AI-generated Dataset using Generative Model and Evaluation using Pre-trained CNN Model

Jia-Jun Ho1, a), Wee-How Khoh1, 2, b), Ying-Han Pang1, 2, c) and Hui-Yen Yap1, 2, d)

1Faculty of Information Science & Technology (FIST), Multimedia University, Jalan Ayer Keroh Lama, Bukit Beruang, 75450, Melaka, Malaysia.

2Centre for Advanced Analytics, CoE for Artificial Intelligence, Multimedia University, Jalan Ayer Keroh Lama, Bukit Beruang, 75450, Melaka, Malaysia.

*b) Corresponding author: whkhoh@mmu.edu.my   
a) jiajunho.926@gmail.com*

*c) yhpang@mmu.edu.my*

*d) hyyap@mmu.edu.my*

**Abstract.** With applications in security, psychology, automation and robotic, and human-computer interaction, facial expression recognition (FER) has become an essential tool for non-verbal communication. Current research often categorizes expressions into micro and macro types, yet existing datasets suffer from limitations such as inconsistent labelling for classes, limited diversity of the databases, and insufficient scale for the currently available datasets. To address these gaps, this work proposes a novel framework combining the Diffusion model with pre-trained convolutional neural networks (CNNs). Leveraging original images from established datasets, including The Chinese Academy of Sciences Micro-expression Ⅱ (CASME Ⅱ), we generate synthetic facial expressions to augment training data, mitigating bias and inconsistency. The synthetic dataset is evaluated using pre-trained VGG16 and ResNet 50 architectures, with performance metrics of accuracy, precision, sensitivity and F1-score compared against some baseline models trained on original data. Additionally, the classification performance for every class is represented in graphical form using a confusion matrix. The novel framework proposed combination in this work can overcome the stated limitations in FER. This approach can generate new facial expression images and create a new database. The newly created facial expression database can be shared with others, contributing to appreciating research value in this field. Although the proposed approach is novel, there are still many possibilities to explore facial expression recognition and approaches to solve the recent research gaps pointed out by the researchers. This work utilizes a novel combination of generative model and pre-trained CNN model to overcome the limits of the current dataset while providing a new database for FER. Our proposed approach has achieved state-of-the-art performance and outperformed some existing state-of-the-art approaches benchmarked in this study.

# Introduction

Facial expression is a kind of non-verbal communication for human that helps humans to express their feelings or messages in a more useful and effective way. There are two categories of facial expressions, which are micro expression and macro expression. Micro expression is a represents the actual feelings of an individual via facial muscle movements. In contrast, macro expression can represent an individual genuine feeling and sometimes can be an acted emotion. Both micro and macro expressions are to help to understand a person's current feelings easily. The common facial expressions are smile, sad, fear, angry, surprise, disgust, confused and neutral. Facial expression recognition (FER) has become more common nowadays and can be found in various fields such as psychological counselling, automotive safety features, stress detection etc. Although FER is commonly used in various areas, but there are still many unknown areas that attract the researchers to explore. Even though FER is commonly practiced by researchers, it is still facing some critical issues which affect the effectiveness and efficiency of FER. For example, the current existing facial expression datasets that frequently utilized to conduce the studies on FER is suffered from several limitations such as inconsistent labelling for classes, limited diversity of the databases, and insufficient scale for the currently available datasets.

In order to reduce the limitations of facial expressions database, generative models have been proposed in some studies. Generative model is a machine learning model that has the ability to produce new data which is similar to the data it was trained on. Generative models focus on understanding the fundamental distribution of input data, in contrast to discriminative models, which emphasise distinguishing between categories or predicting labels. Once the generative model is trained, it can generate completely new, synthetic data that share similar features to the original dataset. There are many types of generative models including Variational Autoencoder (VAE), Autoregressive model, Generative Adversarial Network (GAN), Diffusion model others. Generative models were deployed in this work to overcome the limitations faced by the current existing public facial expression databases. To elevate the performance of FER, pre-trained CNN models are proposed in several most recent works. Pre-trained CNN models are deep learning models which have been trained using ImageNet. One of the strengths of using pre-trained CNN model is it eliminates the need of train from scratch, which is time consuming and cost large amount of resources. Various of popular pre-trained CNN models that can be adopted in this work including VGG16, ResNet 50, AlexNet, SqueezeNet, GoogleNet and others. The idea of this study is to introduce a novel approach for facial expression recognition by combining the generative model and pre-trained CNN model. In short, a generative model is used to generate a new dataset, and the new dataset is used to train and test the pre-trained CNN models. The obtained results are used to determine the performance of the generated dataset.

# Literature Reviews

CNN model is popular and high-performance approach; it is used in most of the studies [1][2][3][4][5][6][7][8][9][10][11][12][13][14][15][16] as the classification models. CNN is one of the best algorithms to solve facial expressions related studies and able produce state-of-the-art performance. Nasri et al. [1] adopted Xception CNN paired with a K-fold cross-validation technique and trained and tested on Empathic, AffectNet and CK+ databases. Lasri et al. [2] proposed a CNN model to identify students' moods based on their facial expressions Furthermore, Pranav et al. [3] employed a 2-dimensional (2D) CNN to identify the facial emotion and evaluated the model by using a self-collected facial emotion database. To enhance the performance of CNN model, a Venturi Architecture for CNN was proposed by Verma et al. [4]. Moreover, Luo et al. [5] made comparisons between the proposed CNN model and the FGD total gradient descent algorithm, SGD random gradient descent technique, and MGD small batch gradient descent algorithm. A feature redundancy-reduced convolutional neural network (FRR-CNN) has been proposed by Xie and Hu [7] to recognize facial expressions and experimented using CK+ and JAFFE. A broad learning based convolutional neural network (CNNBL) was employed by Chen et al. [8] to accelerate the model's learning process. Lastly, Zhang and Shen [17] employed a residual network to recognize micro-expression.

Furthermore, generative model is a relatively new and unique approach in the facial expression recognition area. It is normally used to create fake images by using an original image input. This approach can increase the number of images and overcome the problem or insufficient dataset. In the following study, Wang et al. [18] developed an enhanced StarGAN for face emotion recognition using LAUN. LAUN improved StarGAN are used in this experiment in order to generate a series of higher quality fake facial emotions images for every emotion and VGG16 is the classification model used in this study. Moreover, generative model has also been used in Fan et al. [19] work. A generative adversarial network is employed in an unsupervised domain adaption for the recognition of face emotions. A cross-database method is proposed by using multiple databases and conditional generative adversarial network (CGAN) is the model employed. An unsupervised learning micro-expression generative adversarial network (ULME-GAN) was proposed by [20] to generate micro-expression sequences. AU-matrix re-encoding (AUMR) was used to improve accuracy and employed transfer learning approach to train their proposed generator network. [21] introduced a multi-sequence based micro-expression (ME) generation approach for ME recognition. An improved facial alignment approach by applying heuristic key points to align faces. SAMM, SMIC, and CASME II were the datasets used to conduce the experiments in this work.

# Methodology

In this research, it aims to propose a study on AI-generated Dataset using Generative Model and Evaluation using Pre-trained CNN Model. It is about using a generative AI model to create an AI-generated dataset and use a pre-trained model to assess the AI-generated dataset. CASME Ⅱ is the dataset utilized in this work. This work was divided into two parts by using two datasets for the training of pre-trained CNN models: using original CASME Ⅱ to train the pre-trained CNN model and using AI-generated dataset to train the model. For the first experiment, CASME Ⅱ dataset is directly fit into the pre-trained CNN model for training and testing. This method is to act as a baseline model in order to compare with the other model. For the second experiment, CASME Ⅱ dataset is input into a generative model to generate a series of AI generated images. The generative model utilized in this work is Diffusion model. Diffusion model is a popular generative model used for image generation. It basically slowly adds noise into an image and try to denoise and reconstruct the image. Then, the generated images are pre-processed by resizing the image, rescaling and turning into grayscale. The preprocessing approach is to enhance the quality of the facial expression image to emphasize the key features such as mouth, nose and eyes and standardize the image size since some images have different sizes. This approach provides the pre-trained CNN model has a proper learning criteria and prevent overfitting issue in order to obtain better performance. The pre-processed images are used to train a pre-trained CNN model. ResNet 50 and VGG-16 are the pre-trained CNN models proposed in this work. VGG-16 is deep convolutional neural network architecture which is popular and commonly used in image classification task. It made up of 3 fully connected and 13 convolutional layers, and a SoftMax layer for classification. Moreover, ResNet 50 is made up of 50 layers of convolution layers. It uses its residual connections to learn the features from the input and improve its learning process. VGG-16 and ResNet 50 carried out the classification task to distinct the seven different facial expressions including happiness, sadness, disgust, repression, fear, surprise and others. Lastly, the performance for each model was obtained and compared to each other. Overview of the methodology is shown in Figure 1.

A diagram of a diagram

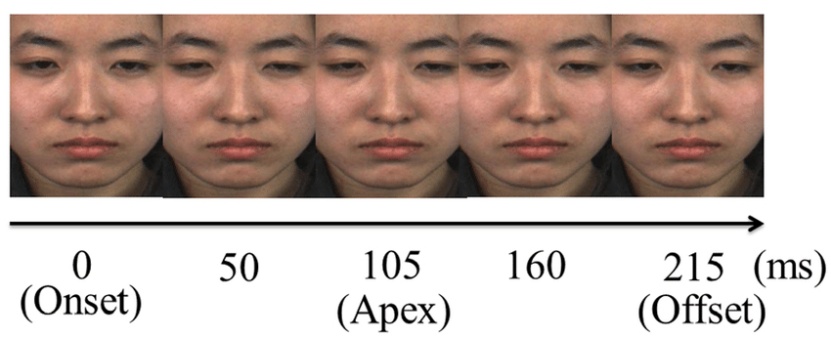
AI-generated content may be incorrect.

**FIGURE 1.** Methodology overview

# Results Analysis and discussion

## Dataset

The dataset used to conduct the experiment in this study is CASME Ⅱ (refer to Figure 2). CASME Ⅱ was first presented by Yan et al. [22]. It is an enhanced form of the database CASME, which was developed by [23]. The Chinese Academy of Sciences Micro-expression (CASME) Databases is an improved data source for spontaneous micro-expression. A controlled setting with the right tools and illumination is used to capture the dataset. At a high resolution of 280×340 pixels, it is captured in video format at 200 frames per second. Six distinct expressions are used to designate the micro-expression, following the guidelines provided by Yan et al. [22].



**FIGURE 2.** Sample images of CASME Ⅱ

## Experimental Setup and Settings

The experiments were carried out on the Intel Core i7-11800H CPU running at 2.30GHz, with 16GB of RAM and the NVIDIA GEFORCE RTX 3060 GPU with 6 GB of dedicated graphic memory. Python is the programming language used and its working environment is JUPYTER Notebook. The images were resized into 160×160 pixels for the VGG-16 before feeding into the model due to computer resource limitations. Then, the datasets were divided into 60% for training, 20% for validation and 20% for testing.

## Experimental Analysis and Discussions

This experiment was conducted using two different datasets: pre-trained CNN model were training and testing using original dataset requested from the owner and pre-trained CNN model were training and testing by using AI generated dataset. For the evaluation of every dataset’s performance, pre-trained VGG-16 and pre-trained ResNet 50 are utilized for the testing of the dataset’s performance. The datasets ate described as follows:

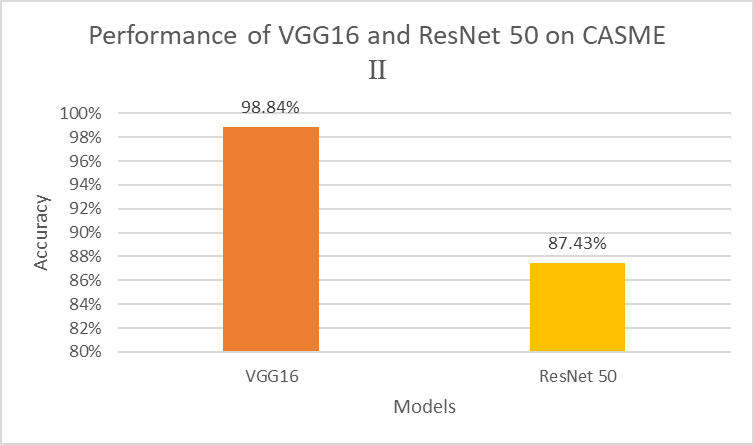
* **Dataset 1**: VGG-16 and ResNet 50 were training on the original CASME Ⅱ obtained from the owner of this dataset.
* **Dataset 2**: VGG-16 and ResNet 50 were training on the AI-generated CASME Ⅱ dataset. AI generated version of CASME Ⅱ is generated a generative model called Diffusion model. This model can generate large number of new images.

CASME Ⅱ is the dataset used for both settings of experiment. The first experiment is VGG-16 and ResNet 50 trained on the original CASME Ⅱ dataset. Moreover, the second experiment used proposed AI-generated dataset to train VGG-16 and ResNet 50.

The results obtained by VGG-16 and ResNet 50 while trained on CASME Ⅱ are represented in Table 1. For VGG-16, it successfully achieves an outstanding performance with 98.84% accuracy. Meanwhile, ResNet 50 has also reached a state-of-the-art performance with 87.43% accuracy. VGG-16 has significantly outperformed ResNet 50 by 11.14% accuracy. The performance of VGG-16 and ResNet 50 on CASME Ⅱ are displayed in Figure 3.

**TABLE 1.** Accuracy obtained for VGG-16 and ResNet 50 while trained on CASME Ⅱ

|  |  |
| --- | --- |
| **Models** | **Accuracy** |
| VGG-16 | 98.84% |
| ResNet 50 | 87.43% |

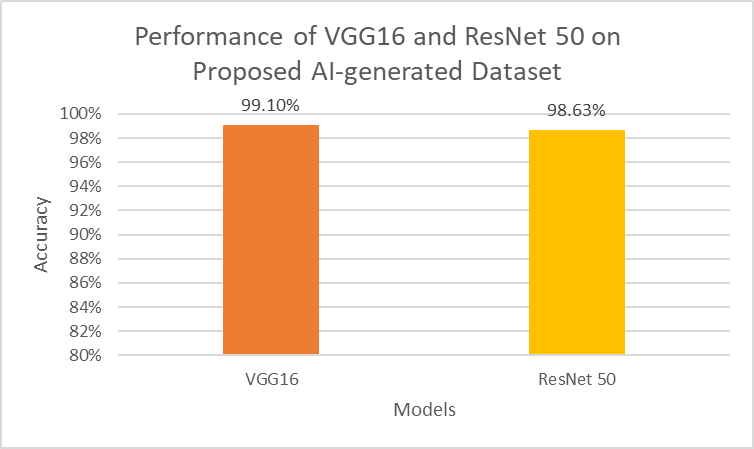


**FIGURE 3**. Performance of VGG16 and ResNet 50 on CASME Ⅱ

The results obtained by VGG-16 and ResNet 50 while trained using our proposed AI-generated dataset are represented in Table 2. VGG-16 has also obtained the best result while trained using our proposed AI-generated dataset and outperformed ResNet 50 with 99.10% accuracy. On the other hand, ResNet 50 also reached a state-of-the-art performance with 98.63% accuracy. ResNet 50 is outperformed by VGG-16 with a minimal difference of 0.47% accuracy. Figure 4 shows the performance of VGG-16 and ResNet 50 on our proposed AI-generated dataset.

**TABLE 2.** Accuracy obtained for VGG-16 and ResNet 50 while training on proposed AI-generated dataset

|  |  |
| --- | --- |
| **Models** | **Accuracy** |
| VGG-16 | 99.10% |
| ResNet 50 | 98.63% |



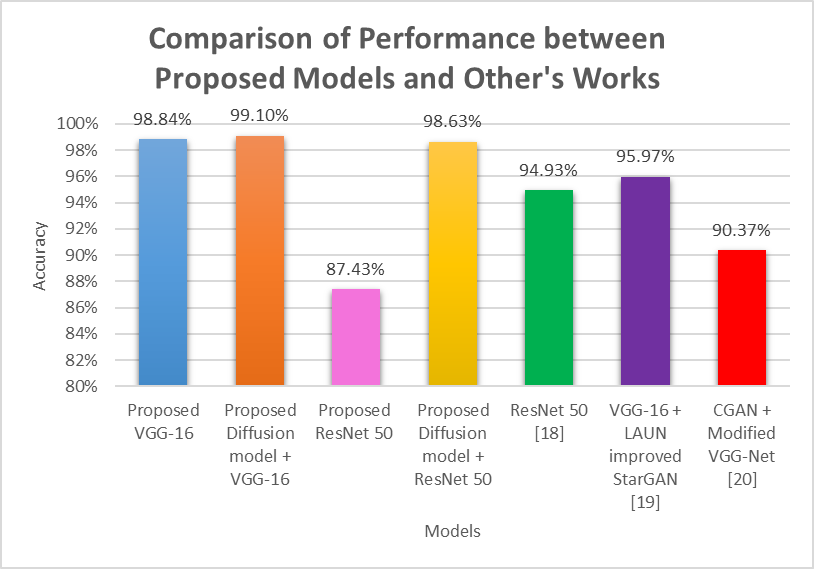
**FIGURE 4**. Performance of VGG-16 and ResNet 50 on the proposed AI-generated dataset

The comparison of our proposed architectures with other works were made in this study. The results are tabulate in Table 3 and illustrate in Figure 5. ResNet 50 proposed by [18] was trained on CASME Ⅱ, VGG-16 combined with LAUN improved StarGAN by [19] was trained using Karolinska Directed Emotional Faces (KDEF) and CGAN with Modified VGG-Net proposed by [20] was trained using combined datasets of Oulu-CASIA and CK+. As shown in the table, our proposed architectures were compared against other’s work that employed the similar architectures including ResNet 50 [18] (94.93%), VGG-16 combined with LAUN improved StarGAN [19] (95.97%) and CGAN with Modified VGG-Net [20] (90.37%). Our proposed VGG-16 trained using our proposed AI-generated dataset achieved outstanding performance as compared to other’s work with an accuracy of 99.10%. Then, it followed by our proposed VGG-16 trained on CASME Ⅱ with 98.84% accuracy. Moreover, our proposed ResNet 50 trained on our proposed AI-generated dataset achieved 98.63%, it performed slightly poorer than VGG-16 trained on our proposed AI-generated dataset and VGG-16 trained on CASME Ⅱ. ResNet 50 [18] achieved 94.93%, VGG-16 combined with LAUN improved StarGAN [19] achieved 95.97% and CGAN with Modified VGG-Net [20] achieved 90.37%. Most of our proposed architectures managed to surpass other’s proposed works except our proposed ResNet 50 trained on CASME Ⅱ. Our proposed ResNet 50 trained on CASME Ⅱ has the lowest accuracy of 87.43%.

**TABLE 3.** Comparison of proposed architecture against other’s works

|  |  |  |
| --- | --- | --- |
| Model | Dataset | Accuracy |
| Proposed VGG-16 | CASME Ⅱ | 98.84% |
| Proposed Diffusion model + VGG-16 | Proposed AI-generated Dataset | 99.10% |
| Proposed ResNet 50 | CASME Ⅱ | 87.43% |
| Proposed Diffusion model + ResNet 50 | Proposed AI-generated Dataset | 98.63% |
| ResNet 50 [18] | CASME Ⅱ | 94.93% |
| VGG-16 + LAUN improved StarGAN [19] | KDEF | 95.97% |
| CGAN + Modified VGG-Net [20] | Oulu-CASIA + CK+ | 90.37% |

Our proposed VGG-16 trained using our proposed AI-generated dataset has the best performance among other models including our proposed models. It shows a promising performance and has potential of applying it in more modern and advanced applications including, real time facial expressions recognition, autonomous driving, psychotherapy etc. Meanwhile, our proposed ResNet 50 trained on CASME Ⅱ has the lowest accuracy. In order to improve its training accuracy, fine-tuning and ablation work such as freezing some convolution layers can be applied. Although our proposed works show a promising performance, there are still many potentials and possibilities to be explored including applying different types of generative model for the image generation, applying different types of deep learning model for the dataset evaluation and applying ablation works to the models.



**FIGURE 5.** Comparison of proposed architectures with other’s works

# CONCLUSION

In this paper, we presented AI Creation of Facial Expression Databases for Advanced Emotion Recognition by using generative model and pre-trained CNN model. Generative model had been utilized to generate a new dataset based on an existing dataset and pre-trained CNN models were used to determine the performance of the dataset. This research included two different settings of experiment: using an original to train the pre-trained CNN models and using AI-generated dataset to train the pre-trained CNN models. The existing dataset used to carry out the experiments is CASME Ⅱ. In conclusion, all the proposed architectures managed to achieve state-of-the-art performance in both experiments. In this study, Diffusion model had been applied to generate a new dataset that contains images with seven different facial expressions based on CASME Ⅱ. The new dataset was then utilized to train VGG-16 an ResNet 50. The VGG-16 and ResNet 50 trained using original CASME Ⅱ was used as a baseline model in order to compare with VGG-16 and ResNet 50 trained on the proposed AI-generated dataset. The obtained performance of the proposed models was compared to each other. VGG-16 trained using proposed AI-generated dataset obtained the best performance with 99.10% accuracy. VGG-16 trained using proposed AI-generated dataset has outperformed the other proposed models in this work. All of our proposed approaches managed to achieve state-of-the-art performances. Although this study presented promising performance but there are still available many potentials and possibilities to be applied in the future work.

# References

1. M.A. Nasri, M.A. Hmani, A. Mtibaa, D. Petrovska-Delacretaz, M.B. Slima, and A.B. Hamida, “Face Emotion Recognition From Static Image Based on Convolution Neural Networks,” in *2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, (IEEE, Sousse, Tunisia, 2020), pp. 1–6.
2. I. Lasri, A.R. Solh, and M.E. Belkacemi, “Facial Emotion Recognition of Students using Convolutional Neural Network,” in *2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS)*, (IEEE, Marrakech, Morocco, 2019), pp. 1–6.
3. E. Pranav, S. Kamal, C. Satheesh Chandran, and M.H. Supriya, “Facial Emotion Recognition Using Deep Convolutional Neural Network,” in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, (IEEE, Coimbatore, India, 2020), pp. 317–320.
4. A. Verma, P. Singh, and J.S. Rani Alex, “Modified Convolutional Neural Network Architecture Analysis for Facial Emotion Recognition,” in *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)*, (IEEE, Osijek, Croatia, 2019), pp. 169–173.
5. Y. Luo, J. Wu, Z. Zhang, H. Zhao, and Z. Shu, “Design of Facial Expression Recognition Algorithm Based on CNN Model,” in *2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA)*, (IEEE, Shenyang, China, 2023), pp. 580–583.
6. N. Naik, and M.A. Mehta, “An Improved Method to Recognize Hand-over-Face Gesture based Facial Emotion using Convolutional Neural Network,” in *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, (IEEE, Bangalore, India, 2020), pp. 1–6.
7. S. Xie, and H. Hu, “Facial expression recognition with FRR‐CNN,” Electronics Letters **53**(4), 235–237 (2017).
8. L. Chen, M. Li, X. Lai, K. Hirota, and W. Pedrycz, “CNN-based Broad Learning with Efficient Incremental Reconstruction Model for Facial Emotion Recognition,” IFAC-PapersOnLine **53**(2), 10236–10241 (2020).
9. J. Shao, and Y. Qian, “Three convolutional neural network models for facial expression recognition in the wild,” Neurocomputing **355**, 82–92 (2019).
10. R.K. Madupu, C. Kothapalli, V. Yarra, S. Harika, and C.Z. Basha, “Automatic Human Emotion Recognition System using Facial Expressions with Convolution Neural Network,” in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, (IEEE, Coimbatore, India, 2020), pp. 1179–1183.
11. A. John, A. Mc, A.S. Ajayan, S. Sanoop, and V.R. Kumar, “Real-Time Facial Emotion Recognition System With Improved Preprocessing and Feature Extraction,” in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, (IEEE, Tirunelveli, India, 2020), pp. 1328–1333.
12. M. Mohammadpour, H. Khaliliardali, S.Mohammad.R. Hashemi, and Mohammad.M. AlyanNezhadi, “Facial emotion recognition using deep convolutional networks,” in *2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, (IEEE, Tehran, 2017), pp. 17–21.
13. M. Mukhopadhyay, A. Dey, R.N. Shaw, and A. Ghosh, “Facial emotion recognition based on Textural pattern and Convolutional Neural Network,” in 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), (IEEE, Kuala Lumpur, Malaysia, 2021), pp. 1–6.
14. M.-I. Georgescu, R.T. Ionescu, and M. Popescu, “Local Learning With Deep and Handcrafted Features for Facial Expression Recognition,” IEEE Access **7**, 64827–64836 (2019).
15. R. Zhi, H. Xu, M. Wan, and T. Li, “Combining 3D Convolutional Neural Networks with Transfer Learning by Supervised Pre-Training for Facial Micro-Expression Recognition,” IEICE Trans. Inf. & Syst. **E102.D**(5), 1054–1064 (2019).
16. Y. Lim, K.-W. Ng, P. Naveen, and S.-C. Haw, “Emotion Recognition by Facial Expression and Voice: Review and Analysis,” JIWE **1**(2), 45–54 (2022).
17. Z. Haiwei, and S. Zhuofan, “Micro-expression Recognition Based on Residual Network,” in *2023 IEEE 5th International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, (IEEE, Dali, China, 2023), pp. 772–775.
18. X. Wang, J. Gong, M. Hu, Y. Gu, and F. Ren, “LAUN Improved StarGAN for Facial Emotion Recognition,” IEEE Access **8**, 161509–161518 (2020).
19. Y. Fan, J.C.K. Lam, and V.O.K. Li, “Unsupervised Domain Adaptation with Generative Adversarial Networks for Facial Emotion Recognition,” in *2018 IEEE International Conference on Big Data (Big Data)*, (IEEE, Seattle, WA, USA, 2018), pp. 4460–4464.
20. J. Zhou, S. Sun, H. Xia, X. Liu, H. Wang, and T. Chen, “ULME-GAN: a generative adversarial network for micro-expression sequence generation,” Appl Intell **54**(1), 490–502 (2024).
21. Y. Chen, C. Zhong, P. Huang, W. Cai, and L. Wang, “Improving Micro-expression Recognition using Multi-sequence Driven Face Generation,” in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (IEEE, Hyderabad, India, 2025), pp. 1–5.
22. W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, and X. Fu, “CASME II: An Improved Spontaneous Micro-Expression Database and the Baseline Evaluation,” PLoS ONE **9**(1), e86041 (2014).
23. W.-J. Yan, Q. Wu, Y.-J. Liu, S.-J. Wang, and X. Fu, “CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces,” in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, (IEEE, Shanghai, China, 2013), pp. 1–7.