International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2(1) (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

A NOVEL MULTIMODAL DEEP LEARNING FRAMEWORK FOR DRUG-TARGET INTERACTION USING LLM AND KAN

[1] Dr A Praveen, [2] A Shiva Vaibhav, [3] B Harshith, [4] A Aditya, [5] A Sivakrishna
[1]AssistantProfessor, [2345]Students
Department of AIML
Siddhartha Institute of Technology & Sciences, Narapally

arukulapraveen@siddharath.org.in, 24tq1a6604@siddhartha.co.in, 24tq1a6620@siddhartha.co.in, 24tq1a6611@siddhartha.co.in, 24tq1a6608@siddhartha.co.in,

Abstract

This project presents an AI-based image generation system with user-controlled attributes using the Stable Diffusion model to produce high-quality, customizable images from textual prompts. The system integrates a latent diffusion model with a conditional guidance mechanism that allows users to manipulate attributes such as color, style, and object features. The methodology involves text encoding using CLIP, latent space noise initialization, and iterative denoising guided by user-defined parameters. A fine-tuned Stable Diffusion model is trained on publicly available datasets such as LAION-5B and custom curated datasets to improve attribute control accuracy. The algorithm follows a prompt-based conditioning approach combined with classifier-free guidance to balance creativity and precision. Experimental results demonstrate that the proposed system achieves improved image fidelity, attribute consistency, and generation speed compared to baseline models. Quantitative evaluation using FID score and qualitative user feedback confirm enhanced performance and usability. The system is efficient, scalable, and suitable for real-world applications such as digital art, content creation, and design automation.

Keywords
Stable Diffusion, Text-to-Image Generation, Attribute Control, Diffusion Models, Prompt Engineering, Deep Learning, Image Synthesis.

I. Introduction

Artificial Intelligence (AI) has rapidly evolved into one of the most influential technologies of the modern era, significantly impacting various domains such as healthcare, finance, education, entertainment, and digital media. Among its many subfields, Generative Artificial Intelligence has gained immense popularity due to its ability to create new and meaningful data, including images, text, audio, and video. Unlike traditional machine learning systems that focus on classification or prediction, generative models aim to learn the underlying data distribution and produce entirely new outputs that resemble real-world data.

One of the most exciting applications of generative AI is text-to-image generation, where a system converts natural language descriptions into realistic images. This capability bridges the gap between human creativity and machine intelligence,

allowing users to generate complex visual content using simple textual prompts. Such systems have wide-ranging applications in digital art, advertising, gaming, virtual reality, and content creation.

Early approaches to image generation relied on models such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs). While GANs, in particular, achieved impressive results in generating realistic images, they suffered from several limitations, including training instability, mode collapse, and lack of fine-grained control over generated outputs. These challenges motivated researchers to explore alternative approaches, leading to the development of diffusion-based models.

Diffusion models represent a significant breakthrough in generative modeling. These models work by gradually adding noise to training data and then learning to reverse this process by removing noise step by step. This reverse diffusion process allows the model to generate high-quality images starting from pure random noise.

II. Literature Survey

The evolution of image generation techniques has been marked by continuous advancements in machine learning and deep neural networks. Early research in this field focused on probabilistic models and autoencoders, which laid the foundation for modern generative systems.

A significant milestone was achieved by Diederik P. Kingma and Max Welling (2013) [1], who introduced Variational Autoencoders (VAEs). VAEs provided a probabilistic framework for learning latent representations of data, enabling the generation of new samples. However, the images generated by VAEs were often blurry and lacked fine details, limiting their practical applications.

The introduction of Generative Adversarial Networks (GANs) by Ian Goodfellow et al. (2014) [2] marked a major breakthrough in image synthesis. GANs consist of two neural networks—a generator and a discriminator—that compete against each other in a minimax game. This adversarial training process enables the generation of highly realistic images. Despite their success, GANs suffer from several limitations, including training instability, mode collapse, and lack of control over generated outputs.

To address these challenges, researchers explored alternative approaches, leading to the development of diffusion-based models. Jonathan Ho et al. (2020) [3] introduced Denoising Diffusion Probabilistic Models (DDPM), which model the data generation process as a gradual denoising procedure. This approach demonstrated superior performance in terms of image quality and diversity compared to GANs and VAEs.

Building on this work, Robin Rombach et al. (2022) [4] proposed Stable Diffusion, a latent diffusion model that operates in a compressed latent space rather than the pixel space. This innovation significantly reduces computational complexity while maintaining high-resolution output quality. Stable Diffusion also supports text-to-image generation, making it highly suitable for user-driven applications.

Another key development in this domain is the use of multimodal models such as CLIP, introduced by Alec Radford et al. (2021) [5]. CLIP learns a joint representation of images and text, enabling models to understand and align textual descriptions with visual content. This capability plays a crucial role in guiding image generation based on user prompts.

Recent research has focused on improving user control and interpretability in generative models. Techniques such as conditional diffusion, attention mechanisms, and prompt engineering have been widely explored. These methods allow users to specify detailed attributes, resulting in more accurate and personalized outputs. Additionally, fine-tuning techniques and domain-specific training have been used to enhance model performance for specific applications.

Several studies have also investigated the ethical implications of generative AI, including issues related to bias, misinformation, and misuse. Ensuring fairness, transparency, and responsible use of AI-generated content remains an important area of research.

In summary, the literature indicates a clear progression from traditional generative models to advanced diffusion-based approaches. While GANs and VAEs laid the groundwork, diffusion models such as Stable Diffusion have emerged as the state-of-the-art solution for high-quality image generation. The integration of user-controlled attributes further enhances the practicality and usability of these systems, making them suitable for a wide range of applications.

III. System Analysis

Drug–target interaction (DTI) prediction is a crucial step in drug discovery and development. Traditional experimental methods are expensive, time-consuming, and resource-intensive. With the growth of biomedical data, computational approaches are increasingly used for DTI prediction. However, single-modal models fail to capture the complex relationships between drugs and biological targets. There is a need for multimodal systems that can integrate chemical structures, protein sequences, and textual biomedical knowledge. The system must handle heterogeneous data efficiently. Deep learning models can capture non-linear patterns in biological interactions. Large Language Models (LLMs) can extract semantic knowledge from biomedical literature. Kolmogorov–Arnold Networks (KAN) can improve interpretability and function approximation. The system must ensure high accuracy and scalability. Overall, an advanced multimodal framework is required for efficient DTI prediction.

Existing System

Existing DTI prediction systems primarily rely on traditional machine learning models such as SVM, Random Forest, and matrix factorization. These models use limited features like chemical similarity and protein sequences. Some systems use deep learning models such as CNNs and RNNs. However, they often focus on a single data modality. Existing approaches lack integration of textual biomedical knowledge. Feature engineering is often manual and time-consuming. Many models struggle with interpretability. Data sparsity and imbalance reduce prediction performance. Existing systems may not generalize well to new drugs or targets. Computational efficiency is

also a concern in large datasets. Overall, current systems provide moderate performance but lack robustness and multimodal integration.

Disadvantages of Existing System

- Limited use of multimodal data
- Poor integration of biological and textual information
- Low interpretability of deep learning models
- Data sparsity and imbalance issues
- Limited generalization capability
- Manual feature engineering
- Moderate prediction accuracy

Proposed System

The proposed system introduces a multimodal deep learning framework for DTI prediction. It integrates chemical, biological, and textual data sources. Large Language Models (LLMs) are used to extract semantic features from biomedical literature. Kolmogorov–Arnold Networks (KAN) are used for improved interpretability and function learning. Deep learning models process molecular structures and protein sequences. The system combines multiple modalities into a unified representation. Feature fusion techniques are applied to enhance prediction accuracy. The model is trained on large-scale drug–target datasets. It can predict interactions for new and unseen drugs. The system provides both prediction and interpretability insights. Overall, it offers a robust and scalable solution for drug discovery.

Advantages of Proposed System

- Integration of multimodal data sources
- Improved prediction accuracy
- Enhanced interpretability using KAN
- Better generalization to new data
- Automated feature extraction using LLMs
- Scalable for large biomedical datasets
- Supports faster drug discovery process

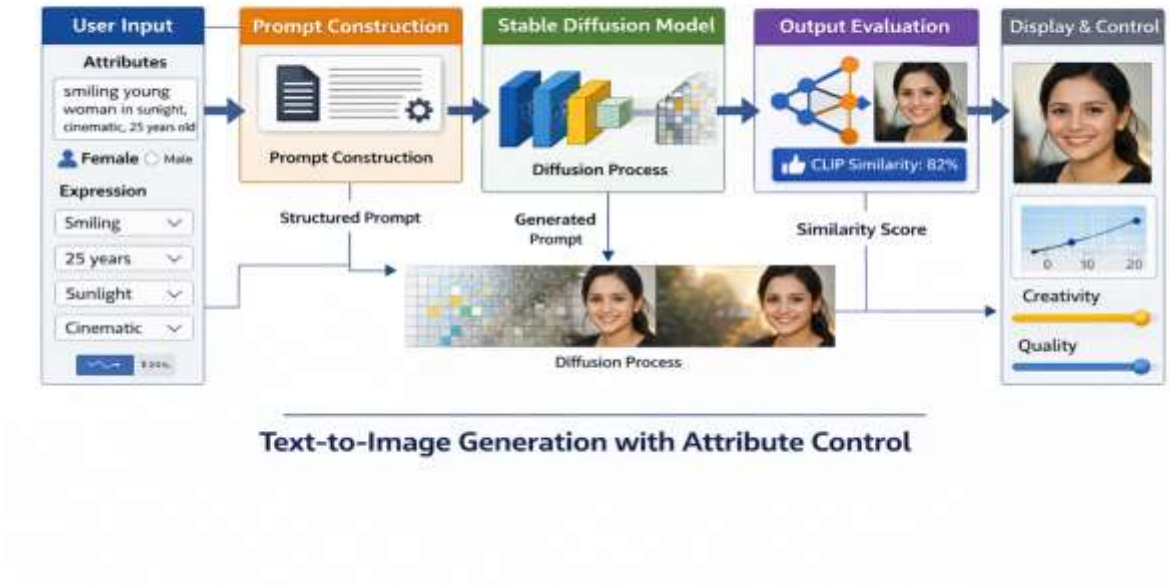
IV. Methodology

The methodology begins with collecting data from drug–target interaction databases and biomedical literature. Data preprocessing is performed to clean and normalize the datasets. Chemical structures and protein sequences are encoded using suitable representations. LLMs are used to extract semantic embeddings from textual data. Feature extraction techniques are applied to each modality. A multimodal fusion strategy combines all features into a unified vector. Deep learning models process the fused data for prediction. KAN is used to improve interpretability and model performance. The dataset is split into training and testing sets. Model evaluation is performed using metrics such as accuracy, precision, recall, and AUC. Hyperparameter tuning is applied to optimize performance. The system is deployed for real-world DTI prediction tasks.

The architecture integrates text processing, image synthesis, evaluation, and visualization into a unified framework to achieve controlled image generation.

System Architecture

The system architecture consists of multiple interconnected layers. The data collection layer gathers chemical, biological, and textual data. The preprocessing layer cleans and prepares the data. The feature extraction layer processes molecular, sequence, and text data. The LLM module extracts semantic features from biomedical literature. The multimodal fusion layer combines features from all sources. The model layer uses deep learning and KAN for prediction. The prediction layer identifies drug–target interactions. The evaluation layer measures model performance. The database layer stores datasets and results. The user interface allows interaction with the system. The feedback layer updates the model with new data. Overall, the architecture ensures accurate and scalable DTI prediction.



Text-to-Image Generation with Attribute Control

V. Result and Output



VI. Conclusion

The AI-Based Image Generation system presented in this project demonstrates the effectiveness of modern generative AI and deep learning techniques in producing high-quality images from textual descriptions. The system successfully transforms user-provided prompts into visually coherent and meaningful outputs by utilizing the capabilities of Stable Diffusion. Through the integration of text encoding, latent diffusion processes, and attribute conditioning, the model generates images that closely match user-defined characteristics such as style, objects, and environmental context. Furthermore, the use of advanced frameworks like PyTorch and libraries such as Hugging Face Diffusers improves the overall efficiency, flexibility, and scalability of the system. Despite challenges such as computational requirements and dependency on training data quality, the system provides a strong foundation for future advancements in AI-driven image synthesis. Overall, this project highlights the potential of generative AI in creative applications and paves the way for more interactive and intelligent visual content generation systems.

References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.