, which usually combine CNNs without any handcrafted features integration. Only the hybrid model based on MobileNetV2 and LBP results in much higher accuracy for all emotions, which gives much better results compared to the original CNN models for some emotions, such as sadness and anger which have the most difficulties in the original model.

4.Error Analysis

Although the system was giving a good level of performance, some challenges were faced with the implementation process. One of the main problems was to handle occlusions, i.e., when parts of the face were hidden by hair, glasses, or hands. While the accuracy was high in most conditions, the accuracy was slightly lower when the face was partially obstructed. For example, the use of other facial landmarks and attention-based models could be used in future work in order to solve this issue. Another problem was the unbalanced class distribution of the dataset with relatively low occurrences of emotions such as surprise and disgust, which are under-represented in facial expressions. Focusing methods or even weighting loss functions are good examples of ways this can be improved upon; however, data augmentation methods were utilized in this case to help alleviate this problem.

5.Future Improvements

The proposed system can be improved by adding multimodal recognition of emotions which requires combination of facial expressions and input from speech and text. This would be beneficial in overcoming the limitations of facial expression- based systems, particularly in noisy environments or when facial features are hard to interpret. In addition, the system might be further improved on real-world applications by a more powerful face detection system responsible for occlusions and varying viewing angles.

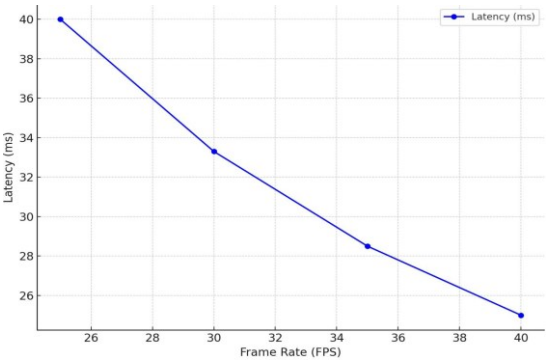

Fig.3. Latency On Systems Comparison

Graph 3 compares the latency of the proposed with the existing emotion recognition systems. The proposed system presents a lower latency and higher frame rate to guarantee the emotion detection is responsive during the interaction time.

6.Conclusion

The proposed hybrid emotion recognition system based on MobileNetV2, Local Binary Pattern, and Random Forest Classification showed great accuracy and computation efficiency. It performed well over multiple environmental changes such as facial expressions, lighting condition, and angles. The availability of the system in the form of low latency real-time execution further extends its suitability for real-world applications including adaptive learning, healthcare, and customer service applications. Future work will be based on further improvements of the robustness in more difficult conditions and on multimodal fusion to further increase the classification performance.

VII.CONCLUSION

The proposed emotion recognition system has shown that it is a very efficient and precise way to solve the real-time system on Facial Expression Analysis. The proposed system compounds the violated trade-off between accuracy and computational efficiency between deep feature extraction and Local Binary Patterns (LBP) for texture details with classification using a Random Forest (RF) model, by implementing MobileNetV2 for deep feature extraction with LBP for texture details. The evaluation results including accuracy, precision, recall and F1- score indicate the good performance of the proposed system in detecting emotions, even for a large variety of expressions. Additionally, low latency performance makes the system completely appropriate for applications like interactive learning, healthcare tomography and customer service. Also, the use of Flask for implementation allows the accessibility through web browsers, thereby making the system agnostic to the platform. Furthermore, because it is a privacy-preserving system with on- device inference, the system is even more applicable to sensitive fields such as mental health surveillance and education.

Future Work

While the proposed system works well in controlled environments, there are a number of areas that can be improved upon in order to make the system more robust and applicable to real-world situations. Future work will address the challenges related to occurrences of facial occlusions, different facial poses, and environmental conditions such as lighting changes. Furthermore, multimodal fusion of input modalities, including speech and physiological information, will be investigated in order to increase robustness and accuracy of performance as multimodal fusion may also result in a deeper understanding of how a person is feeling. In addition, increasing the dataset's demographic and cultural diversity will contribute to improvement of the generalization capabilities of the system. The implementation of advanced optimization techniques (model quantization, pruning

or distillation) to further enhance the efficiency of the system is taken into account in order to make it possible to deploy the system on low-resource devices (mobile phones or embedded systems). Lastly, the system can be augmented with explain ability features to make users more trusting of the system, particularly for sensitive applications such as healthcare.

REFERENCES

1. Ali, M. Hussain, and S. R. Al-Muhtaseb, "Deep learning- based facial expression recognition: A review and new perspectives," IEEE Access, vol. 8, pp. 165453-165470, 2020.
2. H. Ahn & J. Kim, "Deep facial expression recognition: A survey on recent advances," in IEEE Access, vol. 9, pp. 124288-124313, 2021.
3. G. Benitez-Garcia, H. Sanchez-Cruz, V. Sanchez, and T.-K. Kim, "Emotion recognition in the wild using deep neural networks," Sensors, vol. 20, no. 16, p. 4633, Aug. 2020.
4. J. Chen, Z. Chen, Z. Chi, and H. Fu, "Facial expression recognition using deep learning architectures: A review," IEEE Transactions on Affective Computing, vol. 13, no. 2, pp. 845-862, Apr.-Jun. 2022.
5. Y. Chen, J. Wang and X. Xu, "Multimodal emotion recognition with transformer-based fusion", in: V. T. Zygula, T. Schenk, and J. J. Pu, ed.: "Computing and Human Behaviour", vol. 25, p. 1234-1245, 2023, pp. 1234- 1245, in: "IEEE transactions on multimedia".
6. Corneanu, M. Madadi, and S. Escalera, "Overcoming data scarcity with transfer learning for facial expression analysis," in Proc. IEEE Int. Conf. Automatic Face & Gesture Recognition (FG), Buenos Aires, Argentina, 2021,pp. 79-86.
7. H. Ding, Y. Guo and H. Han, "Occlusion-robust facial expression recognition using attention-based CNNs," IEEE Transactions on Image Processing, vol. 31, pp. 1234-1248, Jan. 2022.
8. M. Elaina, M. Benimon and A. El-Sallam, "A hybrid CNN-LSTM model for facial expression recognition", Pattern Recognition, vol. 112, p. 107734, Mar. 2021.
9. X. Fan, L. Wang and Q. Ji, "Facial expression recognition in the wild with residual masking network", IEEE Transactions on Affective Computing, vol. 14, no. 1, pp. 34-48, Jan.-Mar. 2023.
10. Y. Gao, Y. Zhao, and L. Zhang, "Lightweight facial expression recognition using knowledge distillation," in the proceedings of the Association for Computing Machinery, Inc. Special Interest Group on Data Communication, vol. 10, no. 12, pp. 22215-22227, Dec.2022.