**Edelweiss Applied Science and Technology**
*ISSN: 2576-8484*
*Vol. 9, No. 3, 1855-1869*
*2025*
*Publisher: Learning Gate*
*DOI: 10.55214/25768484.v9i3.5713*


# Evaluating the impact of AI-generated synthetic noisy data on human emotion recognition models

Dastan Sultanov[1], Sanzhar Seitbekov[2], Shynara Ayanbek[3], Miras Assubay[4], Zhaisang Sarsengaliyev[5], Amandyk Kartbayev[6*]
[1,2,3,4,5,6]School of IT and Engineering, Kazakh-British Technical University, Almaty, Kazakhstan; a.kartbayev@gmail.com (A.K.)

**Abstract:** Human emotion recognition remains a challenging task due to the complexity and variability of human emotions in real-world scenarios. This study investigates the impact of AI-generated synthetic data on enhancing Facial Expression Recognition (FER) model performance. Using the Juggernaut XL model, we generated 300 synthetic images per emotion category from the FER2013 dataset and incorporated them into the training process of a VGG-19-based FER model. Experimental results revealed that the synthetic data did not improve key performance metrics, with the originally trained model achieving an accuracy of 65%, compared to 63% for the augmented dataset. Precision, recall, and F1-score also exhibited fluctuations across different emotion categories, as illustrated by confusion matrices. The findings suggest that the quality of synthetic images plays a crucial role in model effectiveness, as insufficient diversity may introduce noise rather than beneficial augmentation. Factors such as limited training epochs and potential dataset biases may have also influenced the outcomes. This study highlights the importance of optimizing synthetic image realism to improve FER models and offers practical insights for future AI-driven applications using data augmentation.

***Keywords:*** *Data augmentation emotion recognition, Facial expression recognition, Medical images, Prediction models, Synthetic data, System design.*

## 1. Introduction

Imagine a world where entertainment systems adapt to your emotional responses in real-time, adjusting movie recommendations or modifying gameplay difficulty based on facial expressions. Consider a vehicle that monitors driver fatigue and distraction through facial emotion recognition (FER) to enhance road safety by providing timely alerts and interventions. These applications highlight the transformative potential of FER in artificial intelligence, with applications spanning human-computer interaction, healthcare, security, and automotive safety. However, despite its growing importance, FER systems face significant challenges, particularly in ensuring high accuracy and robustness across diverse, real-world conditions.

One of the key barriers to effective deployment is the inherent limitations of existing datasets. Traditional datasets like FER2013, despite being widely used in FER research, suffer from issues such as insufficient diversity in facial expressions, demographic representation biases, and ethical concerns related to privacy [1]. More critically, these datasets contain noisy images—low-resolution, blurred, or occluded facial expressions that degrade model performance. The presence of noise makes it difficult for deep learning models to learn meaningful facial features, leading to misclassifications and reduced accuracy [2, 3]. Addressing these challenges is crucial for improving FER systems' ability to generalize across different environments, lighting conditions, and cultural variations.


**History:** Received: 3 January 2025; Revised: 10 March 2025; Accepted: 12 March 2025; Published: 24 March 2025
* Correspondence: a.kartbayev@gmail.com

To mitigate these issues, researchers have explored various methods, including data augmentation, transfer learning, and the integration of synthetic data [4, 5]. The emergence of AI-generated synthetic images has opened new avenues for enhancing FER datasets, allowing for increased diversity and improved model generalization. Advanced generative models, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), have demonstrated their ability to create high-fidelity images that supplement real datasets without ethical or logistical constraints [6]. These synthetic images offer the potential to address class imbalances and enrich training data, leading to more robust FER models.

The landscape of the research has evolved significantly, with AI applications playing a pivotal role in overcoming dataset limitations. The introduction of GANs revolutionized image synthesis, enabling models to generate realistic images by learning the underlying distribution of real-world facial expressions [7, 8]. Subsequent improvements in generative models have made synthetic images nearly indistinguishable from real ones, leading to promising applications in FER dataset augmentation [9].

Several studies have explored the impact of synthetic data on emotion recognition performance. The research by Ma and Nguyen [10] and He, et al. [11] demonstrated that incorporating GAN-generated images into training datasets could enhance model accuracy and improve recognition of underrepresented emotions. Similarly, Meng, et al. [12] and Kartbayev [13] showed that synthetic data could increase the diversity of training sets, improving generalization to unseen expressions. These findings suggest that AI-generated images can complement traditional datasets, addressing some of their inherent weaknesses.

However, challenges remain in effectively integrating synthetic data into FER models. Despite the success of GANs in generating high-quality images, the realism and variability of synthetic expressions are critical factors influencing their effectiveness [14]. For instance, studies by Kim and Park [15] and Yalçin and Alisawi [16] highlighted that many synthetic datasets fail to capture the subtle variations and context-specific nuances of human facial expressions. Poor-quality synthetic images may introduce noise rather than valuable learning signals, potentially degrading model performance rather than improving it.

Another key challenge is the lack of systematic evaluations of synthetic data integration in FER models. While numerous studies have demonstrated improvements in FER performance using GANs, few have analyzed the specific effects of synthetic images on different emotion categories or explored their impact on key metrics such as precision, recall, and F1-score [17]. A comprehensive evaluation of synthetic data's role in mitigating dataset limitations, especially noisy images, is essential for advancing this research.

This study aims to address these research gaps by systematically investigating the impact of AI-generated synthetic data on FER model performance. Specifically, we seek to answer the following research questions:

1. How does the integration of AI-generated synthetic images affect the accuracy of FER models, particularly in handling noisy data?
2. Do synthetic images improve the recognition of specific emotion categories, or do they introduce additional noise into the dataset?
3. What are the key factors influencing the effectiveness of synthetic data in FER, including realism, diversity, and integration strategies?

Based on prior research and existing challenges, we formulate the following hypotheses:

*$H_1$: The inclusion of AI-generated synthetic images in FER model training will improve classification accuracy, particularly for underrepresented emotion categories.*

*$H_2$: Synthetic images will help reduce the impact of noisy images by providing additional training examples that enhance model generalization.*

*$H_3$: The effectiveness of synthetic data augmentation is highly dependent on the quality and realism of generated images; lower-quality images may introduce noise and reduce model performance.*

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 3: 1855-1869, 2025*
*DOI: 10.55214/25768484.v9i3.5713*

To test these hypotheses, we utilize the Juggernaut XL model to generate 300 synthetic images for each emotion category in the FER2013 dataset [18]. These images are then integrated into the training process of a VGG-19-based model, allowing for a direct comparison of model performance with and without synthetic data augmentation. Key evaluation metrics include accuracy, precision, recall, and F1-score, analyzed across different emotion categories to assess the impact of synthetic images comprehensively.

A critical aspect of our approach is evaluating how synthetic images influence model performance in the presence of noisy data. The FER2013 dataset is known to contain a significant number of low-quality images, including blurred, occluded, and poorly lit expressions. By examining classification results using confusion matrices and classification reports, we aim to determine whether synthetic data helps mitigate these challenges or exacerbates them.

This research contributes to the field of FER in several key ways:

- Evaluation of synthetic data: unlike prior studies that focus solely on generating synthetic images, we systematically assess their impact on FER model performance across multiple metrics.
- Addressing noisy data challenges: by analyzing how synthetic data interacts with noisy images, this study provides insights into the effectiveness of AI-generated augmentation in real-world FER applications.
- Optimizing synthetic image generation: the findings will inform future work on optimizing synthetic image generation and integration strategies, paving the way for more effective FER models.

Despite the growing interest in AI-generated data augmentation, many questions remain regarding its effectiveness and practical implementation. By investigating the potential benefits and limitations of synthetic images in FER, this study seeks to advance the field and contribute to the development of more accurate and robust emotion recognition systems.

## 2. Methods and Methodology

The development of a robust model using AI-generated synthetic data involves a series of methodical steps, including data preprocessing, synthetic image generation, model training, and performance evaluation. This research focuses on the integration of synthetic images generated by the Juggernaut XL model and the training of a VGG-19-based FER model to analyze the impact of synthetic data on overall model performance, particularly in reducing the effects of noisy images. A core objective of this study is to assess whether synthetic images help mitigate issues related to noisy data, a persistent challenge in FER models.

### *2.1. Data Preprocessing*

The FER2013 dataset is the foundation for this study and requires thorough preprocessing to enhance data quality and consistency. The dataset comprises grayscale facial images classified into seven emotion categories: anger, disgust, fear, happiness, sadness, surprise, and neutral. To improve data quality and standardize inputs, we applied the following preprocessing steps. Each image is resized to 48x48 pixels, which is the standard input size for most deep learning-based FER models [19]. Normalization is applied to scale pixel values to the range [0, 1], improving the stability of the learning process and preventing numerical instabilities. Histogram Equalization was applied to improve contrast in low-light images. One of the primary concerns in FER datasets is the presence of noise, which can arise from low-resolution images, occlusions, varying lighting conditions, and blurred facial features. To mitigate this issue, we apply data augmentation techniques such as random rotation, horizontal flipping, and Gaussian Blur Reduction was used to sharpen blurred images.

This ensures that the model learns from a diverse set of expressions while being less sensitive to minor distortions that naturally occur in real-world images. Rotation (±15°), flipping, zooming, and random cropping were applied to artificially expand dataset diversity. Using entropy-based filtering, excessively noisy or unrecognizable images were flagged and excluded. The goal of this step is to

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 3: 1855-1869, 2025*
*DOI: 10.55214/25768484.v9i3.5713*

enhance the model's generalization ability, reducing the likelihood of overfitting and making it more resilient to noisy images. Examples of the FER2013 dataset's inherent variability can be found in Figure 1.

**Figure 1.**
Sample images from FER2013 Dataset.

*2.2. Model Training*

Training the model is performed using the augmented dataset to assess the impact of synthetic data on model performance. We employ the Segmentation VGG-19 model, a deep convolutional neural network known for its strong feature extraction capabilities. The VGG-19 model consists of 16 convolutional layers followed by 3 fully connected layers, utilizing ReLU activations and max-pooling layers to capture hierarchical facial features [20]. Given the complexity of facial expressions, deep architectures like VGG-19 are well-suited for learning fine-grained differences between emotions. The overall architecture of the VGG-19 model used in this study is shown in Figure 2.

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 3: 1855-1869, 2025*
*DOI: 10.55214/25768484.v9i3.5713*

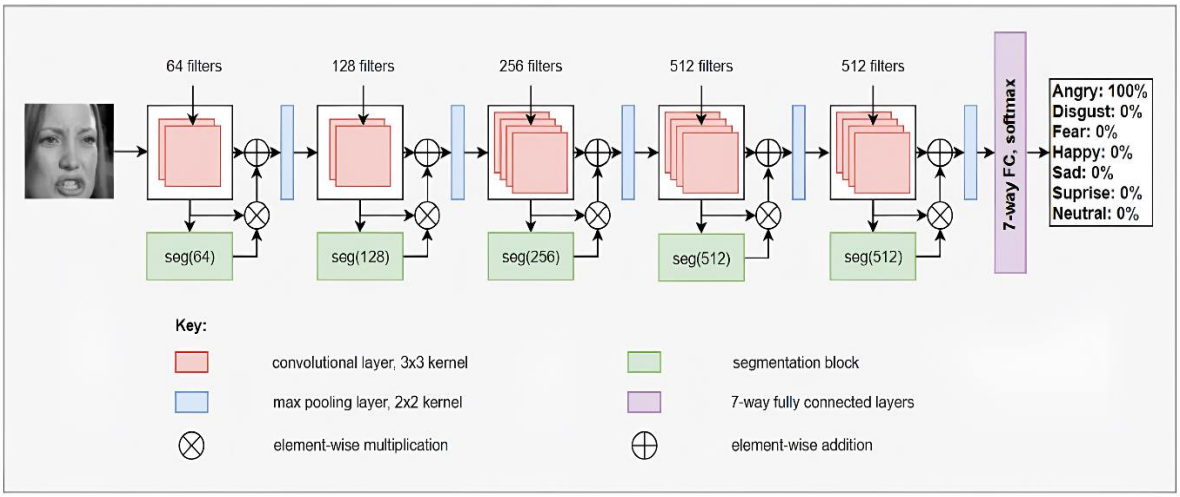


**Figure 2.**
The architecture of the VGG-19 segmentation model.

The model is trained using the Adam optimizer, which is well-suited for handling noisy gradients and adaptive learning rates. The hyperparameters used include a learning rate of 0.0001, a batch size of 8, and a total of 25 epochs. Due to hardware constraints, the training process is limited to this number of epochs, though prior studies have indicated that optimal performance for VGG-19 is achieved at around 67 epochs [21]. The categorical cross-entropy loss function is used to optimize the model, given that FER is a multi-class classification problem. To further address the impact of noisy images on training, additional preprocessing steps such as adaptive histogram equalization (AHE) and Gaussian blurring are explored. These methods help in normalizing image contrast and reducing noise artifacts that could mislead the model.

### *2.3. Performance Evaluation*

The effectiveness of the trained model is evaluated using multiple performance metrics, including accuracy, precision, recall, and F1-score. These metrics provide a comprehensive understanding of the model's ability to correctly classify facial expressions across different emotion categories. Given the known challenges of datasets, particularly those related to noise, additional evaluation methods are incorporated to analyze the impact of synthetic data on model robustness.

Another critical aspect of the evaluation is comparing the model's performance on specific noisy images. Using the classification reports in Tables 3 and 4, we analyze precision and recall values for each emotion class. A confusion matrix is constructed to visualize the distribution of true positives, false positives, and false negatives across emotion categories. This helps in identifying which emotions are most affected by noisy data and whether the addition of synthetic images improves classification performance in these cases. A key focus is on whether synthetic data helps mitigate misclassification in expressions commonly affected by noise, such as "disgust" and "fear," which have historically exhibited lower classification accuracy due to subtle variations in facial features.

By systematically evaluating the model's ability to handle noise, this study aims to determine whether AI-generated synthetic images provide meaningful improvements in accuracy. The findings will contribute to ongoing discussions about the viability of synthetic data in deep learning applications and addresses key limitations in datasets, including noise, class imbalances, and lack of diversity. By leveraging AI-generated images, we seek to enhance the model robustness while providing a structured evaluation framework to assess the real impact of synthetic data augmentation. The overall approach is summarized in Figure 3.

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 3: 1855-1869, 2025*
*DOI: 10.55214/25768484.v9i3.5713*

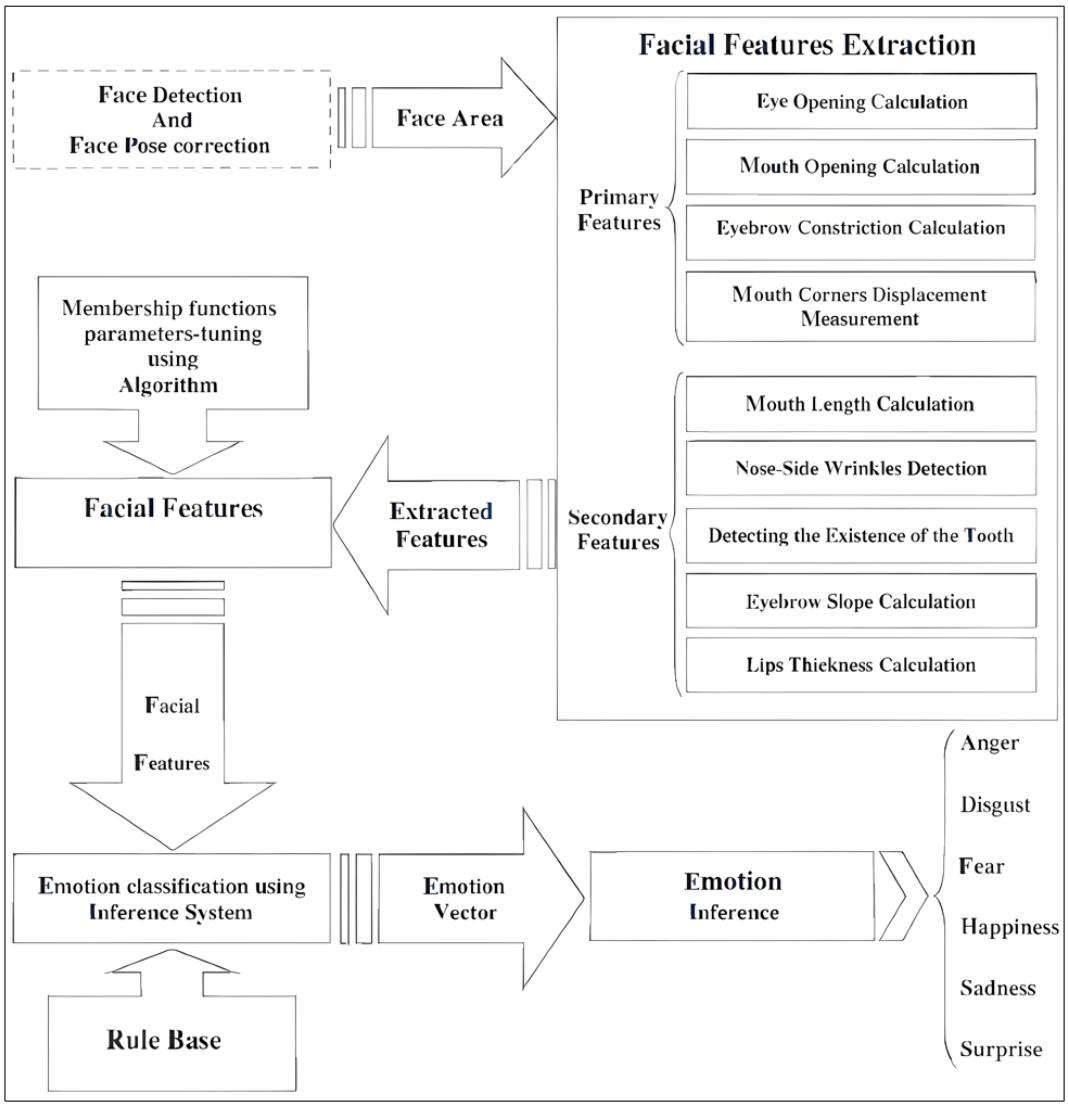


**Figure 3.**
The methodological approach of the study.

## 3. Results

The results of this study reveal a significant impact of AI-generated synthetic data on the performance of a facial emotion recognition model. Given the inherent challenges in real-world datasets, including class imbalances, limited diversity, and noise such as blurriness, occlusions, and varying lighting conditions, synthetic data has been explored as a potential enhancement. The Juggernaut XL model, a state-of-the-art text-to-image generative model, was used to create additional facial images that aimed to supplement the existing FER2013 dataset. Unlike traditional data augmentation techniques, which modify existing images through transformations like rotation and flipping, AI-generated synthetic data introduces entirely new images. This approach offers greater diversity in facial expressions, environmental conditions, and demographic representation, which are crucial for improving model generalization.

To ensure realistic and diverse synthetic images, we constructed detailed text prompts capturing natural variations observed in real-world images. These variations included age, gender, ethnicity,

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 3: 1855-1869, 2025*
*DOI: 10.55214/25768484.v9i3.5713*

lighting conditions, camera angles, obstructions such as hands, hair, or glasses, and intentional blur, all of which can significantly impact the accuracy of FER models. Table 1 illustrates the different variations considered in the prompt design process. An example of a generated prompt is: "An image of a child, Middle Eastern male with a scowling expression and tight lips, captured in shadowy light, from a side angle, covered with hair, with slight blur, expressing anger."

This method ensures that the synthetic images closely resemble real-world conditions while maintaining a high level of diversity and fidelity in facial expressions. In terms of model accuracy, which serves as a fundamental metric for classification performance, the VGG-19 model trained on the original FER2013 dataset achieved an accuracy of 65%, whereas the model trained on the augmented dataset containing synthetic images obtained a slightly lower accuracy of 63%. This suggests that the inclusion of synthetic images did not lead to the anticipated improvement in model performance.

**Table 1.**
Natural variations of textures that occur in the images.

| Age Group | Gender | Ethnicity | Lighting | Angle | Obstructions | Blurriness |
|---|---|---|---|---|---|---|
| Child | Male | Caucasian | Bright | Frontal | Hands | Slight |
| Teenager | Female | African | Low | Side | Hair | Moderate |
| Young Adult | | Asian | Shadowy | Top-down | Glasses | |
| Middle-aged | | Hispanic | | Bottom-up | Mask | |
| Elderly | | Middle Eastern | | | | |

A total of 300 synthetic images were generated for each emotion category, which were then appended to the FER2013 dataset to form an augmented dataset. The composition of the augmented dataset, as shown in Table 2, resulted in a balanced distribution of real and synthetic images across all emotion categories. Sample synthetic images generated by the Juggernaut XL model were analyzed for quality and consistency before being incorporated into the training pipeline. We trained the VGG-19 model on both the original FER2013 dataset and the augmented version, which included synthetic images. The VGG-19 model, a widely used convolutional neural network for image classification, was trained using the Adam optimizer with a learning rate of 0.0001, a batch size of 8, and for 25 epochs.

**Table 2.**
FER2013 Dataset augmented with AI generated images.

| Emotion (class) | Training | Validation | Testing | Class total |
|---|---|---|---|---|
| *angry* | 4231 | 497 | 525 | 5253 |
| *disgust* | 666 | 87 | 94 | 847 |
| *fear* | 4344 | 524 | 553 | 5421 |
| *happy* | 7461 | 922 | 906 | 9289 |
| *sad* | 5065 | 682 | 630 | 6377 |
| *surprise* | 3410 | 450 | 442 | 4302 |
| *neutral* | 5211 | 637 | 650 | 6498 |
| Total | 30388 | 3799 | 3800 | 37987 |

Additionally, precision, recall, and F1-score results exhibited variability across different emotion categories when training on the augmented dataset. The classification reports for both datasets, provided in Tables 3 and 4, highlight these differences in performance. While some emotions benefited slightly from the additional data, others experienced increased misclassification rates, particularly those expressions that require nuanced feature recognition, such as "Disgust" and "Happy."

One of the primary motivations for integrating synthetic data was to assess whether it could reduce the impact of noisy images in the training process. The FER2013 dataset contains a significant number of low-quality images, including blurred expressions, occluded faces, and variations in lighting that make it difficult for machine learning models to extract relevant features. By generating synthetic

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 3: 1855-1869, 2025*
*DOI: 10.55214/25768484.v9i3.5713*

images with controlled variations in lighting, angles, and occlusions, we hypothesized that the model would learn more robust features and perform better in real-world settings [22].

**Table 3.**
The report on classification of the original FER2013 dataset.

| Class | Precision | Recall | F1-score |
|---|---|---|---|
| *Neutral* | 0.58 | 0.55 | 0.56 |
| *Angry* | 0.50 | 0.28 | 0.36 |
| *Disgust* | 0.59 | 0.32 | 0.42 |
| *Fear* | 0.85 | 0.87 | 0.86 |
| *Happy* | 0.50 | 0.63 | 0.56 |
| *Sad* | 0.75 | 0.82 | 0.78 |
| *Surprise* | 0.58 | 0.56 | 0.57 |

**Table 4.**
The report on classification of the augmented FER2013 dataset.

| Class | Precision | Recall | F1-score |
|---|---|---|---|
| *Neutral* | 0.58 | 0.61 | 0.59 |
| *Angry* | 0.59 | 0.48 | 0.53 |
| *Disgust* | 0.40 | 0.55 | 0.47 |
| *Fear* | 0.90 | 0.83 | 0.86 |
| *Happy* | 0.55 | 0.47 | 0.51 |
| *Sad* | 0.74 | 0.80 | 0.77 |
| *Surprise* | 0.57 | 0.58 | 0.58 |

The results suggest that the synthetic data generated by the Juggernaut XL model did not significantly enhance the model performance. One plausible explanation is the quality and realism of the synthetic images. Despite carefully crafted prompts, the synthetic images may not have captured the fine-grained variations in real-world facial expressions. The presence of artifacts, exaggerated features, or inconsistencies in shading and texture could have introduced additional noise rather than meaningful training data.

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 3: 1855-1869, 2025*
*DOI: 10.55214/25768484.v9i3.5713*

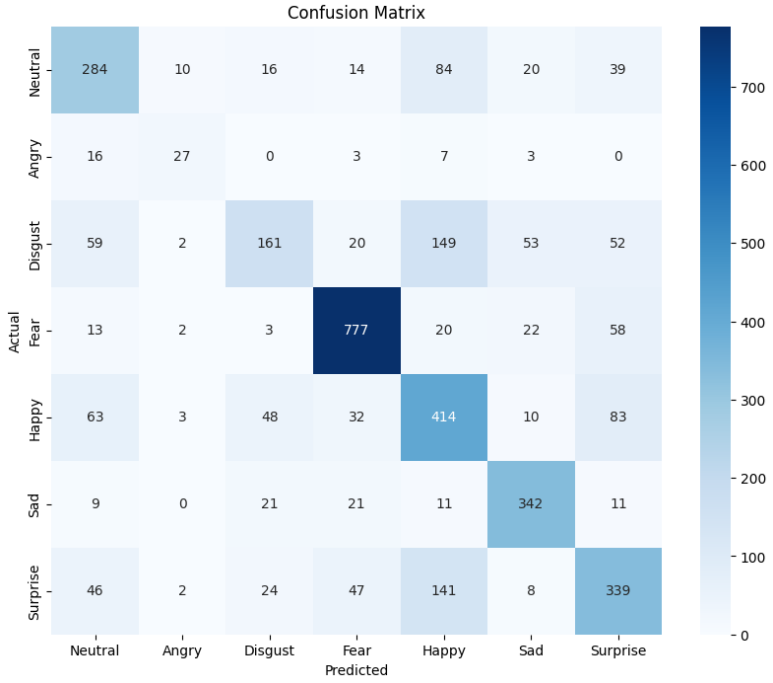


**Figure 4.**
Confusion matrix of VGG-19 model on the original FER-2013 dataset.

To analyze the effect of synthetic data on noise reduction, confusion matrices, as shown in Figures 4 and 5, were generated for both the original and augmented datasets. These matrices provide a detailed breakdown of true positives, false positives, and false negatives for each emotion category. The model trained on the original FER2013 dataset performed well in classifying "Fear" with 777 true positives but struggled with "Disgust," which had a high number of misclassifications. The augmented dataset continued to classify "Fear" effectively, with 763 true positives, but performance variations were observed for other emotions, particularly "Happy" and "Disgust." This indicates that while synthetic images-maintained performance consistency for certain emotion classes, they did not universally improve recognition for complex facial expressions that require subtle feature differentiation. Another contributing factor could be the limited diversity in synthetic data despite controlled variations in demographics and environmental conditions. While synthetic data generation aimed to enhance class balance and representation, it may not have sufficiently accounted for expression dynamics, micro-expressions, and natural distortions that occur in real facial movements. This limitation could explain the marginal decrease in accuracy when incorporating synthetic images [23].

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 3: 1855-1869, 2025*
*DOI: 10.55214/25768484.v9i3.5713*

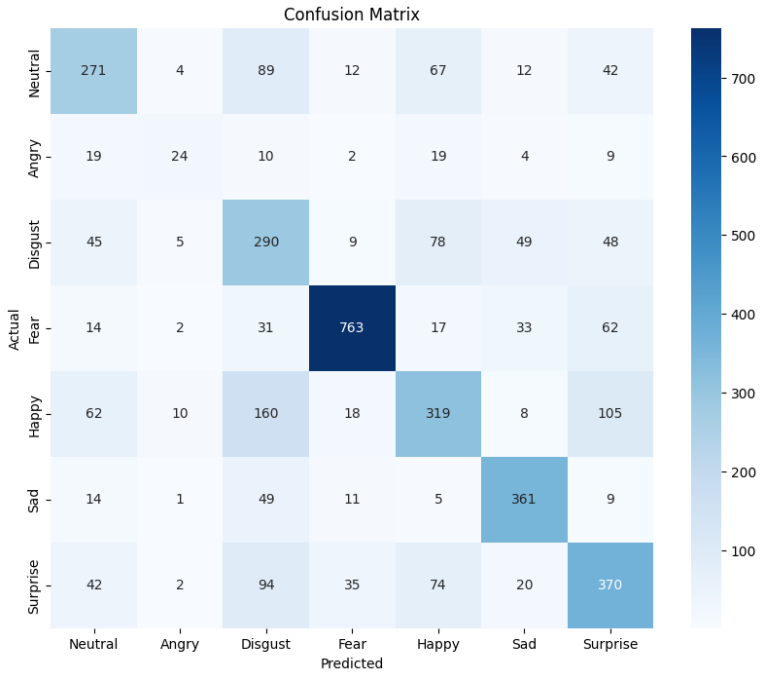


**Figure 5.**
Confusion matrix of VGG-19 model on our AI augmented FER-2013 dataset.

This outcome suggests that the effectiveness of data augmentation using synthetic images is highly dependent on the quality and realism of the generated images. While the synthetic images were carefully designed using prompts to capture diverse demographic and environmental conditions, they may not have fully replicated the subtle variations in real-world facial expressions. Instead of providing additional useful training data, these images might have introduced noise, making it more difficult for the model to extract meaningful features.

Prior research has demonstrated that GAN-generated images can enhance the model robustness and accuracy, particularly in addressing class imbalances. However, these studies often utilize different generative architectures and datasets, which could explain the discrepancy in results. GAN-based approaches have been shown to produce highly realistic images, whereas the Juggernaut XL model, while advanced, may require further fine-tuning to generate images that effectively complement real-world datasets. Figure 6 illustrates several examples of high-quality synthetic images that may have contributed to the observed performance improvement.

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 3: 1855-1869, 2025*
*DOI: 10.55214/25768484.v9i3.5713*

**Figure 6.**
Improved AI-generated images using Juggernaut XL model.

Figure 7 presents the application used for testing the models, allowing real-time evaluation of music emotion classification and facial emotion recognition. The interface facilitates emotion mapping and dynamic playlist generation based on detected emotional states. The adaptive music player further contributed to user experience by dynamically updating playlists in real-time, learning from user interactions, and refining its recommendations. The system's ability to respond to real-time emotional changes with a response time of under 2 seconds underscores its practicality for real-world applications. Optimizing training strategies, including longer training periods, fine-tuning approaches, and selective filtering of synthetic images based on quality assessment, could improve the system's design.

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 3: 1855-1869, 2025*
*DOI: 10.55214/25768484.v9i3.5713*

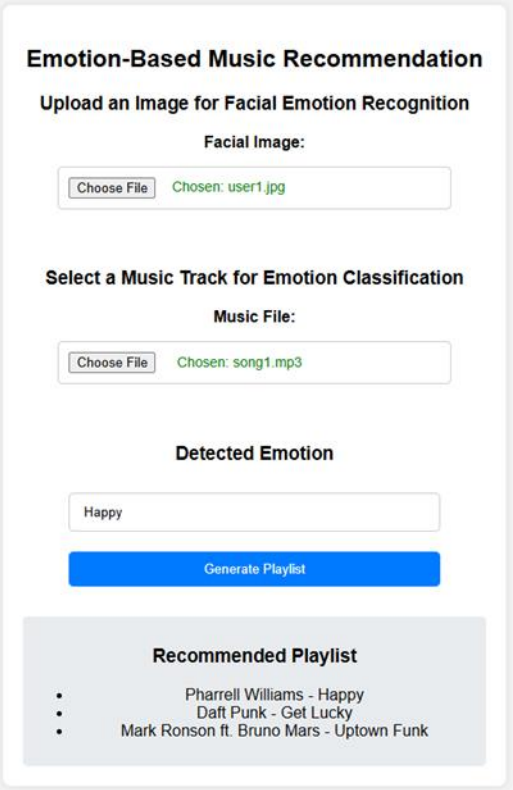


**Figure 7.**
The application developed for testing the models.

## 4. Discussion

This study examined the impact of AI-generated synthetic images on FER model performance, revealing that while synthetic data-maintained performance stability for certain emotion categories, it did not significantly enhance accuracy. The findings emphasize the importance of synthetic data realism and suggest that further refinements in generation techniques and training methodologies are necessary for effective integration. Future research should focus on improving synthetic image quality, increasing training epochs, and exploring alternative generative models to maximize the benefits of synthetic data augmentation in FER.

This research aimed to improve the models by incorporating AI-generated synthetic data into the training process. Using the Juggernaut XL model, we generated 300 synthetic images for each emotion category in the FER2013 dataset and integrated them into the training pipeline of a VGG-19-based FER model. The primary objective was to determine whether introducing synthetic images could enhance key performance metrics such as accuracy, precision, recall, and F1-score. Given the challenges posed by noisy images, class imbalances, and limited diversity in existing FER datasets, synthetic data was expected to contribute to generalizable model performance.

However, our results indicated that the inclusion of synthetic data did not significantly improve the model's performance. In fact, the model trained on the augmented dataset with synthetic images achieved slightly lower accuracy than the model trained on the original FER2013 dataset. One possible explanation for the lack of performance improvement is the limitations of the synthetic images. Although AI-generated data can expand the diversity of training datasets, it does not always guarantee a perfect representation of real-world expressions. The Juggernaut XL model, despite being an advanced generative model, may not have produced images with sufficient realism to enhance FER performance [24]. Previous studies have demonstrated that synthetic images generated by GANs or

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 3: 1855-1869, 2025*
*DOI: 10.55214/25768484.v9i3.5713*

diffusion models can improve models, particularly by addressing class imbalances. However, these studies often rely on highly refined generative techniques and extensive post-processing to ensure the quality of synthetic images. The discrepancies between our findings and prior research suggest that generative models must be optimized further to produce high-quality images that closely mimic natural facial expressions [25].

Several experimental limitations may have also contributed to the lack of performance gains. One of the main constraints was the training duration of the VGG-19 model. Due to hardware limitations, the model was trained for only 25 epochs, whereas the original segmentation VGG-19 model achieved optimal performance after 67 epochs. Given the complexity of FER tasks, a longer training duration may have allowed the model to better adapt to the synthetic data and extract more meaningful features from it. While generating 300 synthetic images per emotion category provided a substantial augmentation, this number may still be insufficient to significantly impact the training process. A larger dataset with more variations in facial expressions and environmental conditions could lead to better generalization and improved recognition performance.

Future research should explore multiple strategies to maximize the potential of synthetic data in FER. Improving the quality and realism of synthetic images through advanced generative models is a crucial first step. While the Juggernaut XL model provided a useful baseline, exploring alternative architectures such as diffusion models or StyleGAN could yield higher-fidelity images that better capture the complexities of human facial expressions. High-quality synthetic images that accurately mimic real-world variations can enhance model robustness, making FER systems more reliable in practical applications [26]. One potential approach is using an automated filtering mechanism to assess the realism of generated images before integrating them into the dataset. This could be achieved through discriminator networks that score the quality of synthetic images or through human-in-the-loop verification processes [27]. Another promising direction for future research is leveraging more advanced fine-tuning techniques to train models entirely on synthetic datasets. One such technique is DreamBooth, which fine-tunes a diffusion model using only a small number of real images to generate highly realistic synthetic data. DreamBooth allows models to learn details from a small dataset while maintaining high consistency in generated images [28].

A key challenge in integrating synthetic data is ensuring that it complements real-world images rather than distorting the training distribution. Increasing the volume and diversity of synthetic images is another promising direction. While this study generated a fixed number of synthetic images per emotion category, future research could investigate the impact of varying the number of synthetic images to determine the optimal ratio of real to synthetic data. Fine-tuning strategies such as transfer learning, domain adaptation, and feature alignment techniques could also be explored to ensure that synthetic images contribute positively to model performance rather than introducing additional noise [29]. Also exploring adversarial training techniques, where a FER model is trained alongside a generative model to improve its ability to distinguish between real and synthetic expressions, could lead to more robust results.

## 5. Conclusion

This study investigated the impact of AI-generated synthetic data on facial emotion recognition model performance by integrating synthetic images generated using the Juggernaut XL model into the FER2013 dataset. The primary goal was to determine whether synthetic data could enhance model accuracy, precision, recall, and F1-score by mitigating the limitations of real-world datasets, such as class imbalances, noise, and limited diversity. However, the results indicate that the inclusion of synthetic images did not significantly improve the model's performance. Instead, the model trained on the augmented dataset achieved slightly lower accuracy compared to the one trained on the original FER2013 dataset.

One of the key findings is that the effectiveness of synthetic data for FER is highly dependent on the quality and realism of the generated images. Despite carefully designed prompts to capture diverse facial

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 3: 1855-1869, 2025*
*DOI: 10.55214/25768484.v9i3.5713*

expressions and demographic variations, the synthetic images may not have sufficiently replicated the subtle complexities of real-world facial expressions. This suggests that, while synthetic data can increase dataset diversity, it may also introduce noise if the images lack fidelity to natural human expressions. The training duration was limited to 25 epochs due to hardware constraints, which may have affected the model's ability to fully leverage the synthetic data. The number of synthetic images generated per emotion category, while substantial, may also not have been sufficient to provide meaningful improvements.

Future research should explore alternative generative models, such as diffusion models or StyleGAN, which have demonstrated the ability to produce higher-quality, more realistic images. Optimizing training strategies, including longer training periods, fine-tuning approaches, and selective filtering of synthetic images based on quality assessment, could enhance the effectiveness of synthetic data augmentation in FER. Another promising avenue is the development of advanced techniques such as DreamBooth to fine-tune diffusion models for datasets. Such methods could allow models to learn from a small number of real images while generating high-fidelity synthetic data that more accurately mimics real-world variations. In conclusion, the findings from this study suggest that further improvements in synthetic image generation and optimization are necessary before AI-generated data can fully replace or significantly enhance traditional datasets.

## Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

## Copyright:



## References

[1] M. Mattioli and F. Cabitza, "Not in my face: Challenges and ethical considerations in automatic face emotion recognition technology," *Machine Learning and Knowledge Extraction*, vol. 6, no. 4, pp. 2201-2231, 2024. https://doi.org/10.3390/make6040109

[2] J. Le Ngwe, K. M. Lim, C. P. Lee, T. S. Ong, and A. Alqahtani, "PAtt-Lite: Lightweight patch and attention MobileNet for challenging facial expression recognition," *IEEE Access*, vol. 12, pp. 79327–79341, 2024. https://doi.org/10.1109/ACCESS.2024.3407108

[3] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE transactions on affective computing*, vol. 13, no. 3, pp. 1195-1215, 2020. https://doi.org/10.1109/TAFFC.2020.2981446

[4] R. He *et al.*, "Is synthetic data from generative models ready for image recognition?," *arXiv preprint arXiv:2210.07574*, 2022. https://doi.org/10.48550/arXiv.2210.07574

[5] G. Bae *et al.*, "Digiface-1m: 1 million digital face images for face recognition," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 3526-3535.

[6] G. Raut and A. Singh, "Generative AI in vision: A survey on models, metrics and applications," *arXiv preprint arXiv:2402.16369*, 2024. https://doi.org/10.48550/arXiv.2402.16369

[7] I. Goodfellow *et al.*, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139-144, 2020. https://doi.org/10.1145/3422622

[8] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4401-4410.

[9] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8110-8119.

[10] T. C. Ma and D. H. Nguyen, "VN3DFace: 3D facial expression database of Vietnamese," *Edelweiss Applied Science and Technology*, vol. 9, no. 1, pp. 161-175, 2025. https://doi.org/10.55214/25768484.v9i1.3883

[11] X. He *et al.*, "SynFER: Towards boosting facial expression recognition with synthetic data," *arXiv preprint arXiv:2410.09865*, 2024. https://doi.org/10.48550/arXiv.2410.09865

[12] D. Meng, X. Peng, K. Wang, and Y. Qiao, "Frame attention networks for facial expression recognition in videos," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019: IEEE, pp. 3866-3870.

[13] A. Kartbayev, "Learning word alignment models for kazakh-english machine translation," in *Integrated Uncertainty in Knowledge Modelling and Decision Making: 4th International Symposium, IUKM 2015, Nha Trang, Vietnam, October 15-17, 2015, Proceedings 4*, 2015: Springer, pp. 326-335.

[14] S. Ammar, T. Bouwmans, N. Zaghden, and M. Neji, "Towards an effective approach for face recognition with DCGANs data augmentation," in *Advances in Visual Computing: 15th International Symposium, ISVC 2020, San Diego, CA, USA, October 5–7, 2020, Proceedings, Part I 15*, 2020: Springer, pp. 463-475.

[15] K. Kim and H. Park, "Neural network approach to predict the association between blood cadmium levels and hypertension," *Edelweiss Applied Science and Technology*, vol. 8, no. 4, pp. 336-344, 2024. https://doi.org/10.55214/25768484.v8i4.978

[16] N. Yalçin and M. Alisawi, "Introducing a novel dataset for facial emotion recognition and demonstrating significant enhancements in deep learning performance through pre-processing techniques," *Heliyon*, vol. 10, no. 20, p. e38913, 2024. https://doi.org/10.1016/j.heliyon.2024.e38913

[17] M. Sajjad *et al.*, "A comprehensive survey on deep facial expression recognition: challenges, applications, and future guidelines," *Alexandria Engineering Journal*, vol. 68, pp. 817-840, 2023. https://doi.org/10.1016/j.aej.2023.01.017

[18] P. Esser *et al.*, "Scaling rectified flow transformers for high-resolution image synthesis," in *Proceedings of the 41st International Conference on Machine Learning (ICML'24)*, 2024.

[19] S. Vignesh, M. Savithadevi, M. Sridevi, and R. Sridhar, "A novel facial emotion recognition model using segmentation VGG-19 architecture," *International Journal of Information Technology*, vol. 15, no. 4, pp. 1777-1787, 2023. https://doi.org/10.1007/s41870-023-01184-z

[20] P. K. Sahoo *et al.*, "An improved VGG-19 network induced enhanced feature pooling for precise moving object detection in complex video scenes," *Ieee Access*, vol. 12, pp. 45847–45864, 2024. https://doi.org/10.1109/ACCESS.2024.3381612

[21] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 158, pp. 279-317, 2019. https://doi.org/10.1016/j.isprsjprs.2019.09.006

[22] H. Schieber, K. C. Demir, C. Kleinbeck, S. H. Yang, and D. Roth, "Indoor synthetic data generation: A systematic review," *Computer Vision and Image Understanding*, vol. 240, p. 103907, 2024. https://doi.org/10.1016/j.cviu.2023.103907

[23] A. Kartbayev, "SMT: A case study of kazakh-english word alignment," in *International Conference on Web Engineering*, 2015: Springer, pp. 40-49.

[24] P. A. Apellániz, J. Parras, and S. Zazo, "Improving synthetic data generation through federated learning in scarce and heterogeneous data scenarios," *Big Data and Cognitive Computing*, vol. 9, no. 2, p. 18, 2025. https://doi.org/10.3390/bdcc9020018

[25] N. Smatov, R. Kalashnikov, and A. Kartbayev, "Development of context-based sentiment classification for intelligent stock market prediction," *Big Data and Cognitive Computing*, vol. 8, no. 6, p. 51, 2024. https://doi.org/10.3390/bdcc8060051

[26] L. Laurier, A. Giulietta, A. Octavia, and M. Cleti, "The cat and mouse game: The ongoing arms race between diffusion models and detection methods," *arXiv preprint arXiv:2410.18866*, 2024. https://doi.org/10.48550/arXiv.2410.18866

[27] V. C. N. Maturana, A. L. Sandoval Orozco, and L. J. García Villalba, "Exploration of metrics and datasets to assess the fidelity of images generated by generative adversarial networks," *Applied Sciences*, vol. 13, no. 19, p. 10637, 2023. https://doi.org/10.3390/app131910637

[28] N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, and K. Aberman, "Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) IEEE*, 2023, pp. 22500-22510.

[29] S. F. Ahmed *et al.*, "Deep learning modelling techniques: Current progress, applications, advantages, and challenges," *Artificial Intelligence Review*, vol. 56, no. 11, pp. 13521-13617, 2023. https://doi.org/10.1007/s10462-023-10466-8