STM JOURNALS

Journal of
Image Processing & Pattern Recognition Progress

ISSN: 2394-1995
Volume 11, Issue 2, 2024
May–August
DOI (Journal): 10.37591/JoIPPRP

https://journals.stmjournals.com/JoIPPRP

Review JoIPPRP

Text to Image Using Machine Learning

Aashutosh Meshram[1], Yogini Kapse[1],
Shreedhar Mhatre[1], Sandeep Patil[2,*]

Abstract

In the era of digital transformation, our project addresses the convergence of computer vision and natural language processing to enhance user interaction and visual content creation. This project comprises three distinct modules: user authentication and session management, image colorization from grayscale inputs, and text-to-image generation. The login registration module provides secure access to the system, ensuring user privacy and data integrity. Once authenticated, users can utilize advanced computer vision techniques through our image colorization module, where grayscale images are transformed into vibrant, realistic color representations using deep learning models and OpenCV. Additionally, our text-to-image generation module leverages state-of-the-art natural language processing models to convert textual descriptions into visually compelling images. These modules are integrated into a Flask-based web application, ensuring seamless user interaction through intuitive interfaces and robust backend processing. The project not only showcases technical proficiency in AI-driven applications but also underscores practical applications in enhancing digital content creation and user experience.

Keywords: Computer vision, natural language processing (NLP), deep learning, flask web application, image colorization, text-to- image generation

INTRODUCTION

In an increasingly digital world, the intersection of artificial intelligence (AI), computer vision, and natural language processing (NLP) has revolutionized how we create and interact with visual content. This project aims to harness these technologies to develop a versatile web application that empowers users with advanced capabilities in image processing and content generation [1, 2].

At its core, the project focuses on two pivotal functionalities: image colorization and text-to-image generation. The image colorization module utilizes deep learning models and computer vision techniques to transform grayscale images into vibrant, realistic color representations. On the other hand, the text-to-image generation module leverages state-of-the-art NLP models to generate visually compelling images directly from textual descriptions, fostering creativity and simplifying content creation processes.

By integrating these modules into a user-friendly Flask-based web application, our goal is to provide a seamless platform where users, from digital artists and marketers to educators and enthusiasts, can effortlessly enhance images and generate personalized visual content. This project not only

***Author for Correspondence**
Sandeep Patil
E-mail: skpatil_skncoe@sinhgad.edu

[1]Student, Department of Electronics & Telecommunication, Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India
[2]Assistant Professor, Department of Electronics & Telecommunication, Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

Received Date: July 03, 2024
Accepted Date: July 20, 2024
Published Date: July 30, 2024

Citation: Aashutosh Meshram, Yogini Kapse, Shreedhar Mhatre, Sandeep Patil. Text to Image Using Machine Learning. Journal of Image Processing & Pattern Recognition Progress. 2024; 11(2): 35–41p.

showcases the practical applications of AI in enhancing user interaction and content creation but also underscores its potential to democratize complex tasks traditionally reserved for experts.

Through this initiative, we aim to demonstrate how advanced AI technologies can reshape digital content creation, making it more accessible, intuitive, and impactful for diverse user needs and application.

MOTIVATION

The motivation behind this project stems from the growing intersection of artificial intelligence. The primary motivation is to empower users with intuitive digital content creation, and user experience enhancement. In the modern digital era, visual content holds significant importance in communication, education, and entertainment. Our project seeks to democratize this process by harnessing advanced computer vision and natural language processing techniques, thereby expanding accessibility to a wider audience tools [3] that simplify complex tasks like image colorization and text-to-image generation. Many users, including artists, educators, marketers, and content creators, often face challenges in transforming grayscale images into vibrant color representations or translating textual descriptions into visual content. Our project aims to address these challenges through the implementation of an automated and user-friendly platform.

Moreover, the project addresses the technological advancement in AI, particularly in deep learning models for image processing and NLP. By implementing state-of-the-art models and frameworks such as OpenCV, Hugging Face Transformers, and Flask, we aim to showcase the practical applications of these technologies in real-world scenarios. This not only demonstrates our technical proficiency but also underscores the potential for innovation in AI-driven applications [4, 5].

Additionally, the project's motivation extends to fostering creativity and exploration. By enabling users to experiment with different color schemes for grayscale images and generate images based on textual prompts, we encourage creative expression and experimentation. This aligns with broader trends in AI democratization, where advancements in machine learning are increasingly accessible for creative endeavors beyond traditional technical domains. Overall, the motivation behind this project is rooted in enhancing user capabilities through AI-driven automation, promoting creativity through simplified tools, and demonstrating the transformative potential of advanced technologies in reshaping digital content creation and user interaction paradigm [6, 7].

LITERATURE SURVEY

Text-to-image generation using machine learning explores how algorithms can create visual representations from textual descriptions, leveraging advancements in deep learning and neural networks to bridge the gap between natural language and visual content as shown in Table 1.

IMPLEMENTATION

Methodology

To achieve the project's objectives, we employ a combination of agile development practices, advanced deep learning techniques, and robust software engineering methodologies. The project utilizes an iterative development cycle, refining and enhancing each phase based on user feedback and testing outcomes. This approach ensures flexibility and adaptability, allowing the project to evolve in response to changing requirements and technological advancements. Central to the project are the methodologies used in implementing the image colorization and text-to-image generation modules. For image colorization, we utilize convolutional neural networks (CNNs) and pre-trained deep learning models, specifically designed to transform grayscale images into realistic color images. The process involves fine-tuning the models using a diverse set of training images and integrating them with OpenCV for efficient image processing. In the text-to-image generation module, we leverage state-of-the-art natural language processing models, such as the Stable Diffusion pipeline from Hugging Face's Transformers library. This includes utilizing pretrained models that can interpret textual descriptions and produce

high-quality images that correspond to the provided prompts. Techniques like mixed precision training and GPU acceleration are employed to optimize performance and reduce computation time.

Table 1. Latest literature review.

S.N.	Title	Author	Year	Methodology
1	Multi-Semantic Fusion Generative Adversarial Network for Text-to-Image Generation [8]	Huang *et al.*	2023	Text-to-image synthesis consistently encounters two primary challenges: ensuring high image quality and aligning images effectively with textual descriptions. Current approaches typically rely on a single sentence for image synthesis, which struggles to capture sufficient semantic detail, often resulting in generated images that diverge significantly from the ground-truth. In our study, we introduce a novel Multi-Semantic Fusion Generative Adversarial Network. This model can amalgamate semantic information from multiple sentences while preserving their distinct nuances to produce precise images. Furthermore, we propose a multi-sentence joint discriminator to ensure that generated images faithfully correspond to all provided sentences. Our experiments on CUB and MSCOCO datasets illustrate the significant advantages of our approach.
2	Black and White Image Colorization using Deep Learning [9]	Raghavan *et al.*	2022	Computerized colorization of black and white images began in the 1970s, marking the start of a significant evolution in image processing. Since then, RGB images have become indispensable in various fields, providing additional information that grayscale images cannot. These areas include determining the chemical composition of materials, chromatographic spectroscopy, and studying plant biology. Our approach to image colorization leverages deep learning techniques, creating a fully automated model that produces high-quality images without human intervention. This system draws inspiration from recent advancements in generative models.
3	Image Generation from Random Noise using Generative Adversarial Networks [10]	Meshram *et al.*	2023	In this research paper, we examine a GAN model that takes random noise as input and produces realistic output images. The generator creates new images from this noise by adjusting its weights based on whether the discriminator can be deceived into classifying the generated images as authentic. We developed architectures for both the generator and discriminator, evaluating the model using two datasets: the Simpson faces dataset and the Abstract paintings dataset.
4	End-to-End Chinese Landscape Painting Creation Using Generative Adversarial Networks [11]	Xue	2021	We introduce Sketch-And-Paint GAN (SAPGAN), the first model capable of autonomously generating Chinese landscape paintings from start to finish, without relying on conditional input. SAPGAN comprises two GANs: SketchGAN for producing edge maps and PaintGAN for translating these edges into fully realized paintings. Our model is trained on a novel dataset of traditional Chinese landscape artworks previously unexplored in generative research. In a Visual Turing Test involving 242 participants, SAPGAN paintings were frequently mistaken for human-created art at a rate of 55%, surpassing artworks generated by standard GANs. This research establishes a foundation for the development of genuinely original machine-generated art.
5	Colorization of Black & White Videos & Photographs [12]	Nimala *et al.*	2022	When a grayscale image is given as input, this paper addresses the challenging problem of predicting the audio description of the image. This problem has been underexplored, so previous methods have been less effective. It may rely on significant user interaction or result in a dull color rendering. We suggest a fully automatic method that produces brightness and realistic coloring. We accept the subtle uncertainty of the problem by setting it as a division function and applying class re-equation to training time to increase color variation in effect. The system is used as a feed pass on CNN during testing and so on trained in over a million color photographs. We test our calculation utilizing a "colorization Turing test", requesting that human members pick between the variety picture of the delivered truth and the base.

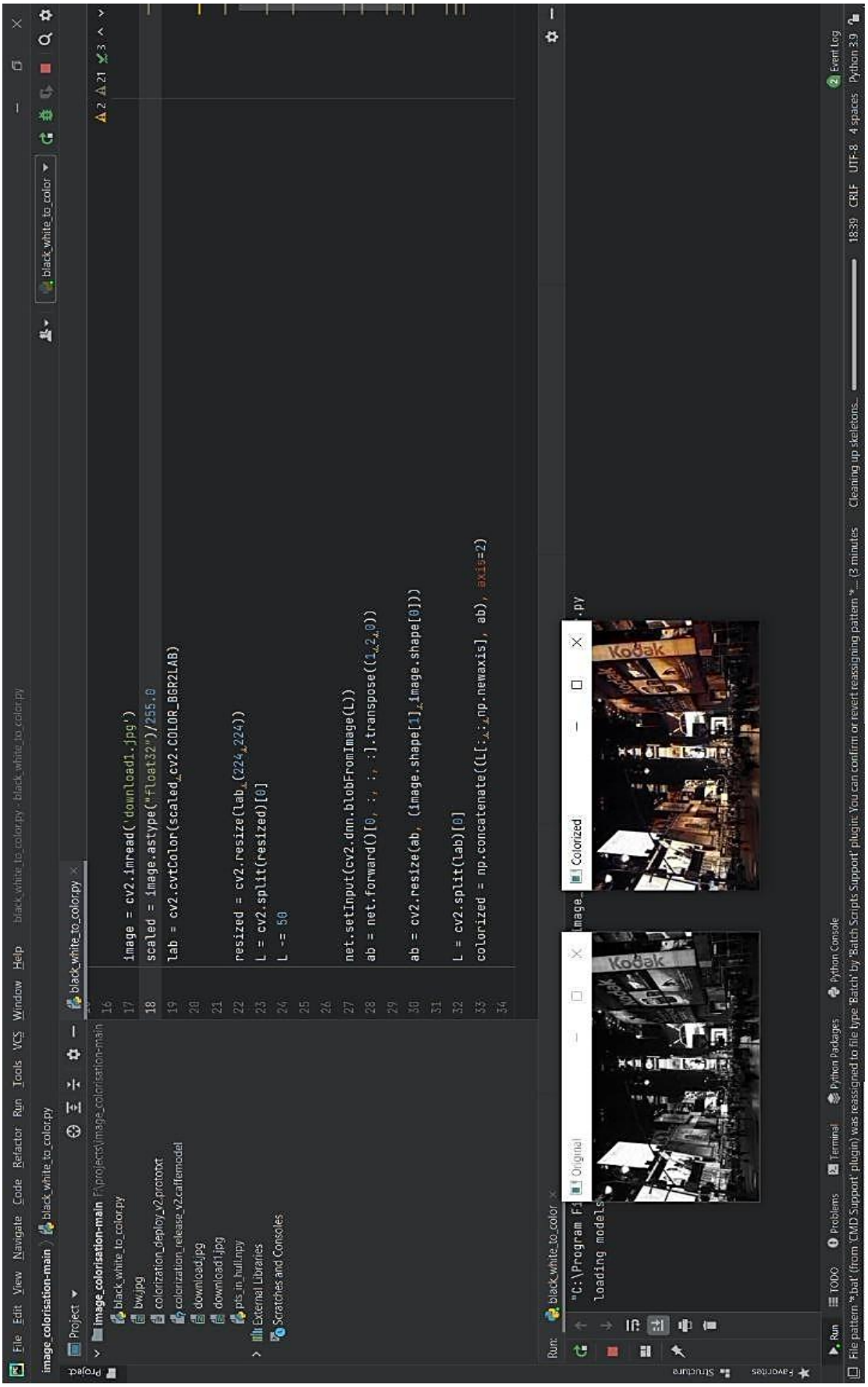

Figure 1. Output Image.

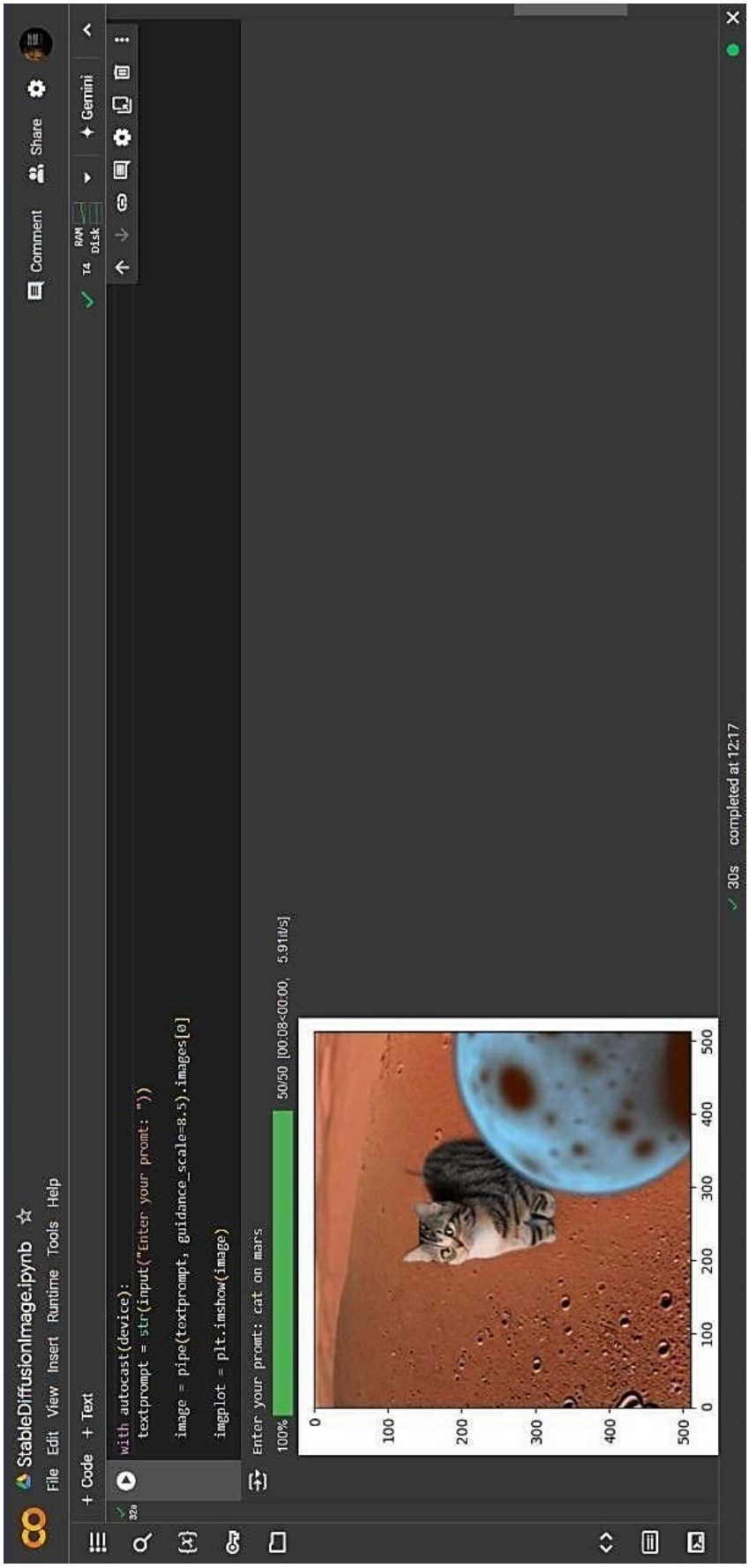

Figure 2. Output Image of proposed system.

Throughout the development process, we emphasize performance profiling and optimization to ensure that the system operates efficiently under various loads. Parallel processing techniques and resource management strategies are implemented to handle computationally intensive tasks, ensuring the application remains responsive and scalable.

RESULTS

The development of this web application, integrating advanced computer vision and natural language processing techniques, has demonstrated the powerful potential of AI-driven technologies in enhancing digital content creation and user interaction. Through the successful implementation of image colorization and text-to-image generation modules, we have equipped users with powerful tools to effortlessly convert grayscale images into vibrant color images and create visual content from textual descriptions as shown in Figures 1 and 2.

CONCLUSION

Throughout the project, rigorous methodologies were employed to ensure the reliability, efficiency, and usability of the application. From the initial stages of model selection and optimization to the iterative development cycle and comprehensive testing processes, each step was meticulously planned and executed to meet our goals and objectives. The deployment of the application in a user-friendly Flask-based web platform further underscores our commitment to delivering a seamless and intuitive user experience.

Verification and validation processes were crucial in upholding the integrity of the system. By continuously gathering user feedback and conducting extensive performance and usability tests, we were able to refine the application and address any issues promptly, ensuring that it meets the needs and expectations of its users.

Acknowledgment

It gives us great pleasure with the preliminary project report on ‘Text to Image using Machine Learning’. We would like to express our sincere gratitude and appreciation to our internal guide Prof. S. K. Patil for giving us all the help and guidance we needed. We are really grateful for kind support. Their valuable suggestions were very helpful. We are also grateful to Prof. Dr. S. K. Jagtap, Head of Electronics and Telecommunication Engineering Department, SKNCOE for her indispensable support and suggestions.

Future Scope

The future scope of this project includes expanding the functionality of the web application to support additional AI-driven features, such as real-time video colorization and more advanced text-to-image generation capabilities. We plan to integrate continuous learning mechanisms to improve model performance over time based on user interactions. Enhancing the application's scalability and incorporating multilingual support will broaden its accessibility and usability across diverse user bases. Furthermore, exploring partnerships with educational institutions and creative industries can open new avenues for collaboration and innovation.

REFERENCES

1. Upendra Hruthik S, Shruti Suman, Vishnu Sai S, Jaya Dharani. (Department of ECE, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India). Real-Time Face Mask Detection using OpenCV and Deep Learning. International Conference on Emerging Technologies: AI, IoT, and CPS for Science & Technology Applications, NITTTR Chandigarh, India. 2021 Sep 06–07.
2. Mishra S, Verma V, Akhtar N, Chaturvedi S, Perwej Y. An intelligent motion detection using OpenCV. Int J Sci Res Sci Eng Technol. 2022 Mar 5; 9(2): 51–63.
3. You Y, Gong S, Liu C. Adaptive moving object detection algorithm based on back ground subtraction and motion estimation. Int J Advancements in Computing Technology (IJACT). 2013; 5(6): 357–63.

4. Murshed M, Ramirez A, Chae O. Statistical background modeling: an edge segment based moving object detection approach. In 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance. 2010 Aug 29; 300–306.
5. Duraimurugan N, Chokkalingam SP. Real-Time Object Detection with Yolo. Int J Eng Adv Technol (IJEAT). 2019 Feb; 8(3S): 578–81.
6. Marengoni M, Stringhini D. High level computer vision using opencv. In 2011 IEEE 24th SIBGRAPI Conference on Graphics, Patterns, and Images Tutorials. 2011 Aug 28; 11–24.
7. Gürel C, Erden A. Face detection algorithm with facial feature extraction for face recognition system. In The 20th Int Conf on Mechatronics and Machine Vision in Practice. 2013; 1–7.
8. Huang P, Liu Y, Fu C, Zhao L. Multi-Semantic fusion generative adversarial network for text-to-image generation. In 2023 IEEE 8th International Conference on Big Data Analytics (ICBDA). 2023 Mar 3; 159–164.
9. Raghavan V, Sree SR, Kaladevi R, Hariharan S, Bhanuprasad A. Black and White Image Colorization using Deep Learning. In 2022 IEEE International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON). 2022 Dec 22; 2: 27–30.
10. Meshram K, Jadhav H, Narsale N, Raut R, Devkar A. Image Generation from Random Noise using Generative Adversarial Networks. In 2023 IEEE International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES). 2023 Apr 28; 381–385.
11. Xue A. End-to-end Chinese landscape painting creation using generative adversarial networks. In Proceedings of the IEEE/CVF Winter conference on applications of computer vision. 2021; 3863–3871.
12. Nimala K, Geetha G, Ponsam JG. Colorization of Black & White Videos & Photographs. In 2022 IEEE International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES). 2022 Jul 15; 1–6.

