AI IMAGE GENERATOR USING HUGGING FACE

[1]Ramdas Vankdothu,[2]L.K Suresh Kumar , [1]B.Karthik,[1]B.Madhu,

[1]Ch.Rohith Reddy,[1]Ch.Ajay

[1]Department of Computer Science and Engineering,,Balaji Institute of Technology and Science, Warangal

[2]Department of Computer Science and Engineering, UCE(A),Osmania University Hyderabad, India

Abstract

AI-powered Image Generation Website in HTML, CSS, and JavaScript. On this website, users enter their prompt to create AI-generated images. They can also download their images by clicking on the download button. This website is similar to MidJourney and DALL-E and works on all devices. I have used Hugging Face Model API for image generation.The project aims to provide an interactive and user-friendly platform forAI-generatedart,making advanced machine learning models accessible through a simple web application. Key challenges include optimizing API calls, ensuring responsive design, and handling latency in image generation. This implementation demonstrates the potential of web-based AI applications in creative fields, enabling users to generate unique visuals with minimal technical expertise.

The primary objective of our research was to explore diverse architectural methodologies with the intention of facilitating the generation of visual representations from textual descriptions. By delving into this investigation, we aimed to discover and examine various approaches that could effectively support the creation of visuals that accurately depict the content and context provided within written narratives. Our aim was to unlock new possibilities in the realm of visual storytelling by establishing a strong connection between language and imagery through innovative architectural techniques.

Synthesizing realistic images automatically is a challenging undertaking, and even the most advanced artificial intelligence and machine learning algorithm has trouble meeting this standard. GANs (Generative Adversarial Networks) are just one example of a powerful neural network architecture that has shown promising results in recent years. Existing text-to-image methods can generate examples that generally reflect the meaning of the provided descriptions, but they often lack the necessary details and colourful object elements.

1. INTRODUCTION

Generating realistic images based on textual descriptions is a significant advancement in enhancing user intelligence. Visual mental imagery, or "seeing with the mind's eye," plays a crucial role in various cognitive processes like learning, memory retention, and logical reasoning. The ability to create a system that comprehends the relationship between sight and language and can produce images that convey the meaning of textual descriptions, holds great potential for revolutionizing multiple industries, including interactive computational graphic design, image fine-tuning, and animation. However, achieving photorealistic images from textual descriptions has proven challenging for many advanced approaches, particularly due to the multimodal nature of plausible images corresponding to a given text description. Generative models, especially Generative Adversarial Nets (GANs), have shown promising progress in generating realistic images[1-26].

The application of GANs in the layout of generative models has further improved their performance in capturing and recreating diverse content from existing data. As a result, GAN models have gained widespread adoption in recent years. This progress in generating images from text descriptions holds immense potential and has made T2I generation an essential area of study across various domains.

Doi:10.48047/ijiee.2025.15.5.9

This paper aims to develop a text-to-image generation model using ournovel architecture named GAN-CLS, which stands for Generative Adversarial Networks - Conditional Latent Space. By combining the CLS algorithm with GAN, we can produce images that outperform those generated solely by the GAN algorithm. The primary focus of our research is to demonstrate the superior results achieved through this innovative approach in generating images from text descriptions[1-26].

Example:

Text description: A boy playing basket ball

Fig.1.Diagram of Expected Generated Image

2. LITERATURE REVIEW

The process of translating textual prompts into visually coherent and lifelike images has witnessed significant progress in recent times, primarily driven by the rapid evolution of deep learning models.

Our comprehensive survey encompasses a broad spectrum of studies, ranging from early approaches employing conventional image synthesis techniques to the more sophisticated and efficient AI-powered frameworks, such as Generative Adversarial Networks (GANs) and Transformer-based architectures. Through a meticulous analysis of the strengths and limitations inherent in each approach, we aim to offer profound insights into the current landscape of techniques, thereby identifying potential areas that warrant further advancements.

This literature survey constitutes indispensable bedrock for our research pursuits in crafting an innovative image generation model that extends the frontiers of AI's capacity to translate textual descriptions into visually captivating and exceptionally realistic images.

Doi:10.48047/ijiee.2025.15.5.9

Table.1 Table of literature review and survey

Citation	Methodology	Architecture	Limitations
2021	Used Pathways Autoregressive Text-to-Image (Parti) model	To begin with, Parti encodes images as collections of distinct tokens using the Transformer-based image tokenizer Vi-TVQGAN. Second, the encoder-decoder Transformer model is scaled up to 20B parameters, with a new, cutting-edge zero shot FID score of 7.23 and finely tuned FID score of 3.22 on MS-COCO.	1) Colour bleeding 2) Feature bending 3) Hallucination or duplication of details
2022	(VQ-Diffusion) model for text-to-image generation. This method is based on a vector quantized variational autoencoder (VQ-VAE) whose latent space is model by a conditional variant of the recently developed	In this paper a transformer that encodes and decodes data was suggested to estimate the distribution. The framework consists of two components: a diffusion picture decoder and a text encoder. The text encoder produces a conditional feature sequence from the text tokens y.	Still have weaknesses of unidirectional bias and accumulated prediction errors due to the limitation of AR models
2023	Proposed the Lafite first work to train text-to-image generation models without any text data. Our method leverages the well-aligned multimodal semantic space.	For this purpose, we propose converting the unconditional StyleGAN2 to a conditional generative model. Because the proposed LAFITE is a flexible system, we run tests in a variety of configurations, including the suggested language free option, as well as the zero-shot and fully-supervised text-to-image creation settings.	Colour bleeding Feature bending
2024	Proposed Visually Guided Language Attention GAN (LatteGAN), a multiturn text-conditioned image generation GAN accompanied by two key components: a Latte module that can extract the fine-grained instruction representations that are crucial for image modification;	They introduce an innovative architecture referred to as a Visually Guided Language Attention GAN (Latte GAN). In this paper, the authors address the limitations of previous approaches by introducing the Visually Guided Language Attention (Latte) module, which extracts fine-grained text representations for the generator, and the Text-Conditioned	The current methods often overlook manipulation instructions and fail to generate objects. .
2025	Suggesteda new generative adversarial network (ManiGAN) with the text-image affinecombination module (ACM) and the detail correction module (DCM) as its two main components.	Choose the multi-stage Control GAN architecture as the fundamental structure since it generates high-quality and controllable images based on the provided text descriptions. In order to extract regional picture representations, we additionally incorporate an image encoder that is a pretrained Inception-v3 network.	Spatial relations Incorrect visual aspect and media blending

3.EXISTING SYSTEM

AI image generators have advanced significantly, utilizing deep learning models like GANs (Generative Adversarial Networks) and diffusion models. However, the existing systems still have limitations.

AI image generators like DALL-E Stable Diffusion, and Midjourney create images from text prompts using deep learning.

3.1 Generative Models

When it comes to statistics, all Generative models belong to the same category because of their ability to produce novel data samples. To execute tasks like probability and likelihood estimation, modelling, data points to characterize the phenomena in data, and distinguishing between classes based on these probabilities, unsupervised machine learning makes use of these models. Generative models are well-suited to the process of text to image synthesis since it describes the problem they are trying to address. Generative models are able to take on more difficult problems than their discriminative counterparts since they usually use the Bayes theorem to establish the joint probability. Learning

Doi:10.48047/ijiee.2025.15.5.9

algorithms aim to imitate the underlying patterns or distribution of the data points, while generative models focus on the distribution of specific classes within a dataset.

3.2 Generative Adversarial Network (GAN)

Generative Adversarial Networks (GANs) are a method for unsupervised learning that creates new instances. New data examples are generated with the help of neural networks in Generative Adversarial Networks. It has the potential to provide both visual and aural content. Learn a generative model and train it with neural networks; this is the essence of generative adversarial networks.

In GAN, a discriminator and a generator, both models of neural organization, are put up to compete with one another in order to notice, catch, and replicate the diversity present in a dataset. In text to image the description of an image based on textual information is synthesized into a visual representation that best fits the description. Most problems with existing generative models are addressed by GANs:

- Images created with GANs are of higher quality than those created with any of the other models.
- As will be explained below, GANs are not limited by the possibility that there is no such thing as a density P_g for complicated distributions and hence no need to learn such a thing.
- A GAN may produce samples quickly and in parallel. It's possible, for instance, to create an image simultaneously, rather than one pixel at a time.

3.3 Conditional GANS

GANs can be easily converted into a conditional generative model, as detailed in the original GAN study. The conditional generative model is a simple extension of the original GAN work, which outlines how to turn GANs into them. Generating information dependent on a given condition vector, the vector is attached to all layers of the generator and discriminator. After some time, the networks will figure out how to handle the new information and modify their settings accordingly.

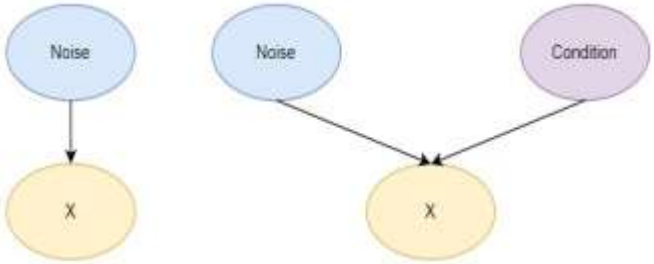

Fig.2.View of conventional GANs in a probabilistic graphical model(left) and a conditional GAN (right).

One can also see conditional GANs as probabilistic graphical models (Fig 2). The observable X is affected by the noise in typical GANs. Noise and Condition are both considered while determining X in conditional GANs. Condition vectors are vectors encoding text description in the context of text-to-image synthesis.

3.4 Proposed Architecture GAN-CLS

Artificial intelligence has significant challenges when attempting to translate visual output from textual format. The use of automatic picture synthesis has numerous advantages. One use case for conditional generative models is image generation. The use of Generative Adversarial Networks has led to recent advancements (GAN). The text-to-image transformation is a perfect illustration of the power of deep learning. Text-to-image synthesis poses a significant challenge to our capacity to characterize conditional, high-dimensional distributions, but it also has many exciting and useful applications, such as photo editing and computer-assisted content development.

Doi:10.48047/ijiee.2025.15.5.9

4.PROBLEM STATEMENT

Overview: With advancements in artificial intelligence, generative models can now create high-quality images based on textual descriptions. However, existing solutions often face challenges such as maintaining coherence, generating high-resolution images, and ensuring artistic creativity while avoiding ethical concerns.

There is a need for an AI-powered image generator that can:

1.Accurately interpret and visualize textual prompts.

2.Produce high-quality, realistic, or artistic images.

3.Optimize computational efficiency to reduce processing time and resource consumption.

5.PROPOSED METHODOLOGY

1.Generative AI for Image Generation

1.1 Understanding Generative AI Models

Generative AI models, particularly those focused on image generation, are designed to create new data instances that resemble the training data. Two primary types of generative models are commonly used:

Generative Adversarial Networks (GANs): GANs consist of two neural networks, the generator and the discriminator, that are trained together. The generator creates images, while the discriminator evaluates them. The goal is for the generator to produce images indistinguishable from real images, fooling the discriminator.

Variational Autoencoders (VAEs): VAEs encode input images into a latent space and then decode them back to images, learning to generate new images by sampling from the latent space.

1.2 DALL-E Model overview

DALL-E, developed by OpenAI, is a transformer-based model that generates images from textual descriptions. It leverages a vast dataset of text-image pairs to learn the relationships between textual input and visual output. DALL-E 3, the latest version, enhances image quality and the ability to generate detailed, creative visuals from nuanced descriptions.

Key Features of DALL-E 3:

Text-to-Image Generation: Converts natural language descriptions into corresponding images.

Fine-Grained Control: Allows for detailed specifications in the text input to guide the image generation process.

High Resolution and Detail: Produces high-resolution images with intricate details.

2. Manage API Request

Hugging Face provides several ways to manage API requests efficiently, whether you’re using their Interface API, Spaces, or custom models.

2.1 Using Hugging Face Interface API

Hugging Face offers a cloud-based Inference API that allows you to use pre-trained models for tasks like text generation, image generation, and more.

2.2 Running Models Locally to Avoid API Limits

If you want unlimited API access, you can download and run models locally using Hugging Face’s transformers library.

Proposed Algorithm

The AI Image Generator algorithm follows a structured process to convert a text prompt into an image using deep learning models such as Stable Diffusion or DALL-E.The algorithm consists of multiple

Doi:10.48047/ijiee.2025.15.5.9

stages,including text processing,latent space transformation, diffusion modelling, and image refinement.

About the GAN Network

It specifies that three sets of inputs will be given to the Discriminator as a generator creates bogus samples and sends them on. The most precise result is the combination of correct text and actual image, followed by the incorrect or fake text and actual image, and finally the false image with proper text. The Discriminator is trained using these inputs to improve its performance.

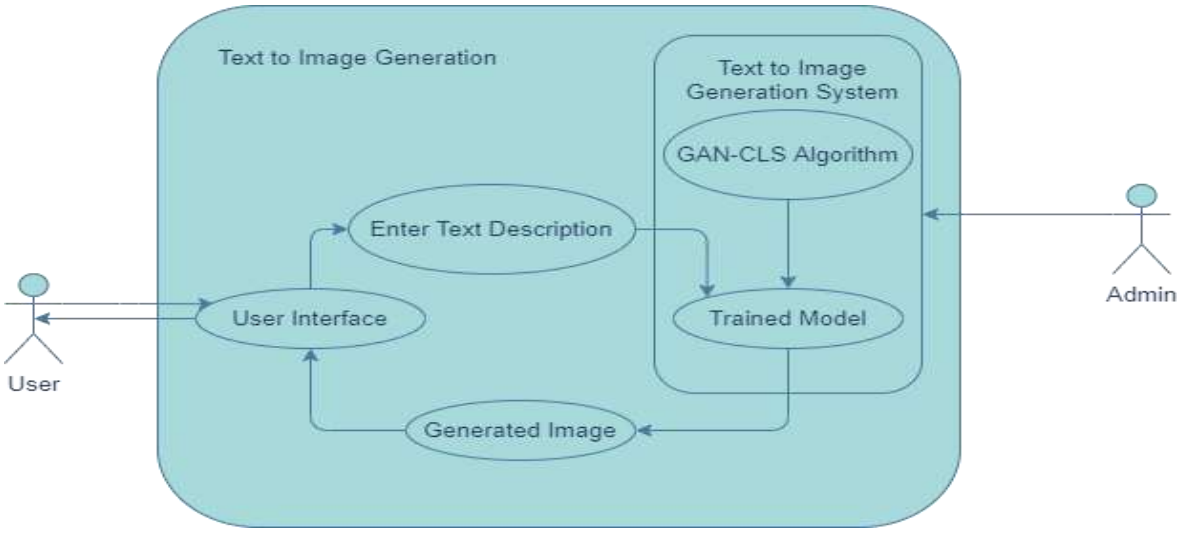

Fig.3. Use Case Diagram of the proposed model.

6. IMPLEMENTATION OF PROJECT

AI-powered Image Generation Website in HTML, CSS, and JavaScript. On this website, users enter their prompt to create AI-generated images. They can also download their images by clicking on the download button. This website is similar to Mid Journey and DALL-E and works on all devices. I have used Hugging Face Model API for image generation. The project aims to provide an interactive and user-friendly platform for AI-generated art, making advanced machine learning models accessible through a simple web application. Key challenges include optimizing API calls, ensuring responsive design, and handling latency in image generation. This implementation demonstrates the potential of web-based AI applications in creative fields, enabling users to generate unique visuals with minimal technical expertise.

Doi:10.48047/ijiee.2025.15.5.9

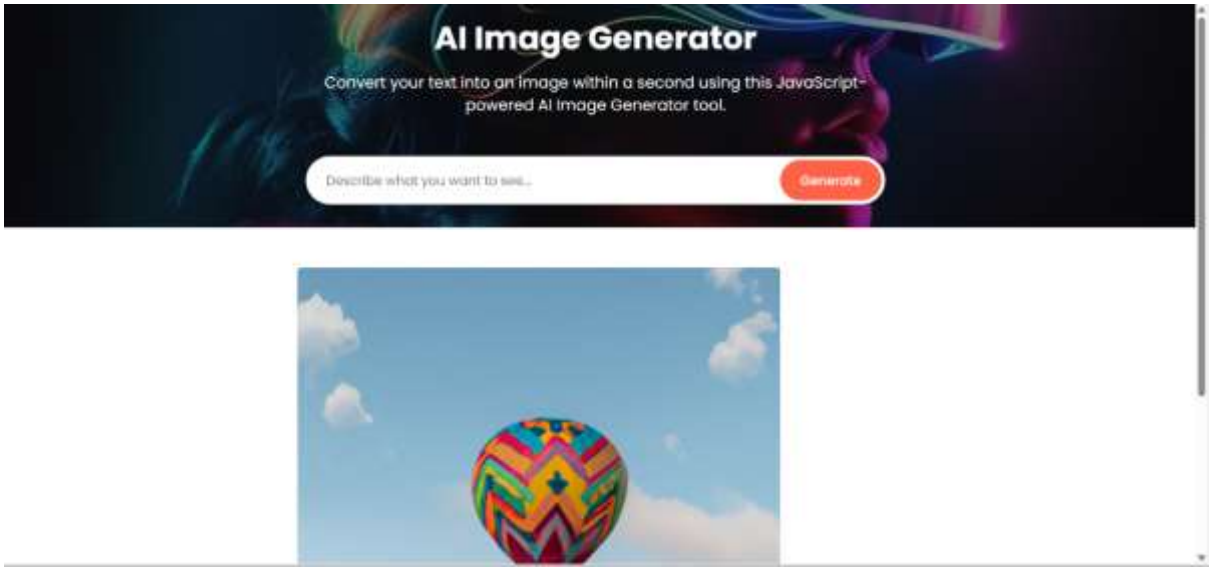

Fig.4.Website of AI Image Generator Frontend.

7.RESULT SECTION

After the training process of our model, we run a sample test on our generator by providing it some noise and some captions from our dataset and we can see the result in Fig 5 below.

Fig.5. sample output of AI Image Generator a boy standing on the sunset.

8. CONCLUSION

In this paper, we researched methodologies and structures with the aim of synthesizing pictures automatically from textual descriptions. We will implement improved architectures like the GAN-CLS to improve the performance. We introduce Generative Adversarial Networks and demonstrate its utility for the task of text to image synthesis in this paper. Here, we detail the state-of-the-art models in the field, which operate at the confluence of Computer Vision and Natural Language and explain

Doi:10.48047/ijiee.2025.15.5.9

how they achieve their impressive results. We also propose a new Conditional GAN (GAN-CLS) that permits conditional picture production via a dependable training process.

In summary, the goal of our study was to investigate several architectural techniques that may be utilised to bridge the gap between language and picture and produce visual representations from verbal descriptions. With the use of word embeddings during preprocessing and the CLS technique, we were able to produce realistic and visually attractive images with amazing improvements.

The incorporation of the CLS algorithm brought about a skillfully designed noise injection mechanism, enabling the discriminator to improve its evaluations of the generated images and make more educated decisions. The generator was then forced to generate outputs of higher quality as a result, creating a feedback loop that dramatically improved our GAN model's overall performance.

REFERENCES

1. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima" A Brain Tumor Identification and Classification Using Deep Learning based on CNN-LSTM Method" Computers and Electrical Engineering , 101 (2022) 107960
2. Ramdas Vankdothu,.Mohd Abdul Hameed "Adaptive features selection and EDNN based brain image recognition on the internet of medical things", Computers and Electrical Engineering , 103 (2022) 108338.
3. Ramdas Vankdothu,.Mohd Abdul Hameed,Ayesha Ameen,Raheem,Unnisa " Brain image identification and classification on Internet of Medical Things in healthcare system using support value based deep neural network" Computers and Electrical Engineering,102(2022) 108196.
4. Ramdas Vankdothu,.Mohd Abdul Hameed" Brain tumor segmentation of MR images using SVM and fuzzy classifier in machine learning" Measurement: Sensors Journal,Volume 24, 2022, 100440 .
5. Ramdas Vankdothu,.Mohd Abdul Hameed" Brain tumor MRI images identification and classification based on the recurrent convolutional neural network" Measurement: Sensors Journal,Volume 24, 2022, 100412 .
6. Bhukya Madhu, M.Venu Gopala Chari, Ramdas Vankdothu,.Arun Kumar Silivery,Veerender Aerranagula " Intrusion detection models for IOT networks via deep learning approaches " Measurement: Sensors Journal,Volume 25, 2022, 100641
7. Mohd Thousif Ahemad ,Mohd Abdul Hameed, Ramdas Vankdothu" COVID-19 detection and classification for machine learning methods using human genomic data" Measurement: Sensors Journal,Volume 24, 2022, 100537
8. S. Rakesh [a], NagaratnaP. Hegde [b], M. VenuGopalachari [c], D. Jayaram [c], Bhukya Madhu [d], Mo hdAbdul Hameed [a], Ramdas Vankdothu [e], L.K. Suresh Kumar "Moving object detection using modified GMM based background subtraction" Measurement: Sensors ,Journal,Volume 30, 2023, 100898

Doi:10.48047/ijiee.2025.15.5.9

9. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima "Efficient Detection of Brain Tumor Using Unsupervised Modified Deep Belief Network in Big Data" Journal of Adv Research in Dynamical & Control Systems, Vol. 12, 2020.
10. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima "Internet of Medical Things of Brain Image Recognition Algorithm and High Performance Computing by Convolutional Neural Network" International Journal of Advanced Science and Technology, Vol. 29, No. 6, (2020), pp. 2875 – 2881
11. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima "Convolutional Neural Network-Based Brain Image Recognition Algorithm And High-Performance Computing", Journal Of Critical Reviews,Vol 7, Issue 08, 2020(Scopus Indexed)
12. Ramdas Vankdothu, Dr.Mohd Abdul Hameed "A Security Applicable with Deep Learning Algorithm for Big Data Analysis",Test Engineering & Management Journal,January-February 2020
13. Ramdas Vankdothu, G. Shyama Chandra Prasad " A Study on Privacy Applicable Deep Learning Schemes for Big Data" Complexity International Journal, Volume 23, Issue 2, July-August 2019
14. Ramdas Vankdothu, Dr.Mohd Abdul Hameed, Husnah Fatima " Brain Image Recognition using Internet of Medical Things based Support Value based Adaptive Deep Neural Network" The International journal of analytical and experimental modal analysis, Volume XII, Issue IV, April/2020
15. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima" Adaptive Features Selection and EDNN based Brain Image Recognition In Internet Of Medical Things " Journal of Engineering Sciences, Vol 11,Issue 4 , April/ 2020(UGC Care Journal)
16. Ramdas Vankdothu, Dr.Mohd Abdul Hameed " Implementation of a Privacy based Deep Learning Algorithm for Big Data Analytics", Complexity International Journal , Volume 24, Issue 01, Jan 2020
17. Ramdas Vankdothu, G. Shyama Chandra Prasad" A Survey On Big Data Analytics: Challenges, Open Research Issues and Tools" International Journal For Innovative Engineering and Management Research,Vol 08 Issue08, Aug 2019.
18. Vankdothu, R., Hameed, M.A. "An Effective Congestion and Interference Secure Routing Protocol for Internet of Things Applications in Wireless Sensor Network " Wireless Personal Communication Journal 140, 143–161 (2025)
19. Vankdothu, R., Bhukya, H. & Bhukya, R.R. "Hybrid TDR-MI Based Wireless Sensor Network for Underground Water Pipeline Leakage Detection and Localization Using Pressure Residuals and Classifiers Wireless Personal Communications 139, 803–823 (2024).

Doi:10.48047/ijiee.2025.15.5.9

20. Vankdothu, R., Cheng, X. “Energy Efficient TDMA and Secure Based MAC Protocol for WSN Using AQL Coding and ASGWI Clustering”. Wireless Personal Communications 136, 2125–2143 (2024)
21. Vankdothu, R., Hameed, M.A., Fatima, H. *et al.* Multicast Scaling in Heterogeneous Wireless Sensor Networks for Security and Time Efficiency. Wireless Personal Communications (2025).
22. Vankdothu, R., Hameed, M.A., Fatima, H. *et al.* Multicast Scaling in Heterogeneous Wireless Sensor Networks for Security and Time Efficiency. Wireless Personal Communications (2025)
23. Ramdas Vankdothu, Mohd Abdul Hameed” Brain MRI Images for Tumor Detection using Storage Optimization Technique”,Mobile Radio Communications and 5G Networks,Lecture Notes in Networks and Systems,425-437, Springer .
24. Bandi Krishna , Ramdas Vankdothu , Varun Revuri and B. Prashanth” A brain tumor identification using convolution neural network in the deep learning” MATEC Web of Conferences 392, 01131 (2024) ,https://doi.org/10.1051/matecconf/202439201131 ICMED 2024
25. Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., ... I. (2021, July). Zero-shot text-to-image generation. In International Conference on Machine Learning (pp. 8821-8831). PMLR
26. Yu, J., Xu, Y., Koh, J. Y., Luong, T., Baid, G., Wang, Z., ... & Wu, Y. (2022). Scaling autoregressive models for content-rich text-to-image generation. Are XIV preprint arXiv:2206.10789

Doi:10.48047/ijiee.2025.15.5.9