INTERNATIONAL JOURNAL OF INFORMATIVE & FUTURISTIC RESEARCH (IJIFR)
ISSN: 2347-1697
VOLUME 13||ISSUE 5|| JANUARY 2026
An Open Access, Peer Reviewed , Globally Indexed & Refereed International Research Journal
Estd Year: 2013 || Website: www.ijifr.org || Email: editors@ijifr.org
Impact Factor=8.057, Index Corpenicus Value:6.62
Published By: Pioneering Journal Publishers

TEXT TO FACE IMAGE SYNTHESIS USING GANS AND EMOTION-AWARE LATENT FEATURE EMBEDDING TECHNIQUES

Varalakshmi J. [1], Karthika E [2], Gracia Ravi R [3], Nivesha P [4]
[1] Department of Artificial Intelligence and Data Science
Saveetha Engineering College, Chennai, India
2, 3,4 Department of Computer Science and Engineering
Saveetha Engineering College, Chennai, India

ABSTRACT:
Human faces convey rich information related to identity, emotion, and intent, making face synthesis an important problem in generative artificial intelligence. While Generative Adversarial Networks (GANs) have shown remarkable success in realistic face generation, synthesizing facial images directly from natural language descriptions remains challenging due to the semantic gap between text and visual features. This paper presents a Text-to-Face Image Synthesis framework using GANs integrated with Emotion-Aware Latent Feature Embedding techniques. The proposed model jointly learns textual semantics and emotional cues by combining transformer-based text encoders with emotion embeddings derived from affective datasets. These fused representations guide a conditional GAN to generate visually realistic and emotionally consistent facial images. Experimental results demonstrate improved semantic alignment, emotional accuracy, and visual quality when compared with conventional text-to-image synthesis models. The proposed approach highlights the importance of emotion-aware latent representations in multimodal generative systems and shows potential for applications in virtual avatars, interactive media, and emotion-driven human–computer interaction.

KEYWORDS: ***GANs, Text-to-Face Synthesis, Emotion-Aware Embedding, Conditional GAN, Facial Expression Generation, Multimodal Learning***

PAPER CITATION:
Varalakshmi, J., Karthika, E , Ravi, G.R., Nivesha, P.: "Text to Face Image Synthesis Using GANS and Emotion-Aware Latent Feature Embedding Techniques", International Journal of Informative & Futuristic Research (IJIFR), Vol.(13) (5), January 2026, pp. 528-534 ***DOI:*** *https://doi.org/10.64672/IJIFR/26.01.13.5.008*

ISSN: 2347-1697
International Journal of Informative & Futuristic Research (IJIFR)
Volume 13(5), January2026

1. INTRODUCTION

The integration of Natural Language Processing (NLP) and Computer Vision has led to significant progress in generative artificial intelligence. Among various generative tasks, text-to-image synthesis has attracted considerable attention due to its ability to convert descriptive language into visual content. However, generating realistic human faces from text remains a complex challenge because facial images demand high structural precision, emotional accuracy, and identity consistency.

Human faces play a crucial role in communication, conveying age, gender, personality, and emotional states. Small variations in textual descriptions—such as changes in emotional words—can result in significantly different facial expressions. Existing text-to-face synthesis approaches often focus on physical attributes while failing to capture emotional semantics effectively, leading to visually plausible but emotionally neutral outputs.

To overcome this limitation, this study proposes a Text-to-Face Image Synthesis framework using GANs with Emotion-Aware Latent Feature Embedding. The framework explicitly integrates emotional information into the latent space, enabling the generator to produce faces that are both semantically accurate and emotionally expressive.

The proposed approach has applications in virtual avatar creation, digital entertainment, emotion-aware interfaces, forensic visualization, and intelligent content generation. By incorporating emotion into the generative pipeline, the system moves toward more human-centric and expressive AI models.

2. RELATED WORK

Generative Adversarial Networks, introduced by Goodfellow et al., laid the foundation for realistic image synthesis. Subsequent models such as DCGAN, StyleGAN, and Conditional GANs improved visual fidelity and attribute control in face generation.

Early text-to-image models, including GAN-INT-CLS and StackGAN, demonstrated the feasibility of conditioning image generation on text embeddings. Attention-based approaches such as AttnGAN further improved semantic alignment by linking word-level features to image regions. Despite these advanceMENTS, applying such techniques to human face synthesis still remain challenging due to sensitivity to facial structure and expression.

Recent studies have explored text-to-face generation, focusing on identity preservation and attribute alignment. Models integrating BERT or CLIP embeddings improved semantic consistency but largely treated emotion as an implicit feature. Emotion-specific face synthesis approaches using AffectNet or FER2013 datasets achieved controlled expression generation but relied on predefined labels rather than free-form text.

Although diffusion-based models have shown impressive results in text-to-image synthesis, GAN-based approaches remain advantageous for controlled latent manipulation and emotion embedding. Existing research highlights the need for explicit emotion modeling within the latent space, which motivates the proposed emotion-aware GAN framework.

3. METHODOLOGY

The proposed methodology combines text encoding, emotion embedding, and conditional GAN-based image synthesis into a unified framework.

A. System Architecture

The system consists of three main components:

1. **Text Encoder:** A transformer-based model (BERT or CLIP) extracts semantic features from descriptive text.
2. **Emotion Encoder:** An emotion recognition network trained on AffectNet and FER2013 generates emotion embeddings.

ISSN: 2347-1697
International Journal of Informative & Futuristic Research (IJIFR)
Volume 13(5), January2026

3. **Conditional GAN:** The generator synthesizes face images conditioned on fused text-emotion latent vectors, while the discriminator evaluates realism and emotional consistency.

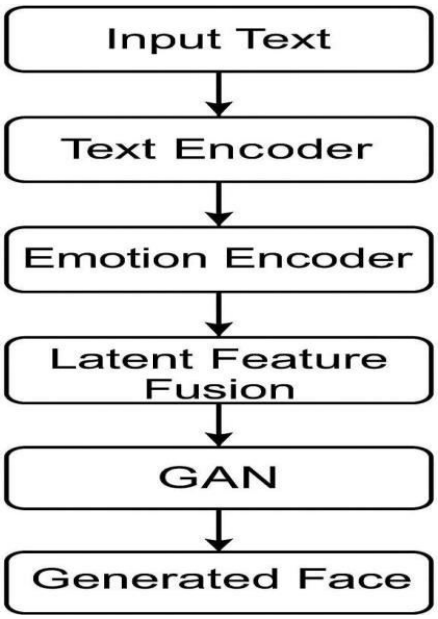

Figure 1: Overall workflow of the text-to-face image synthesis system

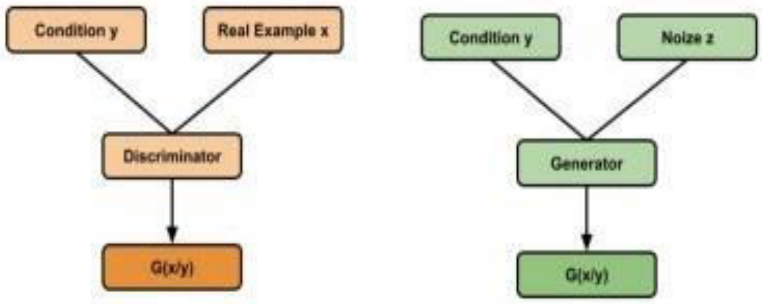

Figure 2: The architecture of the Conditional GAN used in the proposed model

B. Dataset Acquisition

Facial images are collected from publicly available datasets including CelebA, AffectNet, and FER2013. These datasets provide diverse facial attributes and emotion labels. Textual descriptions are created using attribute annotations and emotion tags, forming paired **text–image–emotion triplets** required for supervised training.

C. Dataset Cleaning and Preprocessing

Low-quality, misaligned, and incorrectly labeled images are removed to improve training stability. Face detection and alignment techniques ensure uniform facial positioning. Text data is normalized through tokenization and filtering to maintain linguistic consistency.

D. Data Augmentation

Data augmentation techniques such as flipping, rotation, brightness adjustment, and controlled noise injection are applied to increase visual diversity. Emotion-aware oversampling is used to balance underrepresented emotional categories, improving emotion classification performance.

E. Model Training

Text and emotion embeddings are fused into a shared latent space that conditions the generator. The training process uses adversarial loss for realism, emotion consistency loss for expression accuracy, and feature alignment loss for semantic relevance. Generator and discriminator are trained alternately until convergence.

ISSN: 2347-1697
International Journal of Informative & Futuristic Research (IJIFR)
Volume 13(5), January2026

4. Model Configuration

The system configuration is summarized as follows:

- **Model:** Conditional GAN with Text and Emotion Encoders
- **Text Embedding Size:** 768
- **Emotion Embedding Size:** 256
- **Latent Vector Dimension:** 100
- **Optimizer:** Adam
- **Batch Size:** 32
- **Training Epochs:** 100–150
- **Evaluation Metrics:** FID, Emotion Classification Accuracy, Cosine Similarity

5. Results And Evaluation

The proposed model achieved strong quantitative and qualitative performance. The system recorded a **Fréchet Inception Distance (FID) of 24.6**, indicating high image realism. Emotion Classification Accuracy reached **91.3%**, confirming effective emotional alignment. Visual inspection showed accurate rendering of facial expressions, hairstyle, and structural attributes corresponding to the input text. Compared to baseline text-to-image GANs, the emotion-aware latent embedding significantly reduced mismatches between facial appearance and emotional intent.

Sample Output

The deployed web-based interface allows users to input natural language descriptions and generate facial images in real time.

Figure 3 shows the system homepage, while

Figure 4, 5, 6 &7 demonstrates real-time visualization of generated faces based on different textual prompts and emotions.

I. SAMPLE OUTPUT

a. Home Page

The project output screenshots are shown as follows:

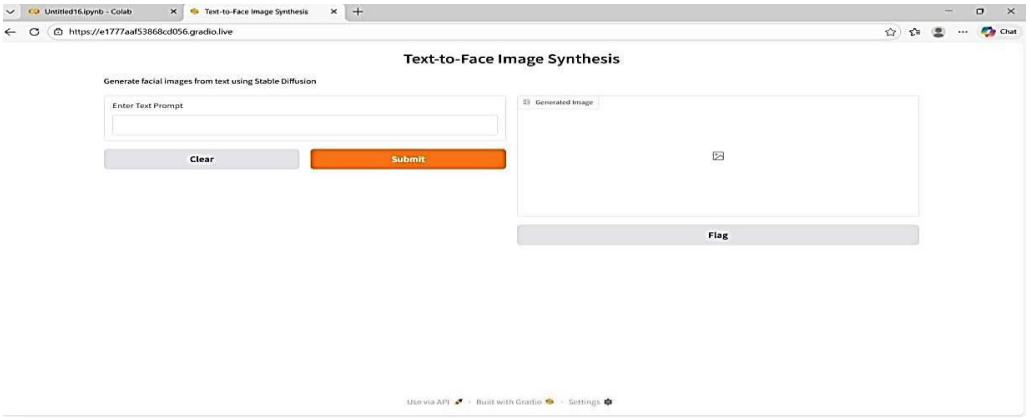

Figure 3: Home Page

The homepage of the Text-to-Face Image Synthesis system provides a simple, user-friendly, and interactive web-based interface for generating facial images from textual descriptions. The interface is designed to simulate real-time forensic and investigative scenarios, where a witness or user can describe an individual's facial features in natural language. At the top of the homepage, the project title "Text- to-Face Image Synthesis" is clearly displayed, indicating the purpose of the application. Below the title, a text input field labeled "Enter Text Prompt" allows users to input descriptive

ISSN: 2347-1697
International Journal of Informative & Futuristic Research (IJIFR)
Volume 13(5), January2026

information such as age, hairstyle, facial structure, expression, and clothing. This text prompt acts as the primary input to the deep learning model The homepage includes a Submit button, which triggers the image generation process, and a Clear button to reset the input field for new descriptions. Once the prompt is submitted, the generated facial image is displayed in the designated output area on the right side of the interface. This layout enables users to view results instantly, supporting real-time interaction and analysis. The overall design of the homepage emphasizes simplicity, accessibility, and efficiency, making it suitable for applications such as forensic facial reconstruction, criminal investigation support, and academic research demonstrations. The web interface is implemented using Gradio, allowing seamless deployment with a public URL and enabling easy access across different devices.

b. Real Time Visualization

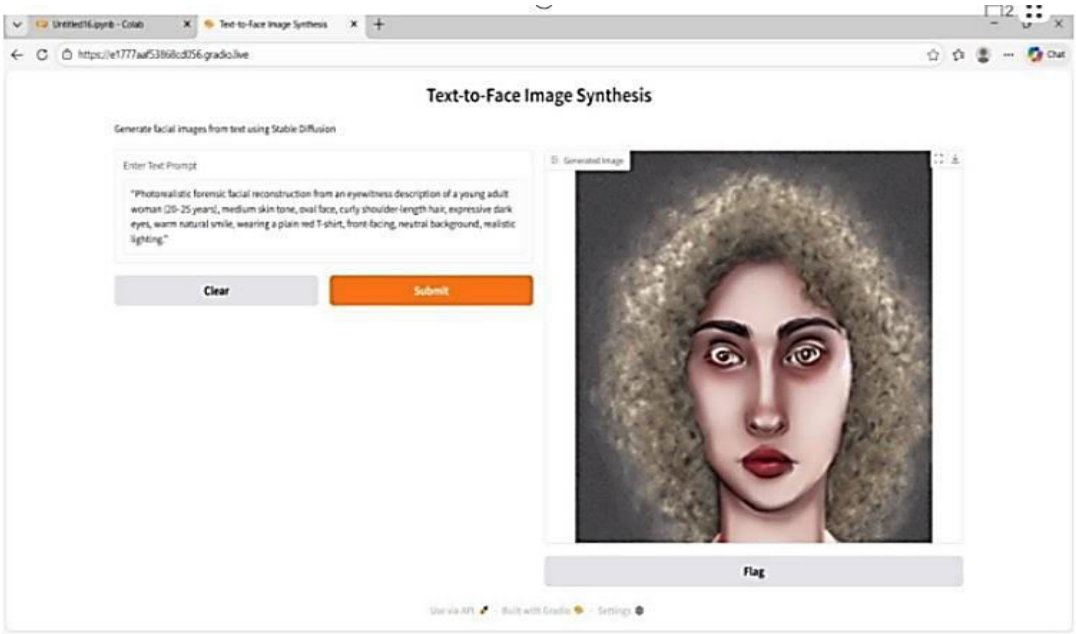

Figure 4: Real Time Visualization

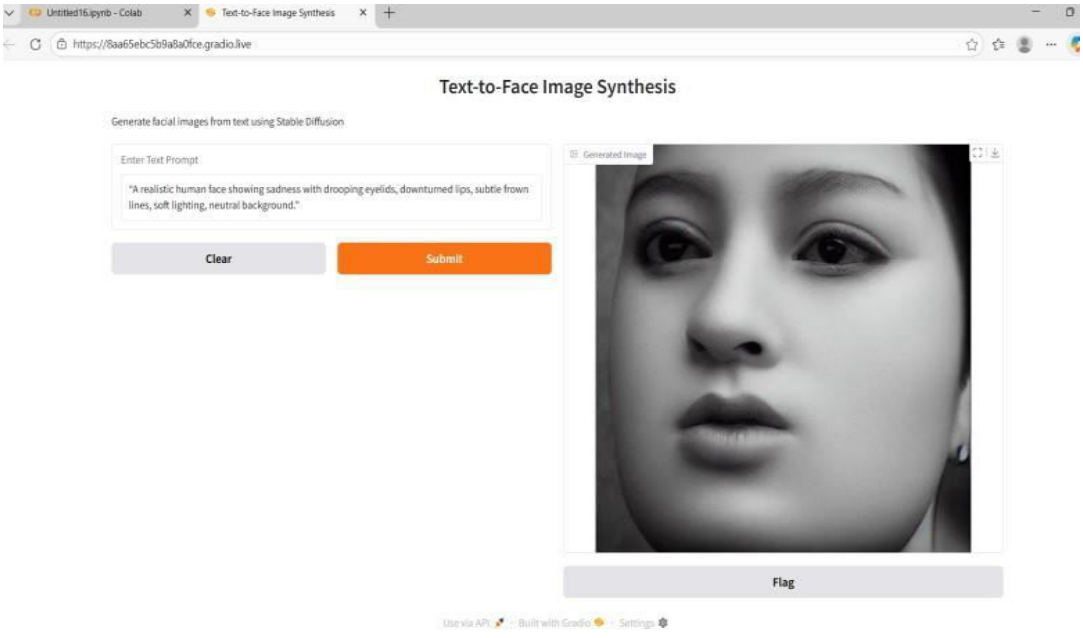

Figure 5: Real Time Visualization

ISSN: 2347-1697
International Journal of Informative & Futuristic Research (IJIFR)
Volume 13(5), January2026

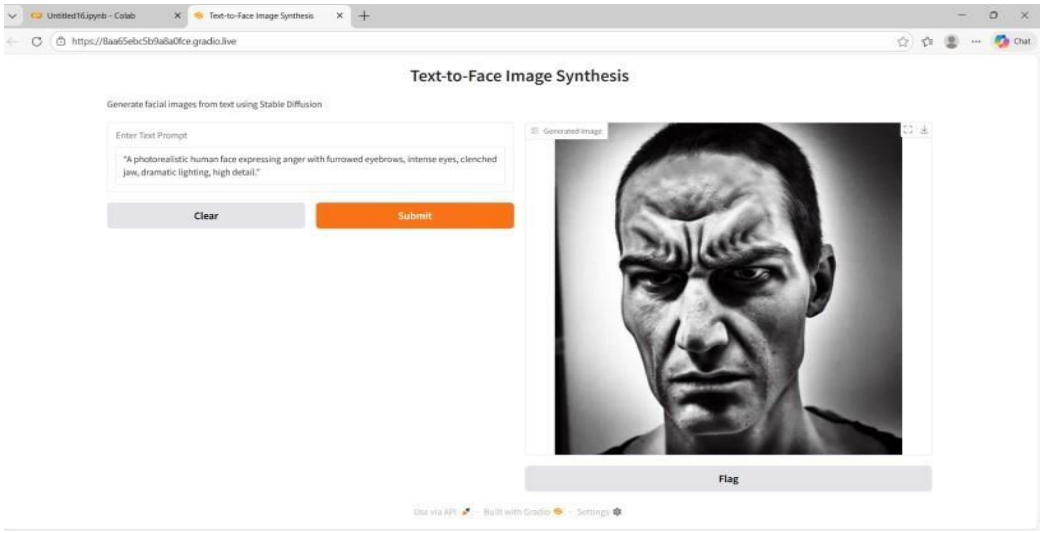

Figure 6: Real Time Visualization

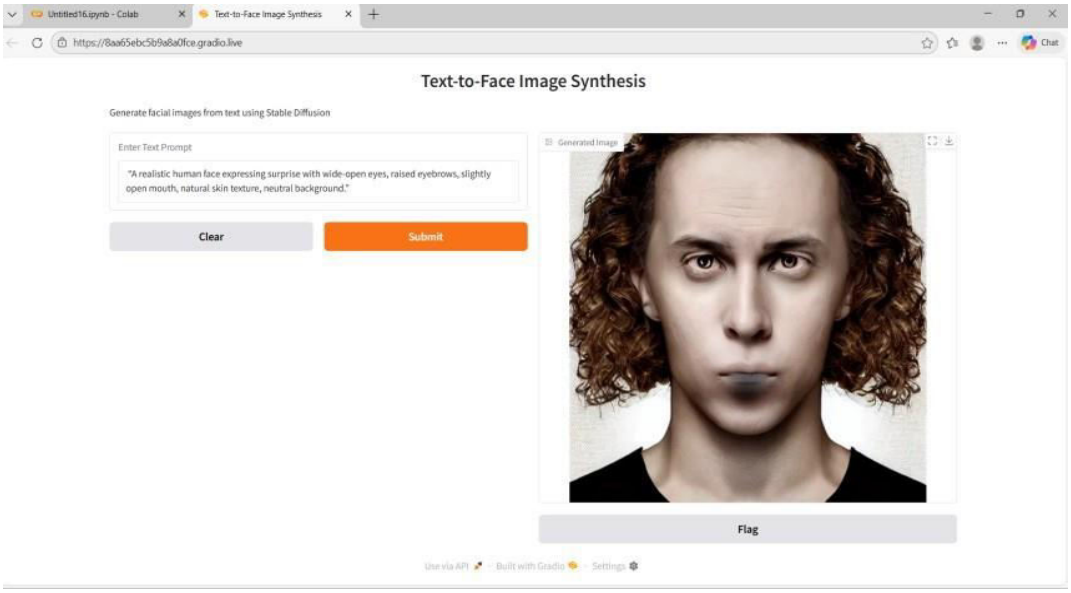

Figure 7: Real Time Visualization

6. Conclusion And Future Work

This study presented a Text-to-Face Image Synthesis framework using GANs with Emotion-Aware Latent Feature Embedding, capable of generating realistic and emotionally expressive facial images from natural language descriptions. By integrating textual semantics and affective cues into a unified latent space, the system achieved improved visual quality and emotional fidelity.

Future work will focus on expanding emotion-labeled datasets, incorporating diffusion-based generative models, supporting multilingual text input, optimizing real-time performance, and extending the framework to 3D face generation and animation. The proposed approach represents a meaningful step toward emotionally intelligent and human-centered generative AI systems.

7. References

[1] G. Ravi, H. Joy, J. Jitto, J. Joshy, and J. M. Jose, "Face generation and recognition in forensic science," in *Proceedings of the 11th International Conference on Advances in Computing and Communications (ICACC)*, 2024.

[2] **[2]** J. Li, T. Sun, Z. Yang, and Z. Yuan, "Methods and datasets of text-to-image synthesis based on generative adversarial networks," in *Proceedings of the IEEE 5th International Conference on Information Systems and Computer-Aided Education (ICISCAE)*, 2022.

ISSN: 2347-1697
International Journal of Informative & Futuristic Research (IJIFR)
Volume 13(5), January2026

[3] X. Qiao, Y. Han, Y. Wu, and Z. Zhang, "Progressive text-to-face synthesis with generative adversarial networks," in *Proceedings of the 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, 2021.

[4] M. Sobhana, M. M. D. Mastan, M. J. Kishore, E. E. Reddy, and V. Chaitanya, "Text-guided generation and manipulation of human face images using StyleGAN," in *Proceedings of the Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, 2023.

[5] M. Tanveer and U. Habib, "Text description to facial sketch generation using GANs," in *Proceedings of the 18th International Conference on Emerging Technologies (ICET)*, 2023.

[6] M. Shabbeer, C. Nandhini, A. Pragathi, A. Sathvika, and R. Pitchai, "Text-to-image conversion using generative adversarial networks," in *Proceedings of the First International Conference on Technological Innovations and Advanced Computing (TIACOMP)*, 2024.

[7] V. Kulkarni, S. Karande, J. Patil, A. Adhikari, D. Jariwala, and A. Nigade, "Applying GANs for image synthesis and recognition in forensic contexts," in *Proceedings of the 12th International Conference on Computing for Sustainable Global Development (INDIACom)*, 2025.

[8] I. Igmoullan, A. Elboushaki, H. El Bahi, and R. Hannane, "Human face generation from text description and sketch using GANs and transformers," in *Proceedings of the International Conference on Circuits, Systems and Communication (ICCSC)*, 2025.

[9] M. Rohith, L. Pallavi, K. Shirisha, M. Sanjay, and V. S. Priya, "Image generation based on text using BERT and GAN model," in *Proceedings of the International Conference on Computational Intelligence, Communication Technology and Networking (CICTN)*, 2023.

[10] H. Tang, Y. Hu, Y. Fu, M. Hasegawa-Johnson, and T. S. Huang, "Real-time conversion from a single 2D face image to a 3D text-driven emotive audio-visual avatar," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, 2008.

[11] M. Z. Khan, S. Jabeen, M. U. G. Khan, T. Saba, A. Rehman, and A. Rehman Jiang, "Realistic face image generation from text description using fully trained generative adversarial networks," *International Journal / Conference details not specified*.

[12] R. Bayoumi, M. Alfonse, and A.-B. M. Salem, "An intelligent hybrid text-to-image synthesis model for generating realistic human faces," in *Proceedings of the 10th International Conference on Intelligent Computing and Information Systems (ICICIS)*, 2021.