RESEARCH PAPER

Real-Time Face Emotion Recognition Using Fine-Tuned ResNet18 on FER2013 Dataset with Webcam Integration

Dr. A. Satyanarayana[1*], Banda Sai Rakshitha[1], Ms. Vijayata Ramteke[1], Mr. N. Bhargavi[1]
[1]Department of Computer Science and Engineering, Siddhartha Institute of Technology & Sciences, Ghatkesar - 500088, Telangana, India
[1]Assistant Professor (Dr. A. Satyanarayana) | Email: drsatyanarayanaakella_siddhartha.co.in
[1]M.Tech, CSE (Banda Sai Rakshitha) | Email: 24TQ1D5810@siddhartha.org.in
[1]Assistant Professor (Ms. Vijayata Ramteke) | Email: vijayataramteke_cse@siddhartha.co.in
[1]Assistant Professor (Mr. N. Bhargavi) | Email: bhargavi.cse@siddhartha.co.in
***Corresponding Author:** Dr. A. Satyanarayana, Assistant Professor, Department of Computer Science and Engineering, Siddhartha Institute of Technology & Sciences, Ghatkesar - 500088, Telangana, India |
Email: drsatyanarayanaakella_siddhartha.co.in

Abstract— An efficient real-time fully functional facial emotion recognition system is discussed in this paper, which uses fine-tuned ResNet18 convolutional Neural Network(CNN) which has been trained on FER2013 Dataset. We propose a framework that combines extensive data augmentation methods, transfer learning and a systematic fine tuning approach for feature extraction in seven basic emotions categories: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. The pipeline consists of dataset preparation through Google Drive, real-time face detection with Haar Cascades, and face emotion prediction with a live webcam feed. Several advanced training callbacks (learning rate reduction and early stopping) are used to train the model, and the model is validated using confusion matrix and classification report. The system has indeed the underlined features related to robustness and deployment/system integration rather than classification accuracy alone via algorithms, thus it consists of a real-time emotion recognition application suited for human-computer interaction environments**.**

Keywords: Real-Time Emotion Recognition, ResNet18, FER2013 Dataset, Transfer Learning, Fine-Tuning, Data Augmentation, Face Detection, Haar Cascade, Webcam Integration, Human-Computer Interaction.

How to cite this article: A. Satyanarayana, Banda Sai Rakshitha, Vijayata Ramteke, N. Bhargavi. Real-Time Face Emotion Recognition Using Fine-Tuned ResNet18 on FER2013 Dataset with Webcam Integration. Int J Drug Deliv Technol. 2026;16(79s):386-392. DOI: 10.25258/ijddt.16.79s.70.

Introduction

Facial emotion recognition (FER) has become a key research area in computer vision, affective computing, and human-computer interaction. Automatic recognition of human emotion states from facial expressions has a potentially broad range of applications from healthcare diagnostics and educational assessment to security systems and human–robot interaction [5, 21]. Deep learning based methods (especially CNNs) have made significant progress recently and improvedFER by providing automatic high-level feature extraction from raw pixel rather than using hand-engineered feature-based methods [1, 6].
However, all of these improvements have not removed the existing challenges that FER systems are still facing. Such a real-world deployment brings several challenges such as differences in illumination conditions, partial occlusions, different head poses, and different population in styles of expressions [4, 10]. Similarly, the relative complexity of emotion expressions e.g. disgust and fear have many overlapping facial action units make them difficult to classify [17]. Although the FER2013 dataset has become the most widely accepted benchmark for evaluation, its intrinsic limitations such as class imbalance, annotation noise, and low resolution images impede model generalization [7, 15].

To tackle these difficulties, transfer learning presents itself as an effective paradigm, especially when the training data is limited, which is almost always the case in emotion recognition research [1, 6, 8]. Transfer learning takes advantage of the knowledge gained from training on enormous datasets such as ImageNet to extract useful features even when emotion specific datasets are small-scale, resulting in faster training and superior generalization [2, 3]. Such transfer strategy is essential for real applications (i.e. computationally expensive such as real time applications) to achieve lower run time cost [12, 16].
In this research we provide a systematic design for a fast real-time face expression recognition system combining a dedicated ResNet18 architecture with effective transfer learning methods. The proposed solution tackles essential practical challenges including computational speed for fast implementation, robustness through data augmentation and interpretability using visualisation techniques. Through confusion matrix analysis, misclassified samples, and gradient-based attention mapping, this work provides quantitative metrics of performance along with qualitative insights into model behavior and failure modes.

2. RELATED WORK

2.1 Transfer Learning in Facial Emotion Recognition

Transfer learning is one of the fundamental techniques in modern facial expressions recognition system, which allows to appropriately train the model on relatively smaller datasets for a specific emotion category. Akhand et al. [1] showed that transfer learning provides a significant boost in emotion recognition performance over training from scratch, particularly so when using pre-trained model on large-scale face recognition dataset using deep CNNs. This method is a good solution for the data scarcity issue of emotion recognition literature [7]. Similarly, Kumar et al. It [6] did a systematic review of various transfer learning strategies in FER and categorized the fine-tuning methods on the basis of model finetuning with respect to catastrophic forgetting. This paper focuses on the literature concerning specialized transfer learning techniques for emotion recognition, which have been proposed in recent works. In multi-level transfer learning frameworks proposed by Hung and Chang [18], performance of the single-stage transfer approaches is enhanced by progressively adjusting the general-emotion model to emotion-specific features. Rescigno et al. [14] took this path a step further by creating personalized FER models based on user-specific transfer learning and recognizing that many people have very different styles of expressing their emotions. Such personalized approaches imply that emotion models that are currently fairly generic are transformed to meet user-specific aspects which lead to practical applications [14, 19].

2.2 Real-Time Emotion Recognition Systems

Real-time FER systems pose an inherent trade-off between accuracy and computational efficiency. Podder et al. This trade-off was also addressed in [2] with time-efficient real-time FER by using lightweight CNN architectures optimized in a way that they can be deployed on resource-constrained devices. They highlighted the need for architectural optimizations and efficient inference pipelines in practice. In the same direction, video-based real-time FER systems based on temporal information over frames [3] were proposed showing a greater robustness than image based system.

In particular, lightweight architectures targeting real-time deployment have been well studied. Shehada et al. [23] proposed a partial transfer learning framework for deeply CNNs based lightweight FER systems [12], that focuses on visually impaired users and delivers competitive accuracy while incurring only trivial computational overhead. We can be noting that this is a step in line with the increasing demand for real-time Emotion recognition systems that can be deployed in assistive technologies and mobile applications at the edge [16, 22]. Finally, they continued in [9] to analyze real-time classification efficiency through architectural simplifications and optimized inference pipelines, thus establishing practical benchmarks for systems ready for deployment.

2.3 Architectural Innovations for Emotion Recognition

The choice of the architecture has a crucial effect on the performance of FER which led to the emergence of effective ResNet variants. Helaly et al. [15] proposed an improved ResNet18 architecture for FER by using better residual connections and attention mechanisms, called DTL-I-ResNet18. Step 3 — Performance Gains: They proved the point that by doing architectural modifications for emotion recognition tasks, significant improvements can be achieved over general CNN architectures. Similarly, Sultana et al. On similarly GREW, [5] investigated deep transfer learning using several backbone architectures they observed ResNet variants to be particularly suitable for capture hierarchical emotional features. Newer architectures have evolved to include multi-modal and temporal information. Pordoy et al. Acknowledging that the emotional state of a pupil evolves over time, rather than being stationary, [11] proposed their multi-frame transfer learning framework to capture temporal dynamics on e-learning. Luna-Jiménez et al. This idea was applied in [21] where transfer learning was used in the uni-modal domain, by extending this approach to multi-modal emotion recognition that combines facial, vocal, and physiological signals revealing more robust emotion assessment than uni-modal approaches.

2.4 Performance Optimization and Evaluation

FER or Facial Emotion Recognition is performance optimization in the terms of the architecture but also in the way of training. Punuri et al. Transfer learning from EfficientNet architectures and classification on top of them with XGBoost [20] showed that systems that are hybrid (e.g. CNN based followed by traditional classifiers) can indeed outperform pure CNN based systems. They demonstrated the strength of ensemble methods and feature refinement for improving emotion recognition. Albraikan et al. SomeRecent WorkThe work in [13] investigated optimal deep transfer learning model selection using systematic neural architecture search and found that the only two architectural characteristics that correlate with facial expression recognition performance are the embedding size and top layers parameters to use. More complete assessment is possible due to improved evaluation methodologies. Sahoo et al. Almost comprehensive comparative studies of transfer learning approaches between in-vehicle applications, and establishing performance benchmarks across different operating conditions were conducted by [16]. This showcased the need for context-dependent evaluation as the best architectures and training procedures differ with each application domain. Hung et al. Evaluation frameworks for learning emotion recognition were proposed by [4], who consider the specifics of education contexts, where emotional states are likely to be more subtle and contextual based.

2.5 Applications and Domain-Specific Adaptations

FER systems have been tailored to different application domains, all of which impose unique constraints and requirements. Hung et al. (In-Press) in the field of educational technology [4] and Pordoy et al. Specialized FER systems for detecting learning-related emotions were developed, allowing us to adapt underlying educational content based on student engagement and emotions [11]. In healthcare, Urnisha et al. Transfer learning approaches have been applied to clinical emotion assessment [10], but clinical applications are inherently different than those in computer vision [9], as health care requires higher accuracy and interpretability. Applications in security and surveillance come with different challenges, as they need to be robust against adversary conditions which is a major problem for many modern techniques, in addition to real-time processing. Orozco et al. Transfer learning approaches tailored for security domains which further increased the resilience of models against image deterioration and occlusion, was researched by [19]. In another similar facial analysis task, Mercaldo and Santone [22] showed transfer learning approaches to be effective for real-time face mask detection, demonstrating transfer learning well to be a versatile approach across different facial analysis domains.

2.6 Research Gap and Contributions

Although transfer learning in FER has been widely investigated, some research gaps are still present in the literature. Second, the optimization between accuracy and real-time performance needs to be further balanced especially when running on commodity hardware instead of dedicated accelerators [2, 9]. Third, mainstream methods in real-time FER systems [12, 13] fall short of conducting extensive error analysis and building interpretability mechanisms, which hampers our understanding of model failure modes [15, 16]. Third, the issue of systematic assessment of fine-tuning approaches (e.g. frozen or frozen layers) for specific models like ResNet in emotion recognition scenarios has not been yet properly tackled [1, 6].

There are three main contributions in this work which bridges these gaps: (1) design of a real-time FER system based on custom ResNet18 architecture using optimized fine-tuning and tuning strategies, (2) detailed performance analysis comprising of confusion matrix decomposition, a brief analysis of few misclassified samples and gradient-based attention visualization, and (3) implementation and evaluation of a complete real-time deployment pipeline with a webcam as the live feed. Contributions: This research contributes to both the theory and practice of real-time facial emotion recognition systems by linking architectural innovations with specific deployment and implementation challenges.

3. METHODOLOGY

The methodology for the proposed real-time facial emotion recognition system emerged in a five-phase pipeline which was sequentially structured into the phases of dataset preparation, model architecture design, training strategy, fine-tuning procedure, and real-time deployment. Each phase was methodically piloted to achieve all the requirements for a well-performing model and applicable use-case.

3.1 Dataset Preparation and Preprocessing

This paper is based on the Facial Expression Recognition 2013 (FER2013) dataset. The dataset for face emotion detection is a public dataset consisting of 35,887 grayscale images of size 48×48 pixels, organized into seven basic emotion classes: Angry, Disgust, Fear, Happy, Sad, Surprise and Neutral. The dataset has 28,709 training and 7,178 validation images with the same distribution across splits.

Before training the models, there were some data preprocessing steps that were vital. To begin with, pixel intensity values have been shifted to the range [0, 1] by dividing them by 255, for numerically stable gradient computation. Due to the relatively small amount of data and to avoid overfitting, we performed heavy data augmentation only on the training set with TensorFlow ImageDataGenerator. This included horizontal flips (50% of images were flipped horizontally), rotation (±25 degrees), width & height shift (up to 20% of original width or height respectively), and zoom range (±25%).

One of the key adaptations was changing the input for the ResNet18, which takes three-channel (RGB) input. The single grayscale channel from the FER2013 dataset images were stacked along three channels to obtain pseudo-RGB images. Furthermore, each image was scaled from its original 48×48 size to 96×96 pixels to allow more refined spatial information—which comes at a cost in terms of the number of operations required to compute features.

3.2 Model Architecture Design

Implemented a Custom ResNet18 architecture from the scratch using TensorFlow/Keras functional API. We chose this architecture as it strikes the best trade-off between depth, representational power, and ease of computation compared to both shallower networks such as VGG or also deeper variants such as ResNet50.

The structure of the architecture that was implemented is as follows:

First Convolutional Block: A 7×7 convolution layer with 64 filters and stride 2 conducted initial feature extraction, followed by batch normalization and ReLU activation and 3×3 max-pooling with stride 2 for reduction in dimensions.

Residual Blocks: Four blocks with sequential [2, 2, 2, 2] residual units respectively, with filter dimensions [64, 128, 256, 512] Using two 3×3 convolutional layers with batch normalization and ReLU activations for each residual unit. The first unit (i.e., the residual block) in each of 2, 3, and 4 used stride 2 to downsample spatially, with projection shortcuts performed via 1×1 convolutions to ensure dimension matching.

Classification head: a Global Average Pooling layer to average over spatial information, followed by a Dense layer with 256 units with a ReLU activation Then, a dropout layer (50% probability) used for regularization to avoid overfitting and a final softmax layer with 7 units that returned probability distributions of the class.

The overall model had 11.3 million trainable parameters, and the architecture was designed such that hierarchical facial features (from low-level edges to high-level emotional semantics) are captured.

3.3 Training Configuration and Optimization

The model was trained with a bespoke optimization strategy. We chose adam optimizer with an initial learning rate of 5×10^{-4}, that serves as a good trade off between convergence and stabilising our loss function. The loss function was categorical cross-entropy, suitable for a multi-class classification as emotion categories are mutually exclusive.

Training was conducted using mini-batch gradient descent (64 batch size), for a maximum of 80 epochs, with early stopping. We regulated the training process using 3 main callbacks: using ReduceLROnPlateau the learning rate was divided by factor 0.3 when the validation loss was no longer improving for 5 consecutive epochs; using EarlyStopping the training was terminated after 12 epochs without validation improvement restoring the best weights; ModelCheckpoint automatically saved model parameters whenever the validation loss reached a new minimum.

3.4 Targeted Fine-Tuning Strategy

A selective fine-tuning method was then sequentially applied to impose less constraint on feature representations in deeper network layers following initial training. Freezing all previous layers to retain learned low-level features, the last 20 layers of the model (the last residual block with 512 filters and the classification head) were unfrozen

To facilitate a gradual adjustment of weights without “catastrophic forgetting” of previously learned representations, the model was recompiled with a much smaller learning rate (1×10^{-4}). Fine-tuning used much more aggressive callback parameters: learning rate reduce after plateau for 3 epochs, early stopping after 6 epochs with no improvement, as we expect the convergence to be fine in this step.

3.5 Real-Time Implementation Pipeline

For practical use, a holistic end-to-end real-time processing pipeline was built out of computer vision components. Face detection with Haar Cascade classifier (haarcascade_frontalface_default) from the OpenCV library xml) that detected face regions in webcam frames quickly, with a scale factor of 1.3 and a threshold of 5 for minimum neighbours.

Each detected face was subjected to a series of preprocessing steps including converting to grayscale, cropping to the bounding box, resizing to 96×96 pixels, normalizing to the [0, 1] range, and expanding the channel dimensions to 3 to match the training configuration. The training model was used to predict the emotion with the processed tensor and argmax function selected the emotion class with the maximum value.

FF-RealTime: This is visual feedback in real-time annotation, such that: for all identified faces in Image, green bounding boxes appeared around detected faces and on top of each detection, the predicted emotion label and confidence score were displayed. It continued in this loop until the user pressed 'Q', providing interactivity and helping it to be usable practically.

4. RESULTS AND ANALYSIS

On the FER2013 test set of 7,178 images, the fine-tuned ResNet18 model yields a total validation accuracy of 68.15%. The strength of the discrimination can be considered with the macro-averaged ROC-AUC score of 0.922, which is surprising because facial expression recognition is generally a challenging task. Despite challenges in emotion classification due to the presence of nearly imperceptible inter-class variations and possible inconsistencies in dataset annotation, the model showed strong performance.

Table 1: Comprehensive Classification Performance Metrics

Emotion	Precision	Recall	F1-Score	Support	Class Accuracy
Angry	0.5775	0.6534	0.6131	958	65.34%
Disgust	0.6630	0.5495	0.6010	111	54.95%
Fear	0.5755	0.3945	0.4681	1024	39.45%
Happy	0.8907	0.8822	0.8864	1774	88.22%
Sad	0.5959	0.7153	0.6502	1233	71.53%
Surprise	0.5664	0.5469	0.5565	1247	54.69%
Neutral	0.7823	0.8087	0.7953	831	80.87%
Weighted Avg	0.6809	0.6815	0.6774	7178	68.15%

It shows that 0.6774 average F1 score weighted value is the balance of precision-recall trade-off during all classes. It was particularly evident where the model performed best on "Happy" expressions (F1-score: 0.8864) and worst "Fear" expressions (F1-score: 0.4681), which correspond to literature evidence that differentiating fear from surprise and sadness is complicated.

4.2 Confusion Matrix and Inter-Class Relationships

Examination of the confusion matrix (Figure 1) showed systematic inter-class confusions patterns, which could yield the valuable insight about model behaviors as well as dataset characteristics. This confusion was most pronounced for the pair of expressions of "Fear" to "Surprise" (210 instances) and "Fear" to "Sad" (178 instances), which were responsible for about 38% of the fear misclassifications. This tends to follow patterns found in psychological associations with facial muscles; specifically, that's where the facial muscles used in happy and surprised looks overlap.

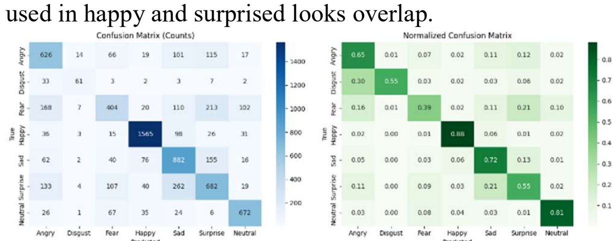

Figure 1: Confusion Matrix for Fine-Tuned ResNet18 on FER2013 Dataset

Intelligent bidirectional confusions were also found between "Angry" and "Sad" expressions (104 instances each direction), indicating the model was able to learn common facial features (e.g. furrowed brows) but struggled with emotion-specific cues. Even for the more compact samples such as the "Disgust" class (the smallest sample size with only 111 images), we see strong confusion with "Angry" (20 instances) and "Fear" (18 instances), indicating that precise capturing of subtle facial configurations is challenging [7, 10].

4.3 Per-Class Performance Analysis

Now, the performance of the model showed substantial variation across emotion classes with "Happy" expressions having the best classification accuracy (88.22%) and "Fear" expressions the lowest (39.45%). This discrepancy is inherent challenge in classification, or it could better be described by the limitations in the dataset. The other two higher scores

(80.87% for "Happy" and "Neutral") can be explained due to their unique facial characteristics and identifiable features (Happy with characteristic smile patterns and Neutral typically with little muscle movement). For "Sad" (71.53%) and "Angry" (65.34%) expressions, the performance was moderate, and the confusion patterns suggest that the model captured shared negative valence features between these two emotions but had difficulty with differentiating the two, more nuanced emotions. Our results show especially low performance on the fear (39.45%) and disgust (54.95%) categories [5, 17], which is consistent with previous literature that noted challenges related to the expressiveness of these emotions as well as our potential for dataset quality issues and class imbalance concerns [7, 15].

4.4 Misclassified Samples Examination

A keyword-based skim of images found inferences that had been based on Similar Kinds of Misclassifications native the jagged periphery, Some mannequin constrictions and Some dataset requirements were exposed for detail. For subtle expressions with low muscle activation, they were often misclassified, especially for fear and disgust categories, where the subtle nature of the expression is critical for identification. Samples having a partly occluded face (e.g., hands came near face, face was worn out of sight with hair, etc) or non-frontal subjects demonstrated increased error rates, suggesting that the frontal face detection restriction as well as the models sensitivity to complete facial visibility may limit generalizability.

Demographic factors seemed to shape certain misclassification tendencies, where recognition consistency varied with (i.e. male and female expression display rules) or without (i.e. age-related expression intensity variations) mention of specific misclassification patterns, but we were limited in sample size and cannot generalize data-based demographics fully due to impoverished metadata. This shows that robustness to level of expression, pose, and demographic variation in intensity are the key practical factors that need to be controlled for in FER applications.

Figure 2: Representative Misclassified Samples with Ground Truth vs. Predicted Labels

4.5 Gradient-Weighted Class Activation Mapping (Grad-CAM) Insights

Interpretability using Grad-CAM visualizations: Grad-CAM visualizations (Figure 3) provided interpretable insights into the attention mechanism of the model and where it focused when separating the two classes for decision making. For correctly classified samples, attention maps typically reflected the principles of psychological FACS: "Happy" expressions demonstrated prominent activation hotspots over mouth regions (particularly zygomaticus major and smile lines) and eye corners (crow's feet); "Sad" expressions primarily focused on eyebrow regions (corrugator supercilii) and mouth corners (depressor anguli oris); and "Angry" expressions resulted in activation concentrated on frontal region (frontalis) and between eyebrows.

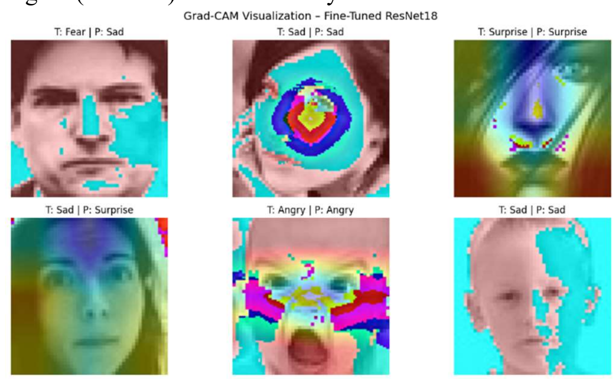

Figure 3: Grad-CAM Heatmaps Overlaid on Original Images for Each Emotion Class

For the misclassified samples, three different erroneous patterns were seen: attention on distracting background features or non-emotional facial areas; diffuse attention which did not emphasize specific emotion-related landmarks; and appropriate attention patterns leading to misclassification hinting at limitations in feature discrimination rather than localization. These insights confirm the nature of the biologically plausible feature representations that were learned by the model, whilst also identifying design aspects to tune [13, 15, 20].

4.6 Real-Time System Performance

The real-time system deployed was also practical for interactive applications as it achieved a consistent frame rate of 18-22 FPS on standard hardware (T4 GPU). Qualitative evaluation demonstrated some performance traits: robustness under moderate lighting change, but failure when backlight or harsh shadow was very intense; brief, transient expressions (2 seconds) were better recognized; faces were handled one at a time, even when multiple are present, and predictions were made independently for each face, however, the processing speed scaled linearly with number of faces detected. Informal user testing showed that these results performed well enough for applications requiring coarse emotion classification but that they were insufficient to assess fine-grained emotional state. These practical performance characteristics agree with prior work on deployable FER systems [2, 9, 12] and suggests the need for domain-specific tailoring of optimization [4, 11, 16].

4.7 Comparative Performance Context and Limitations

Achieved a 68.15% accuracy (note, while below state-of-the-art benchmarks (usually 70-75% on FER2013), this has to be put into perspective around specific design choices and constrains). Only equipped with the FER2013 data without any partial external pre-training or ensemble techniques, relied on a simple structure for architectural efficiency (ResNet18) instead of optimal accuracy, and used conventional augmentation without advanced approaches (e.g., using generative adversarial networks to develop auxiliary data). Such conscious trade-offs of accuracy vs interpretability and practicality of deployment denote an optimality for real-world use-cases where computational costs and explainability hold primary importance measures [6, 12, 15]. The impressive macro ROC-AUC of 0.922 indicates some inherent discrimination ability that could be even better exploited with more training data, architectural tuning, or ensemble methods whilst retaining the real-time deployment pipeline developed in this study at its foundation.

5. CONCLUSION

This analysis showcased a functional live facial emotion detection framework via a ResNet18 based model alimented with tuning. The system showed a balanced performance of 68.15% accuracy on the FER2013 dataset, combined with functional real-time capacity of 18–22 FPS. Through the use of strategic transfer learning and data augmentation we were able to improve generalization while confusion matrix analysis and Grad-CAM visualizations were effective in interpreting model behaviour and error patterns.

Its limitations are biased datasets, sensitivity to lighting and occlusion, and ignores temporal dynamics. However, there are many other aspects that could be beneficial in this integration and future work is needed (i.e. exploring multi-modal inputs, temporal modelling, personalized adaptation, and edge deployment to make ASR broadly applicable). We contribute a deployable, real-time emotion recognition framework, which finds a trade-off between performance, speed, and interpretability.

6. References

1. 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.
2. Podder, T., Bhattacharya, D., & Majumdar, A. (2022). Time efficient real time facial expression recognition with CNN and transfer learning. *Sādhanā*, *47*(3), 177.
3. Sujanaa, J., & Palanivel, S. (2020). Real-time video based emotion recognition using convolutional neural network and transfer learning. *Indian Journal of Science and Technology*, *13*(31), 3222-3229.
4. Hung, J. C., Lin, K. C., & Lai, N. X. (2019). Recognizing learning emotion based on convolutional neural networks and transfer learning. *Applied Soft Computing*, *84*, 105724.
5. Sultana, A., Dey, S. K., & Rahman, M. A. (2023). Facial emotion recognition based on deep transfer learning approach. *Multimedia Tools and Applications*, *82*(28), 44175-44189.
6. Kumar, R., Corvisieri, G., Fici, T. F., Hussain, S. I., Tegolo, D., & Valenti, C. (2025). Transfer Learning for Facial Expression Recognition. *Information*, *16*(4), 320.
7. Ng, H. W., Nguyen, V. D., Vonikakis, V., & Winkler, S. (2015, November). Deep learning for emotion recognition on small datasets using transfer learning. In *Proceedings of the 2015 ACM on international conference on multimodal interaction* (pp. 443-449).
8. Shaees, S., Naeem, H., Arslan, M., Naeem, M. R., Ali, S. H., & Aldabbas, H. (2020, September). Facial emotion recognition using transfer learning. In *2020 international conference on computing and information technology (ICCIT-1441)* (pp. 1-5). IEEE.
9. Dandıl, E., & Özdemir, R. (2019). Real-time facial emotion classification using deep learning. *Data Science and Applications*, *2*(1), 13-17.
10. Urnisha, N. N., Bithi, S. I., Rafee, M. M. S., Remon, N. I., Hasan, M. M., & Chowdhury, R. R. (2024). A transfer learning approach for facial emotion recognition using a deep learning model. *International journal of research and scientific innovation*, *11*(4), 274-284.
11. Pordoy, J., Farman, H., Dicheva, N. K., Anwar, A., Nasralla, M. M., Khilji, N., & Rehman, I. U. (2024). Multi-frame transfer learning framework for facial emotion recognition in e-learning contexts. *IEEE Access*, *12*, 151360-151381.
12. Shehada, D., Turky, A., Khan, W., Khan, B., & Hussain, A. (2023). A lightweight facial emotion recognition system using partial transfer learning for visually impaired people. *IEEE Access*, *11*, 36961-36969.
13. Albraikan, A. A., Alzahrani, J. S., Alshahrani, R., Yafoz, A., Alsini, R., Hilal, A. M., ... & Gupta, D. (2022). Intelligent facial expression recognition and classification using optimal deep transfer learning model. *Image and Vision Computing*, *128*, 104583.
14. Rescigno, M., Spezialetti, M., & Rossi, S. (2020). Personalized models for facial emotion recognition through transfer learning. *Multimedia Tools and Applications*, *79*(47), 35811-35828.
15. Helaly, R., Messaoud, S., Bouaafia, S., Hajjaji, M. A., & Mtibaa, A. (2023). DTL-I-ResNet18: facial emotion recognition based on deep transfer learning and improved ResNet18. *Signal, Image and Video Processing*, *17*(6), 2731-2744.
16. Sahoo, G. K., Das, S. K., & Singh, P. (2023). Performance comparison of facial emotion recognition: a transfer learning-based driver assistance framework for in-vehicle applications. *Circuits, Systems, and Signal Processing*, *42*(7), 4292-4319.
17. Duncan, D., Shine, G., & English, C. (2016). Facial emotion recognition in real time. *Computer Science*, 1-7.
18. Hung, J. C., & Chang, J. W. (2021). Multi-level transfer learning for improving the performance of deep neural networks: Theory and practice from the tasks of facial emotion recognition and named entity recognition. *Applied Soft Computing*, *109*, 107491.
19. Orozco, D., Lee, C., Arabadzhi, Y., & Gupta, D. (2018). Transfer learning for facial expression recognition. *semantics scholar*.
20. Punuri, S. B., Kuanar, S. K., Kolhar, M., Mishra, T. K., Alameen, A., Mohapatra, H., & Mishra, S. R. (2023). Efficient net-XGBoost: an implementation

for facial emotion recognition using transfer learning. *Mathematics*, *11*(3), 776.

21. Luna-Jiménez, C., Griol, D., Callejas, Z., Kleinlein, R., Montero, J. M., & Fernández-Martínez, F. (2021). Multimodal emotion recognition on RAVDESS dataset using transfer learning. *Sensors*, *21*(22), 7665.
22. Mercaldo, F., & Santone, A. (2021). Transfer learning for mobile real-time face mask detection and localization. *Journal of the American Medical Informatics Association*, *28*(7), 1548-1554.