

A Transfer Learning Approach for Facial Emotion Recognition Using a Deep Learning Model

Nahia Nowreen Urnisha¹, Sanjida Islam Bithi², Md. Mushtaq Shahriyar Rafee³, Nasif Istiak Remon⁴, Md. Mahfujul Hasan⁵, Rajarshi Roy Chowdhury⁶

^{1, 2, 3, 4, 5} Department of Computer science and Engineering, Metropolitan University, Sylhet, Bangladesh

⁶Department of Computer science, American International University-Bangladesh, Dhaka, Bangladesh

DOI: <https://doi.org/10.51244/IJRSI.2024.1104022>

Received: 12 March 2024; Revised: 23 March 2024; Accepted: 27 March 2024; Published: 02 May 2024

ABSTRACT

The facial expression recognition (FER) system is the process of identifying the emotional state of a person. Emotion recognition from facial expressions is a rapidly growing area of research with numerous applications, including psychology, marketing, and human-computer interaction. This paper presents a novel approach to FER employing deep learning techniques, specifically leveraging the power of transfer learning methodology. To address the challenge of accurately recognizing different emotions, including angry, disgusted, afraid, happy, sad, surprised, and neutral, from facial expressions, a pre-trained model MobileNetV2 architecture has been fine-tuned for improving accuracy. It is a lightweight convolutional neural network architecture, specifically designed for efficient on-device inference. For evaluating the performance of the proposed FER model, the state-of-the-art FER-2013 dataset, along random images and video clips has been employed. Experimental results demonstrate that the proposed FER model attained a remarkable accuracy rate exceeding 99% when tested on diverse sets of random images and video clips. Moreover, the system achieved a notable accuracy of 61% when evaluated against the FER-2013 dataset. Overall, this approach presents a significant advancement in the field of real-time facial expression recognition using the MobileNetV2 architecture with FER2013 dataset, which may improve the quality of human-computer interactions.

Keywords: Facial Expression Recognition (FER), Convolutional Neural Network (CNN), MobileNetV2, FER2013, Transfer Learning.

INTRODUCTION

Facial expression is a pivotal aspect of understanding human emotions and intentions, a subject that has undergone extensive study and witnessed considerable progress in recent decades. Due to complexity and diversity inherent in facial expressions it remains challenging to achieve high accuracy [1], [2]. Recent strides in human-centered computing have given rise to the creation of autonomous systems designed for precise and dependable facial expression analysis. A prevalent approach for such analysis and classification involves the utilization of deep learning techniques, including recurrent neural networks (RNNs), convolutional neural networks (CNNs) [3] [4], and various CNN architectures like MobileNetV2, AlexNet, VGG [5], and ResNet [6]. These models are typically trained across a diverse array of classes. This broad training enables them to potentially develop universal features, excelling in the discrimination between

various objects. However, the intricacies of emotion recognition pose distinct challenges, requiring the model to adeptly distinguish elusive facial features. FER systems can also enhance cyber security issues by providing an additional layer of authentication beyond traditional methods like passwords or PINs [7]. By analyzing facial expressions, these systems can detect unauthorized access attempts by identifying suspicious behavior or anomalies in user interactions. Furthermore, integrating facial emotion recognition can help prevent unauthorized users from accessing sensitive information or performing malicious activities by adding a biometric verification step to the authentication process [8], [9]. Mehrabian's findings emphasize the significant role of facial expressions in the communication and exchange of information [10]. The exploration of facial expression recognition (FER) has received substantial interest across various fields, including psychology, computer vision, and pattern recognition. As technology advances swiftly, the integration of FER schemes has become widespread across numerous domains. These include human-computer interaction, virtual reality, augmented reality, advanced driver assistance systems, education, and entertainment. Consequently, designing an efficient FER model is imperative to enhance classification performance in these diverse applications.

In this study, a facial expression recognition system has been proposed employing transfer learning approach, whilst both public and private datasets have been utilized for evaluating the model performances. To tackle the challenge of accurately distinguishing various emotions, including anger, disgust, fear, happiness, sadness, surprise, and neutrality, from facial expressions, a pre-trained MobileNetV2 architecture is fine-tuned for improved accuracy. Experimental results have shown that the proposed FER system achieved accuracies surpassing 94% and 61% on the training and test datasets, respectively, using the state-of-the-art FER2013 dataset. Furthermore, the system demonstrated over 99% accuracy in identifying facial expressions from real-time videos and random images. The main contributions of this work are:

1. Developing an effective DL-based facial expression recognition (FER) model capable of identifying emotions from videos and images.
2. Creating private datasets comprising images and video clips to assess and evaluate the performance of the proposed FER model.
3. Evaluating the fine-tuned transfer learning-based FER model using both the private datasets and a widely used public dataset (FER-2013 dataset).

The remainder of this paper is organized as follows. **Section 2** describes related works on FER systems and deep learning approaches. Facial expression recognition datasets, MobileNetV2 architecture and the proposed model are given in **Section 3**, followed by an exploration of the model performances using different public and private datasets and deep learning (DL) algorithm in **Section 4**. Finally, **Section 5** concludes the paper with future direction of work.

REALATED WORK

Automatic facial expression recognition system describes two vital aspects, including facial feature representation and identification problems. Facial feature representation extracts a set of features from the original face images using different types of algorithms, including histogram of oriented gradient (HOG), scale-invariant feature transform (SIFT), gabor filters, and local binary pattern (LBP), for describing individual faces [11], [12]. On the other hand, for facial expression identification or classification task different types of machine learning (ML) algorithms are used, such as convolutional neural network (CNN) and transfer learning. Many researchers have utilized various CNN architectures [3], [4], [11], [12], [13], [14], [15], [16], [17], [18], including VGGNet, deep convolutional neural network (DCNN), and CNN architectures, for the facial expression recognition task, due to its automatic feature extraction capability and computational efficiency. In this study, mainly focused on the different methods of facial expression recognition methods.

Khairuddin Y. and Chen Z. [5] proposed a facial emotion recognition model based on deep learning, employing the VGGNet architecture with carefully tuned hyperparameters. In their approach, the authors incorporated Cosine Annealing to enhance the proposed model performance. Notably, the proposed model demonstrated a state-of-the-art single-network accuracy of 73.28% when tested on the FER2013 dataset, which encompasses 7 distinct emotions. The authors refrained from utilizing additional training data during the model training process. In reference [18], the authors presented a FER model that utilizes an architecture similar to VGGNet for the identification of facial expressions. The proposed FER model achieves an accuracy of 69% when tested on the FER2013 dataset, which comprises six different facial expressions, excluding instances of Disgust. To address overfitting concerns during training, the authors applied a data augmentation technique, and a total of 75 epochs were employed to enhance overall classification performance.

TABLE 1: Some of the key existing FER works

Source	Problem	Algorithm	Dataset	Types of Emotion
[5]	Facial Emotion Recognition	VGGNet	FER2013	7
[18]	Facial Emotion Recognition	Similar to VGGNet	FER2013	6
[3]	Facial Emotion Recognition Video Stream (FER)	CNN	FER2013 Dataset ¹	7
[4]	Facial Emotion Recognition	DCNN	Dataset ²	5
[19]	Facial Emotion Recognition, Face Recognition	CNN	Ck+48 and Private	5

Note: Dataset¹ – private dataset, Dataset² – manually collected using Camera

Shah D. et al. [3] introduced a FER model employing a CNN-based architecture to identify real-time facial emotions. The model demonstrated an accuracy exceeding 63% when evaluated on the FER2013 dataset, with the training accuracy surpassing 85%. Additionally, the model demonstrated the ability to identify real-time facial emotions from a video stream, specifically a webcam stream. In this process, the model initiates by detecting faces within the video stream using the Haar cascade algorithm, and subsequently employs the CNN architecture to identify various emotions with high accuracy. In reference [4], the authors presented a FER model based on the analysis of the deep CNN architecture for identifying facial emotions. A total of 2550 images with 5 different categories, including angry, happy, neutral, sad, and surprised, employed for evaluating the proposed model performances, whilst images are captured manually using a 48-megapixel (MP) camera. The model demonstrated an average accuracy of 78.04% when evaluated on this private dataset.

Dwijayanti S. et al. [19] introduced a modified version of the CNN architecture designed for the identification of individual faces and their corresponding facial expressions. The authors employed both a public dataset (ck+45 with 5 emotions [20]) and a private dataset consisting of 30 electrical engineering students to evaluate the performance of the proposed model. Experimental results indicated that the proposed model achieved accuracies of 87% and 67% for face and emotion recognition, respectively, using these datasets. Furthermore, the authors presented a method for measuring the distance between object and robot.

TABLE 2: Some of the existing FER models Strengths and Weaknesses

Source	Problem	Strengths	Weaknesses
[5]	FER	Higher accuracy	Longer training time
[18]	FER	Moderate accuracy	Limited number of facial emotions, longer training time
[3]	FER	Moderate accuracy	Longer training time
[4]	FER	Higher accuracy	Limited number of facial emotions, less training time, private dataset
[19]	FER, FR	Moderate accuracy	Limited number of facial emotions, longer training time

Note: FER – Facial Emotion Recognition, FR – Face Recognition, TFA- Transfer Learning Approach

Table 2 provides a concise overview of prior research works, highlighting their respective strengths and weaknesses. This analysis serves to pinpoint areas of improvement for enhancing the performance of FER models.

PROPOSED METHODOLOGY

A. Dataset and Data Preprocessing

Evaluation of the proposed model has been utilized multiple datasets, including FER2013 [21], random images from online, and two private video clips datasets. The FER2013 dataset comprises a total of 35,887 grayscale images (48 x 48 pixels), categorized into seven different emotions: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. Sample images are presented in Figure 1 and Table 3 provides a detailed distribution of the number of images corresponding to each expression category.

TABLE 3: Image Distributions

Category	Number of Images	Category	Number of Images
Happy	8989	Angry	4953
Neutral	6198	Surprise	4002
Sad	6077	Disgust	547
Fear	5121	—	—



Fig. 1 Sample images of seven types of facial emotions (FER-2013 [21])

The min-max normalization technique has been employed to normalize the pixel values of grayscale images

(48 x 48 pixels) to a range of 0 to 1, effectively standardizing the dataset. To align with the requirements of the pre-trained MobileNetV2 [15], [22] model, which operates most efficiently on 224 x 224 pixels images, the grayscale images have undergone a transformation to this resolution before being utilized in the proposed model. This resizing ensures a smooth integration with the pre-trained MobileNetV2 model, optimizing the processing efficiency for enhanced performance. Table 4 presents a brief description of the datasets utilized in this study.

TABLE 4: Details of the datasets

Dataset	Type	No of Instances	Source
FER-2013	Public	35,887	[21]
Random Images	Public	50	—
Video Clips	Private	2	—

B. MobileNetV2 Architecture

In the realm of image processing and computer vision, CNN has gained significant attention owing to their substantial economic potential and high accuracy rates. Notable CNN architectures [14], [22], [23], including AlexNet, VGG16, InceptionV3, ResNet, and MobileNetV2, have emerged as popular choices for image classification or identification tasks, each known for its distinct characteristics. While architectures like AlexNet, VGG16, InceptionV3, and ResNet are renowned for their large and deep network structures, they come with increased processing time and computational costs. Conversely, MobileNetV2 stands out due to its inverse residual structure and linear bottleneck design, which effectively reduces convolution calculations. Figure 2 illustrates the conceptual layout of this architecture. In this study, a fine-tuned MobileNetV2 architecture has been utilized for identifying facial expressions, which is selected not only simplicity of its structure but also efficient memory management. It's particularly well-suited for mobile applications.

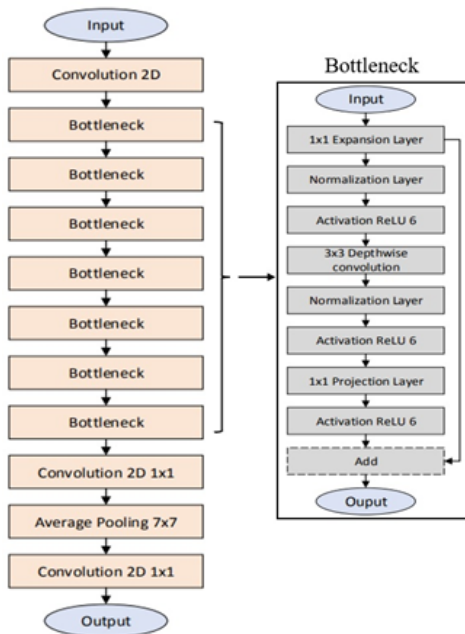


Fig. 2 An abstract design of a MobileNetV2 architecture [24]

The MobileNetV2 [22] architecture is a popular option for transfer learning when tackling the task of facial emotion recognition. Transfer learning [25] is a sophisticated machine learning (ML) approach [23], [26], [27], [28], [29], [30], wherein a pre-trained model is leveraged for solving similar types of new tasks. Traditional ML approaches necessitate training and testing datasets from the same domain, whereas the

transfer learning approach enables the utilization of different datasets for testing with similar characteristics. This methodology proves beneficial in addressing substantial challenges, including data limitations and the high cost of training in ML approaches. The MobileNetV2 model is pretrained with the ImageNet dataset, which consists of a total of 1.4 million images with 1000 object classes, including Persian cat, mountain bike, hot dog, etc. [15].

C. Proposed FER Model

The architecture of the proposed facial expression recognition scheme is illustrated in Figure 3, comprising four main sections. The initial stage of the process involves gathering datasets, encompassing both images and video clips, as depicted in the first section. Subsequently, the collected raw dataset undergoes preprocessing procedures in the second section to refine its quality and prepare it for analysis. This raw dataset comprises a combination of publicly available dataset, such as FER-2013 dataset as well as a private dataset (students video clips), contributing to the diversity of the dataset pool. Within this model framework, the initial 48 x 48 pixel images are transformed into uniform 224 x 224 images utilizing the `cv2.resize()` module, facilitating compatibility with the pre-trained MobileNetV2 architecture. Diverse modules, such as `cv2.resize()` and Haar-cascade, are harnessed to execute distinct tasks like image cleaning, resizing, and image recognition. These modules play crucial roles in ensuring the efficacy and accuracy of image processing tasks within the model architecture. Additionally, an object detection algorithm like Haar-cascade [31] is applied to identify faces in real-time videos. In section three, the configuration of a pre-trained MobileNetV2 architecture, where hyperparameters are fine-tuned for both training and testing phases, integrating the concept of transfer learning. MobileNetV2 [22] is a state-of-the-art CNN architecture designed for efficient mobile and embedded vision applications, featuring inverted residual blocks and linear bottlenecks. It achieves a balance between accuracy and computational efficiency, making it suitable for various real-time image classification and object detection tasks on resource-constrained devices.

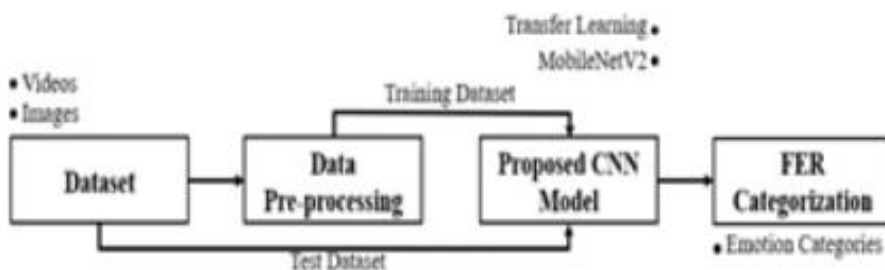
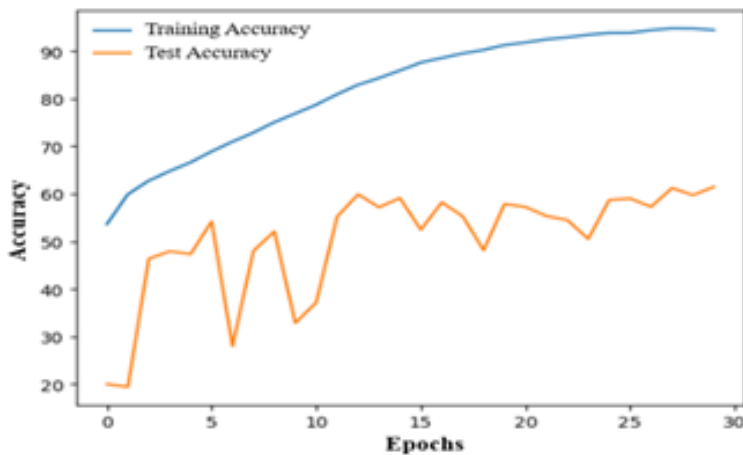


Fig. 3 The proposed FER model using Transfer Learning

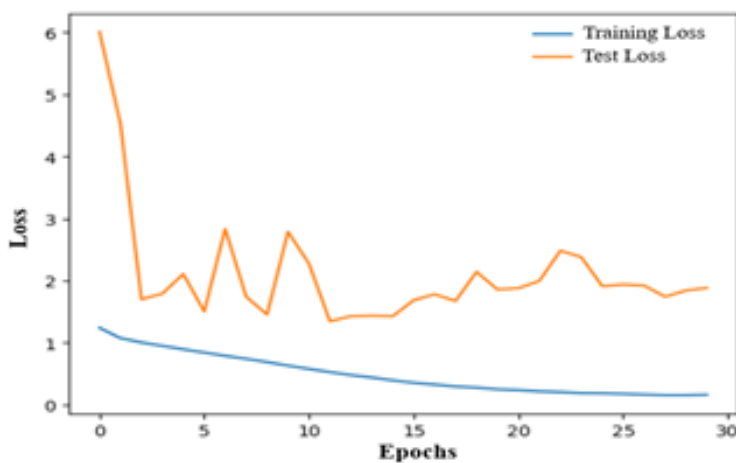
The top layers of the MobileNetV2 architecture have been omitted, as these were initially designed for the ImageNet dataset with 1000 classes. Instead, fully connected layers with seven output nodes (representing 7 classes: angry, disgusted, afraid, happy, sad, surprised, and neutral) have been incorporated into the model to align with the objectives of the proposed facial expression recognition scheme. The top layers consist of a series of fully connected layers that connect to the final output layer, responsible for making decisions based on features learned from preceding layers. Although removing the top layers renders the model unable to make the final decision, the output consists of feature maps from the last convolutional layer. To compile the MobileNetV2 model, the following configurations have been employed to enhance overall performance: loss function: sparse categorical cross entropy (suitable for multi-class classification tasks), optimizer: Adam (an extended version of stochastic gradient descent known for faster computation time and requiring fewer parameters for tuning) and metrics: accuracy (measuring the ratio of correctly predicted samples over the total number of samples). Finally in section four, the proposed model's performance is assessed using various public and private datasets with 30 epochs.

RESULTS AND DISCUSSION

The proposed FER model has been evaluated on a Windows operating system (OS), utilizing Python as the programming language. The implementation involves a fine-tuned MobileNetV2 architecture, incorporating various libraries and modules, including Tensor Flow, Keras, Pandas, Matplotlib, Numpy, Haar-cascade, and OpenCV, for both training and testing purposes. The evaluation of the proposed model performances is conducted using the FER-2013 [21] dataset, alongside online images (random images) and two private video clips datasets.



(a) Training and testing accuracies



(b) Training and testing losses

Fig. 4 FER-2013 dataset: (a) Training and testing accuracies in identifying different expressions from images in each epoch; and (b) Training and testing losses in identifying different expressions in each epoch

The proposed FER model has been trained using the FER-2013 dataset, employing 30 epochs to capture significant features, as depicted in Figure 4 (a). A total of 28,709 facial expression images (distributed across emotions as follows: Happy 7215, Neutral 4965, Sad 4830, Fear 4097, Angry 3995, Surprise 3171, Disgust 436) and 7,178 images (distributed across emotions as follows: Happy 1774, Neutral 1233, Sad 1247, Fear 1024, Angry 958, Surprise 831, Disgust 111) have been utilized for training and testing, respectively. From the experiment it has been observed that the training accuracy stabilizes after approximately 25 epochs, while test accuracy exhibits random fluctuations throughout the training period, as illustrated in Figure 4 (a). Conversely, Figure 4 (b) reveals a significant decline in the loss value for both the training and test datasets.

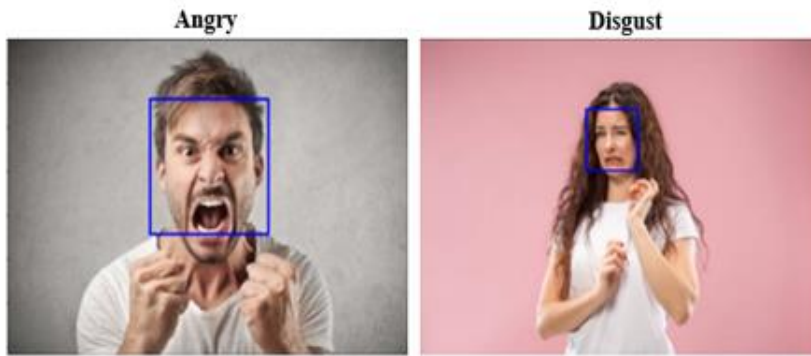


Fig. 5 FER identification performances using random images.

Figure 5 illustrates the classification performance of the proposed facial expression recognition (FER) model in identifying various types of emotions from random images. The experimental results have demonstrated that the proposed FER model gains an accuracy exceeding 99% when evaluated on the selected images (two images) sourced from an online platform.



Fig. 6: FER identification performances using video clips

In Figure 6, the facial expression identification performances of the proposed FER model are showcased on real-time video clips. The results indicate classification performances exceeding 99% in accurately identifying individual facial expressions from these video clips. In this experiment, the classification performances are higher compared with other experimental results. It may be the reason for image quality recorded from video clips. This suggests that image quality may significantly impact the accuracy of classification outcomes, demanding further investigation into its influence on model performance. This experiment employs two distinct video clips for facial expression recognition, featuring content generated by two different individuals. Notably, for privacy concerns, the video frames in Figure 6 have been intentionally obscured.

TABLE 5: Comparison of the proposed FER model with the existing approaches

Source	Task	Dataset	Performance
[5]	Facial Emotion Recognition	FER-2013	73.28% ⁷ cat 50 Epoch

[18]	Facial Emotion Recognition	FER-2013	69% 6 cat 75 Epochs
[3]	Facial Emotion Recognition	FER-2013	63.2% 7 cat 50 Epochs
	Video Streams (FER)	Dataset ¹	None ¹
[4]	Facial Emotion Recognition	Dataset ²	78.04% 5 cat 11 Epochs
[19]	Facial Emotion Recognition	Ck+ 48 and Private	67% 5 cat 500 Epochs
	Face Recognition		
*	Facial Emotion Recognition	FER-2013	61% 7 cat 30 Epoch
	Random Images (FER)	Dataset ^a	100% 7 cat 30 Epoch
	Video Clips (FER)	Dataset ^b	100% 7 cat 30 Epoch
Note: * The proposed FER model, Dataset ^a – Random images (online), Dataset ^b – Video clips, Dataset ¹ – private dataset, Dataset ² – manually collected using Camera, cat – Categorization			

Table 5 provides a comparative overview of several existing facial expression recognition approaches alongside the proposed deep learning-based FER model. Notably, prior FER approaches have typically relied on either large dataset [12], a substantial number of epochs (e.g., 75 and 500 epochs [8]), or a limited number of expressions (5 and 6 [8] [9] [12]) to enhance classification performance. In contrast, the proposed FER model utilized a smaller dataset featuring 7 expressions. Remarkably, the proposed FER model not only demonstrates the ability to recognize expressions but also excels in accurately classifying expressions from random images and video clips.

CONCLUSION

A real-time facial emotion recognition system involves the immediate and automated identification of emotions from a person’s face, offering insights into their emotional state. This technology has applications in diverse fields, including human-computer interaction, virtual reality, mental health monitoring, and personalized user experiences. In this study, the proposed real-time facial emotion recognition model, based on the MobileNetV2 architecture, demonstrates accuracies exceeding 99% on random images and video clips datasets. However, the model performance experiences a slight decline when tested on the state-of-the-art FER-2013 dataset, despite achieving a training accuracy of over 94%. The key strength of the proposed model lies in its balance of high accuracy and low computational cost, making it an ideal solution for mobile devices and real-time video applications.

In future, different transfer learning architectures, including MobileNetV2 architecture, and various hyperparameters need to be explored in depth for improving emotion classification or identification performances.

ACKNOWLEDGEMENT

The authors express deep gratitude to the Department of Computer Science and Engineering, School of Science and Technology, Metropolitan University, Sylhet, Bangladesh, for their invaluable support in conducting this research.

DECLARATIONS

The authors declare that they have no known competing financial interests or personal relationships which have influenced the work reported in this manuscript.

REFERENCES

1. C. Shan, S. Gong, and P. W. Mcowan, "Robust facial expression recognition using local binary patterns," 2005. [Online]. Available: <http://www.kyb.tuebingen.mpg.de>
2. Y. Huang, F. Chen, S. Lv, and X. Wang, "Facial expression recognition: A survey," *Symmetry (Basel)*, vol. 11, no. 10, Oct. 2019, doi: 10.3390/sym11101189.
3. D. Shah, K. Chavan, S. Shah, and P. Kanani, "Real-Time Facial Emotion Recognition," in *2nd Global Conference for Advancement in Technology, GCAT 2021*, Institute of Electrical and Electronics Engineers Inc., Oct. 2021. doi: 10.1109/GCAT52182.2021.9587707.
4. E. Pranav, S. Kamal, C. S. Chandran, and M. H. Supriya, "Facial Emotion Recognition Using Deep Convolutional Neural Network," in *6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, IEEE, Apr. 2020.
5. Y. Khairuddin and Z. Chen, "Facial Emotion Recognition: State of the Art Performance on FER2013," 2021.
6. S. Aneja, N. Aneja, B. Bhargava, and R. R. Chowdhury, "Device fingerprinting using deep convolutional neural networks'," 2022.
7. I. Atanasov and D. Pilev, "CYBER-PHYSICAL SECURITY THROUGH FACIAL RECOGNITION AND SENSOR DATA ANALYSIS," *Journal of Chemical Technology and Metallurgy*, vol. 59, no. 2, pp. 465–472, 2024, doi: 10.59957/jctm.v59.i2.2024.27.
8. D. Mehta, M. F. H. Siddiqui, and A. Y. Javaid, "Facial emotion recognition: A survey and real-world user experiences in mixed reality," *Sensors (Switzerland)*, vol. 18, no. 2, Feb. 2018, doi: 10.3390/s18020416.
9. A. Goni, Md. U. F. Jahangir, and R. R. Chowdhury, "A Study on Cyber security: Analyzing Current Threats, Navigating Complexities, and Implementing Prevention Strategies," *International Journal of Research and Scientific Innovation*, vol. X, no. XII, pp. 507–522, 2024, doi: 10.51244/ijrsi.2023.1012039.
10. M. Albert and R. James A, "An approach to environmental psychology.," The MIT Press. Accessed: Nov. 30, 2023. [Online]. Available: <https://psycnet.apa.org/record/1974-22049-000>
11. J. Chen, Z. Chen, Z. Chi, and H. Fu, "Facial Expression Recognition Based on Facial Components Detection and HOG Features," 2014.
12. M. Drashti, H. Bhatt, M. Kirit, R. Rathod, M. Shardul, and J. Agravat, "A Study of Local Binary Pattern Method for Facial Expression Detection," *International Journal of Computer Trends and Technology*, vol. 7, no. 3, 2014, [Online]. Available: www.internationaljournalsrsg.org
13. S. M. Hassan, A. K. Maji, M. Jasiński, Z. Leonowicz, and E. Jasińska, "Identification of plant-leaf diseases using cnn and transfer-learning approach," *Electronics (Switzerland)*, vol. 10, no. 12, Jun. 2021, doi: 10.3390/electronics10121388.
14. Y. Gulzar, "Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique," *Sustainability (Switzerland)*, vol. 15, no. 3, Feb. 2023, doi: 10.3390/su15031906.
15. M. Akay *et al.*, "Deep Learning Classification of Systemic Sclerosis Skin Using the MobileNetV2 Model," *IEEE Open J Eng Med Biol*, vol. 2, pp. 104–110, 2021, doi: 10.1109/OJEMB.2021.3066097.
16. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," 2012. [Online]. Available: <http://code.google.com/p/cuda-convnet/>
17. I. M. Revina and W. R. S. Emmanuel, "A Survey on Human Face Expression Recognition Techniques," *Journal of King Saud University – Computer and Information Sciences*, vol. 33, no. 6.

- King Saud bin Abdulaziz University, pp. 619–628, Jul. 01, 2021. doi: 10.1016/j.jksuci.2018.09.002.
18. A. Sinha and A. RP, “Real Time Facial Emotion Recognition using Deep Learning,” *International Journal of Innovations & Implementations in Engineering*, Dec. 2019, [Online]. Available: <https://www.researchgate.net/publication/339998686>
 19. S. Dwijayanti, M. Iqbal, and B. Y. Suprpto, “Real-time Implementation of Face Recognition and Emotion Recognition in a Humanoid Robot Using a Convolutional Neural Network,” *IEEE Access*, 2022, doi: 10.1109/ACCESS.2022.3200762.
 20. G. Sharma, “CK+48 5 emotions,” Kaggle. Accessed: Nov. 30, 2023. [Online]. Available: <https://www.kaggle.com/datasets/gauravsharma99/ck48-5-emotions>
 21. Goodfellow et al., “FER2013 Dataset.” Accessed: Nov. 22, 2023. [Online]. Available: <https://paperswithcode.com/dataset/fer2013>.
 22. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.
 - R. R. Chowdhury, A. C. Idris, and P. E. Abas, “A Deep Learning Approach for Classifying Network Connected IoT Devices Using Communication Traffic Characteristics,” *Journal of Network and Systems Management*, vol. 31, no. 1, p. 26, Mar. 2023, doi: 10.1007/s10922-022-09716-x.
 23. A. Tragoudaras et al., “Design Space Exploration of a Sparse MobileNetV2 Using High-Level Synthesis and Sparse Matrix Techniques on FPGAs,” *Sensors*, vol. 22, no. 12, Jun. 2022, doi: 10.3390/s22124318.
 24. K. Weiss, T. M. Khoshgoftaar, and D. Wang, *A survey of transfer learning*. Springer International Publishing, 2016. doi: 10.1186/s40537-016-0043-6.
 25. R. R. Chowdhury, A. C. Idris, and P. E. Abas, “Internet of Things Device Classification using Transport and Network Layers Communication Traffic Traces,” *International Journal of Computing and Digital Systems*, vol. 12, no. 1, pp. 545–555, 2022, doi: 10.12785/ijcds/120144.
 26. R. R. Chowdhury, A. C. Idris, and P. E. Abas, “Internet of things: Digital footprints carry a device identity,” 2023, p. 40003. doi: 10.1063/5.0111335.
 27. R. R. Chowdhury, A. C. Idris, and P. E. Abas, “Device identification using optimized digital footprints,” *IAES International Journal of Artificial Intelligence*, vol. 12, no. 1, pp. 232–240, Mar. 2023, doi: 10.11591/ijai.v12.i1.pp232-240.
 28. R. R. Chowdhury, A. C. Idris, and P. E. Abas, “Identifying SH-IoT devices from network traffic characteristics using random forest classifier,” *Wireless Networks*, 2023, doi: 10.1007/s11276-023-03478-3.
 29. R. R. Chowdhury and P. E. Abas, “A survey on device fingerprinting approach for resource-constraint IoT devices: Comparative study and research challenges,” *Internet of Things (Netherlands)*, vol. 20. Elsevier B.V., Nov. 01, 2022. doi: 10.1016/j.iot.2022.100632.
 30. L. Cuimei, Q. Zhiliang, J. Nan, and W. Jianhua, “Human face detection algorithm via Haar cascade classifier combined with three additional classifiers,” in *13th International Conference on Electronic Measurement & Instruments*, 2017, pp. 483–487.