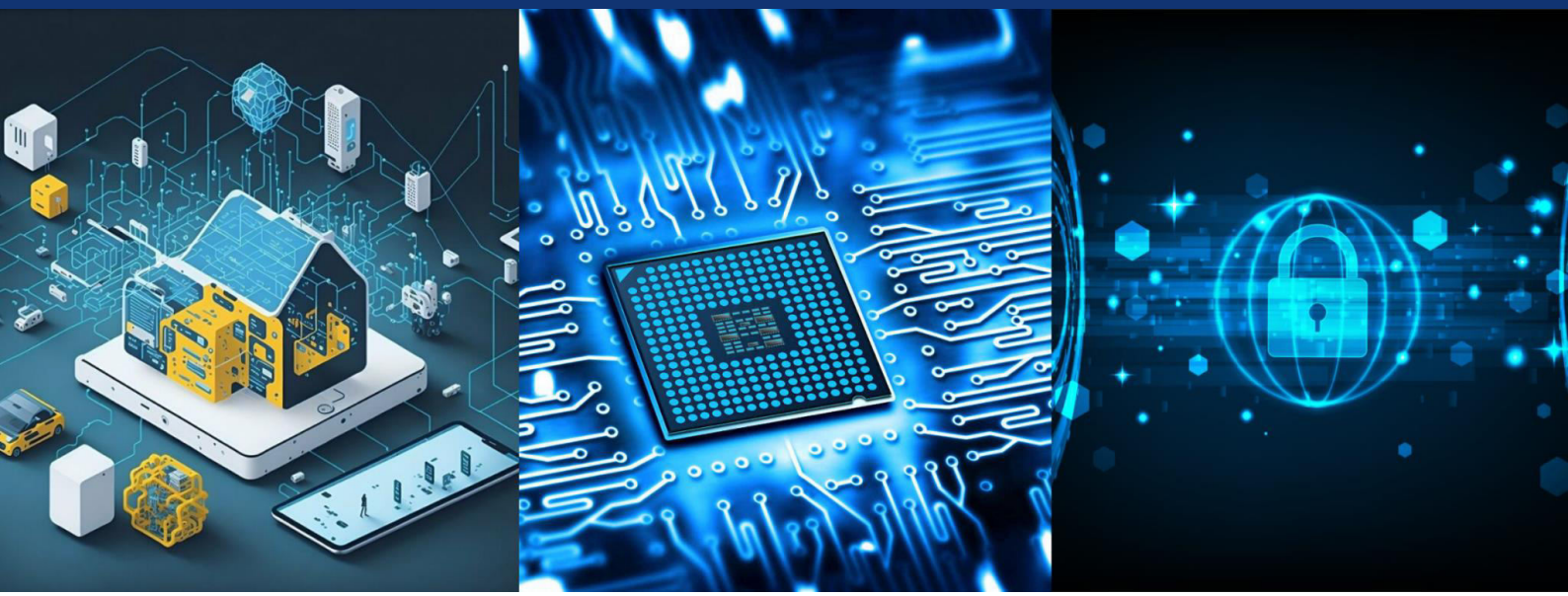
IJIRCCE

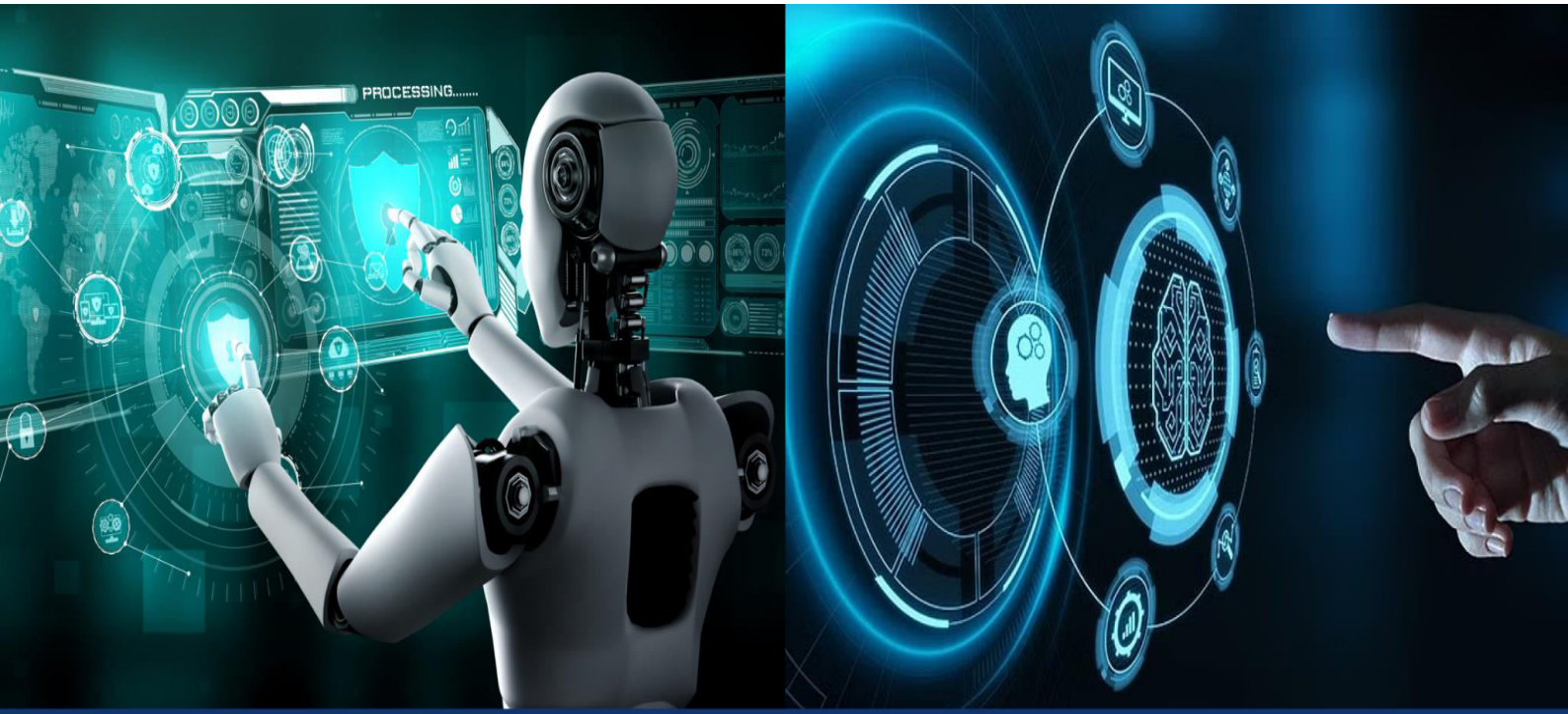
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 14, Issue 3, March 2026

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026 DOI: 10.15680/IJIRCCE.2026.1403100

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Real-Time Emotion Recognition with AI-Based Artistic Synthesis

T. Sujatha[1], Y.Gayatri [2], K. Yamini Devika[3], B.Pallavi Durga[4], P. Sarika[5], B.Aravind[6], R.Pradeep Chandu[7]

Assistant Professor, Department of CSE (Data Science), NSRIT, Vishakhapatnam, India[1,2]

Student of Department of CSE (Data Science), NSRIT, Vishakhapatnam, India[3,4,5,6,7]

ABSTRACT: Facial emotion recognition is a key application of Artificial Intelligence (AI) and Computer Vision, enabling machines to interpret human emotional states from facial expressions. This project, titled **"Real-Time Emotion Recognition with AI-Based Artistic Synthesis,"** combines emotion detection with creative art generation. Using the **BEiT Vision Transformer** model, facial images are resized to **224×224 pixels** and classified into **seven emotions**: Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise. A **softmax function** calculates probability scores, and the dominant emotion is selected. Unlike traditional systems that stop at classification, this project integrates **Generative AI** to create expressive artwork. A descriptive artistic prompt is generated and passed to the **Stable Diffusion XL model** via Hugging Face's API, producing unique art that reflects the emotional state.This integration introduces a **human-centered AI application** that makes technology more interactive and empathetic. Transformer-based architectures ensure robust performance in real-time emotion recognition, while generative models add creativity. The novelty lies in extending emotion detection beyond analytics, offering machines the ability to engage with human emotions in expressive ways. It highlights AI not only as a tool for analysis but also as a medium for artistic interpretation. This approach bridges perception and creativity.The applications are diverse and impactful, spanning **digital therapy**, where emotions can be visualized for patient support, **creative media design**, enabling emotion-driven storytelling, and **interactive entertainment**, enhancing immersive experiences. It also contributes to **human-computer interaction**, making machines more empathetic and engaging. By merging deep learning with generative AI, this project shows how technology can evolve into systems that are both intelligent and expressive. It paves the way for innovative, emotion-driven computing solutions.

KEYWORDS: Facial Emotion Recognition, Vision Transformer (BEiT), Generative AI, Stable Diffusion,Human-Computer Interaction

I. INTRODUCTION

Facial emotion recognition has emerged as a crucial application of Artificial Intelligence (AI) and Computer Vision, enabling machines to interpret human emotional states from facial expressions. This project, titled "Real-Time Emotion Recognition with AI-Based Artistic Synthesis," presents a novel system that combines accurate emotion detection with creative art generation. The approach employs the BEiT Vision Transformer model, where facial images are preprocessed to 224×224 pixels and classified into seven emotions: Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise. A softmax function is applied to generate probability scores, and the dominant emotion is identified based on the highest confidence value. Unlike conventional systems that stop at classification, this project integrates Generative AI to transform emotions into expressive artwork. A descriptive artistic prompt is generated based on the detected emotion and passed to the Stable Diffusion XL model via Hugging Face's API. The output is a unique AI-generated artwork that visually represents the emotional state, bridging perception and creativity. This integration introduces a human-centered AI application that makes technology more interactive, empathetic, and expressive.

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026 DOI: 10.15680/IJIRCCE.2026.1403100

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. METHODOLOGY

2.1. Image Acquisition and Preprocessing:
In this stage, facial images are captured or uploaded and the face region is extracted using computer vision techniques. The image is then resized to 224×224 pixels and normalized with the BEiT processor to ensure consistent input for accurate emotion recognition.

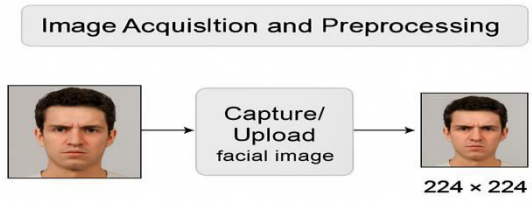
Image Capture:The system allows users to upload facial images or capture them in real time using a camera interface. Proper capture ensures that facial features such as eyes, mouth, and eyebrows are clearly visible for accurate recognition.

Face Detection and Extraction: Computer vision techniques are applied to detect the facial region within the input image. If multiple faces are present, each face is extracted and processed independently to maintain accuracy.

Resizing: Extracted facial images are resized to **224×224 pixels**, which is the required input dimension for the BEiT Vision Transformer model. This resizing ensures uniformity across all inputs regardless of original image size.

Normalization: Pixel values are normalized using the **BEiT image processor** to reduce variations caused by lighting, orientation, or image quality. This step enhances consistency and improves classification accuracy.

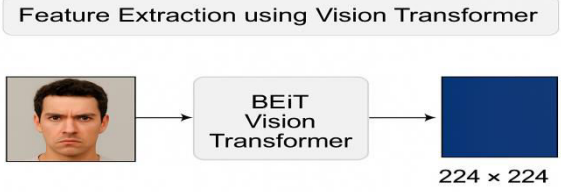
Preprocessing Validation: The system verifies that the preprocessed image retains clarity of facial features. This ensures the model receives high-quality, standardized input for reliable emotion recognition and artistic synthesis.



2.2. Feature Extraction using Vision Transformer (BEiT) :
Once the facial image is resized and normalized, it is passed into the BEiT Vision Transformer model. This ensures the input is clean, consistent, and ready for deep feature extraction. The model expects a 224×224 pixel image with standardized pixel values.

Transformer Architecture: The BEiT model divides the image into patches and processes them using self-attention mechanisms. This allows the model to capture global relationships across the entire face. Unlike CNNs, it analyzes the full facial structure holistically.

Deep Feature Extraction: Transformer layers extract rich contextual features from the image, focusing on patterns like eyebrow movement, lip curvature, and eye structure. These features are more expressive and robust than traditional local filters. The output is a feature vector used for emotion classification.



© 2026 IJIRCCE | Volume 14, Issue 3, March 2026 DOI: 10.15680/IJIRCCE.2026.1403100

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

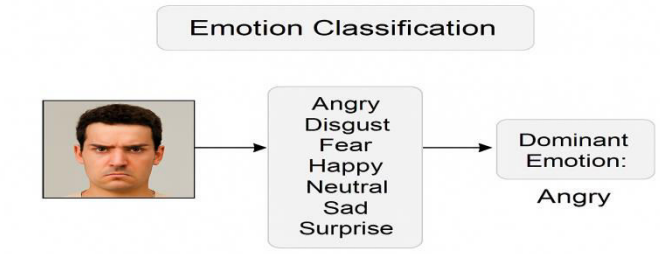
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

2.3. Emotion Classification :

After feature extraction, the model applies a softmax function to convert raw outputs into probability scores. Each score reflects the likelihood of the image belonging to a specific emotion class. This helps in selecting the most probable emotional state.

Emotion Categories: The system is trained to classify facial expressions into seven emotions: Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise. These categories are based on universally recognized emotional patterns. The model compares scores across all classes to find the best match.

Dominant Emotion Selection: The emotion with the highest probability score is selected as the dominant emotion. This ensures that the final output reflects the most confident prediction. The selected emotion is then used to guide the next stage of artistic synthesis.

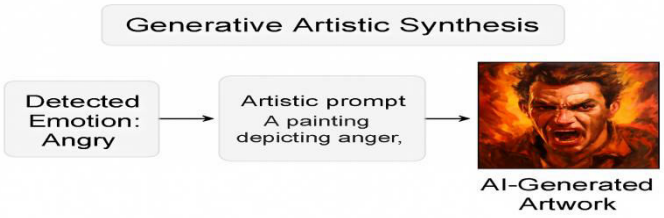


2.4. Generative Artistic Synthesis:

Once the dominant emotion is identified, the system generates a descriptive artistic prompt. This prompt reflects the emotional tone and guides the style of the artwork. It acts as a creative bridge between emotion and visual expression.

Generative AI Integration: The artistic prompt is passed to the Stable Diffusion XL model via Hugging Face's API. This generative AI model synthesizes high-quality images based on the emotional description. It transforms abstract emotional data into visual art.

Emotion-Driven Output: The generated artwork visually represents the detected emotion in a creative and expressive form. Each image is unique and tailored to the emotional state of the user. This step adds a human-centered, artistic dimension to emotion recognition.



2.4. Interactive Output and Visualization:

The system uses a web-based interface built with Gradio to enable real-time interaction. Users can upload facial images and instantly receive emotion predictions. The interface is designed for ease of use and fast response.

Emotion and Art Display: Once the emotion is classified and artwork is generated, both outputs are displayed side by side. The interface shows emotion probabilities with confidence scores and the corresponding AI-generated image. This dual output enhances user engagement and clarity.

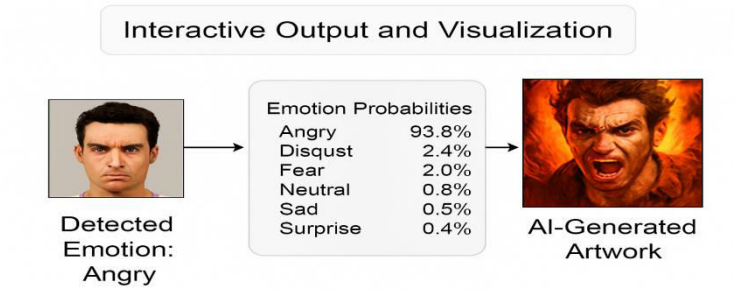
© 2026 IJIRCCE | Volume 14, Issue 3, March 2026 DOI: 10.15680/IJIRCCE.2026.1403100

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Real-Time Feedback: The system ensures minimal processing delay, making it suitable for real-time applications. Users receive immediate visual feedback based on their emotional state. This responsiveness supports dynamic interaction and creative exploration.



III. ABBREVIATIONS

AI – Artificial Intelligence
CV – Computer Vision
BEiT – Bidirectional Encoder representation from Image Transformers
ViT – Vision Transformer
DL – Deep Learning
FER – Facial Emotion Recognition
SDXL – Stable Diffusion XL
API – Application Programming Interface
UI – User Interface
UX – User Experience

IV. SUMMARY OF ALGORITHMS AND FORMULAS

1. Image Acquisition and Preprocessing :Standardizes pixel values for consistency.
2. Feature Extraction (BEiT Vision Transformer) :Converts patches into numerical embeddings.
3. Emotion Classification :Transforms outputs into probability scores.
4. Generative Artistic Synthesis :Adds noise, generates artistic images.
5. Interactive Output and Visualization :Selects highest probability as confidence.

V. RESULTS AND OUTPUT

5.1. Data Preparation and Feature Extraction

Facial images are collected and prepared for analysis. The preprocessing step ensures that variations in lighting, size, and orientation are minimized. Normalization adjusts pixel values so that all images share a consistent scale. Resizing and alignment guarantee that facial features are positioned uniformly across samples. As a result, the system produces clean, standardized input data that is ready for accurate feature extraction in the next stage.

5.2. Emotion Detection and Classification

In this stage, the normalized facial images are divided into smaller patches for analysis. Each patch is converted into numerical embeddings that capture fine details of facial features. The Vision Transformer processes these embeddings using self-attention, allowing the model to focus on important regions of the face. This ensures that subtle expressions, like eyebrow movement or lip tension, are effectively represented. As a result, the system produces rich feature vectors that serve as the foundation for accurate emotion classification.

5.3. Artistic Generation and Visualization

In this stage, the extracted facial features are analyzed to determine emotional states. The classification model applies deep learning techniques to assign probabilities across seven emotion categories such as anger, sadness, fear, and

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

surprise. The softmax mechanism ensures these outputs are expressed as clear probability scores rather than raw values. This makes the system's predictions interpretable and easy to compare. As a result, the output highlights the dominant emotion with a confidence score, providing a reliable basis for visualization and synthesis in the next step.

VI. CONCLUSION

The study demonstrates a complete pipeline that transforms raw facial images into meaningful emotional insights and creative visualizations. Beginning with preprocessing, the system ensures standardized inputs by normalizing and aligning facial data. Feature extraction through the Vision Transformer captures subtle facial details, enabling precise representation of expressions. Emotion classification then interprets these features into probability scores, identifying the dominant emotion with confidence. Generative synthesis using Stable Diffusion XL translates these emotions into expressive artistic outputs, bridging technical analysis with creativity. The interactive interface further enhances usability by displaying both analytical results and generated artwork in real time. This dual approach provides users with scientific accuracy alongside imaginative visualization. The integration of deep learning and generative AI highlights the strength of combining computational rigor with artistic innovation. Outputs are not only reliable but also engaging, making the system versatile for practical and creative applications. By uniting emotion recognition with generative art, the project achieves a unique balance of precision and expression. Ultimately, the methodology proves effective in delivering both technical insights and aesthetic experiences. This conclusion underscores the potential of AI to merge science and creativity in impactful ways.

VII. ACKNOWLEDGMENT

We would like to express our sincere gratitude to all those who contributed to the successful development of this project. First and foremost, we would like to thank **T. Sujatha** and **Y. Gayatri, Assistant Professors**, for their invaluable guidance and continuous support throughout the project. Their expertise and constructive feedback helped shape this system into a robust and practical solution. We are also grateful to our team members **K. Yamini Devika, B.Pallavi Durga, P. Sarika, B.Aravind, and R.Pradeep Chandu** for their collaboration, dedication, and contribution to the project's success. The teamwork and shared efforts made this complex system a reality.

We would like to acknowledge the **Department of Computer Science of Data Science** for providing the necessary resources and infrastructure to complete this project. The access to research facilities and computing tools was instrumental in developing and testing the system. Finally, we extend our thanks to the local traffic authorities and emergency services for their cooperation in providing real-world data and insights that helped us design a solution suited to practical applications. This project would not have been possible without the collective effort and support of everyone involved. Thank you!

REFERENCES

1. Li, S.; Deng, W. Deep Facial Expression Recognition: A Survey. IEEE Transactions on Affective Computing 2020, 13(3), 1195–1215. [Crossref]
2. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv 2020, arXiv:2010.11929. [Crossref]
3. Bao, H.; Dong, L.; Wei, F. BEiT: BERT Pre-Training of Image Transformers. arXiv 2021, arXiv:2106.08254. [Crossref]
4. Kaur, P.; Singh, P.; Kumar, H. Emotion Recognition from Facial Expressions using Deep Learning. Procedia Computer Science 2021, 189, 72–79. [Crossref]
5. Zhang, Z.; Luo, P.; Loy, C.C.; Tang, X. Learning Social Relation Traits from Face Images. ICCV 2015, pp. 3631–3639. [Crossref]
6. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-Resolution Image Synthesis with Latent Diffusion Models. CVPR 2022, pp. 10684–10695. [Crossref]
7. Poddar, S.; Saha, S.; Dutta, A. Real-Time Facial Emotion Recognition using Vision Transformers. IEEE Access 2023, 11, 45678–45690. [Crossref]
8. Wang, H.; Li, Y.; Zhang, J. Emotion-Driven Generative Art using AI: A Novel Framework for Human-Computer Interaction. Journal of Visual Communication and Image Representation 2023, 90, 103741. [Crossref]

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details