IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Utilizing Generative AI in Genomics

Kumari Puja Sharma, Nikitha Vinod, Annapureddy Neha Reddy March 20, 2024

Abstract

The convergence of trustable artificial intelligence (AI) for critical applications and 6G security represents a critical juncture in the development of advanced technology systems. Trustable AI ensures that AI systems deployed in critical applications are transparent, ethical, reliable, and secure. These principles are essential in domains such as autonomous vehicles, healthcare, and financial systems, where human safety and infrastructure are at risk. Simultaneously, 6G security addresses the challenges of securing the next generation of wireless communication technology, which promises ultra-high data rates and massive connectivity. To safeguard 6G networks and systems, end-to-end encryption, robust authentication, and resilience are paramount. This paper explores the interconnectedness of trustable AI and 6G security, highlighting the need for trans- parency, ethical considerations, reliability, cybersecurity measures, and regulatory frameworks to ensure the trustworthiness and security of future technology ecosystems.
Keywords:generative neural networks, artificial human genomes, genomic diagnostics, precision medicine, machine learning, deep learning, genetic mechanisms

1 Introduction

The availability of genetic data has surged in recent years, propelled by advancements in sequenc- ing technologies and reduced costs. This wealth of data holds immense potential for various fields, from medicine to evolutionary studies. However, challenges persist, including the prohibitive costs of data acquisition and the under-representation of certain populations in genetic databases, particularly those of non-European ancestry. Moreover, concerns regarding data privacy further impede access to valuable genetic resources. In this context, generative models within the realm of machine learning offer a promising solution. Generative models, utilized in unsupervised machine learning, are adept at uncovering intrinsic properties of data and generating new data points based on these properties. Despite their widespread application in diverse domains, their potential in genetics, particularly in population genetics and genome-wide association studies (GWAS)[1], remains largely untapped.

Generative models, such as generative adversarial networks (GANs) and restricted Boltzmann machines (RBMs), have demonstrated prowess in learning complex data distributions and generating synthetic data across various domains. GANs, consisting of a generator and a discriminator, engage in a zero-sum game during training to produce authentic-looking data instances. RBMs, on the other hand, learn probability distributions through input data, comprising visible and hidden layers. Despite being overshadowed by newer approaches like GANs, RBMs have proven effective for diverse tasks, including image and document classification. In this study, we propose to harness the capabilities of GANs and RBMs to create Artificial Genomes (AGs) that faithfully mimic real genomes and capture population structures and other salient features. By exploring the potential of AGs as proxies for private datasets and addressing various genomic tasks such as imputation and selection scans, we aim to leverage generative AI to overcome the challenges posed by limited access to genetic data and privacy concerns, ultimately advancing genetic research and applications.

2 Literature Review

The article ”Creating Artificial Human Genomes Using Generative Neural Networks 2” by Tallec et al[1]. in Neural Networks extends their prior research on employing generative neural networks to fab- ricate synthetic human genomes. This study likely delves deeper into methodological advancements and assesses the efficacy of generative models in replicating complex genetic structures. Situated within the context of AI applications in genomics, the research contributes to addressing data scarcity issues and improving genetic analyses. With a multidisciplinary authorship, the study underscores the significance of synthetic genome generation in diverse fields such as population genetics and personal- ized medicine. Published in Neural Networks, the work provides insights into the technical aspects of synthetic genome creation, catering to the interests of the machine learning community.

In the realm of genetic research, Yelmen et al. (2021)[2] present a pioneering approach in ”Creating artificial human genomes using generative neural networks,” leveraging the power of generative models to produce synthetic genetic data. Their work, published in PLoS Genetics, showcases the potential of generative neural networks in mimicking real genomic characteristics, offering novel avenues for genetic analysis and overcoming data scarcity issues. Building upon this foundation, Yelmen and Jay (2023)[3] provide a comprehensive overview of ”Deep generative models in functional and evolutionary genomics” in the Annual Review of Biomedical Data Science. By elucidating the various applications of deep generative models in genomics, they shed light on the transformative impact of these techniques in advancing our understanding of functional genomics and evolutionary processes.

In the domain of clinical diagnostics, Dias and Torkamani (2019)[4] explore the intersection of artificial intelligence and genomic diagnostics in their article published in Genome Medicine. Their investigation delves into the integration of AI technologies in clinical settings, highlighting the potential for enhanced diagnostic accuracy and personalized treatment approaches. Furthermore, Chen et al. (2020)[5] present an innovative strategy for ”Population-scale genomic data augmentation based on conditional generative adversarial networks,” offering a solution to the challenges of limited genetic data availability. Their research, published in the Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, demonstrates the utility of generative adversarial networks in augmenting genomic datasets and facilitating population-scale analyses.

In the evolving landscape of digital pathology, Waqas et al. (2023)[6] discuss the transformative potential of ”Generative artificial intelligence and foundation models” in revolutionizing digital pathol- ogy practices. Published in Laboratory Investigation, their work highlights the role of generative AI and foundation models in enhancing the precision and scalability of pathology analyses, ultimately paving the way for more efficient and accurate diagnostic workflows. These diverse studies collectively underscore the broad-ranging applications of generative models and artificial intelligence across genetic research, clinical diagnostics, and digital pathology, signifying their growing importance in shaping the future of biomedical sciences.

3 Methodology

The methodology for creating artificial human genomes using generative neural networks (GNNs) is multifaceted and draws upon insights from various seminal works in the fields of genomics, artificial intelligence (AI), and computational biology. Building upon the foundational research by Yelmen et al. (2021)[2], which introduces the concept of using deep generative models for generating synthetic genomic data, our approach primarily revolves around the utilization of state-of-the-art generative adversarial networks (GANs) and restricted Boltzmann machines (RBMs). These deep generative models enable the creation of artificial genomes by learning complex data distributions from existing genomic datasets, thereby facilitating the synthesis of new instances that closely resemble real genomic sequences.

Incorporating the advancements highlighted in Dias and Torkamani (2019)[4], our methodology en- sures the integration of AI techniques into genomic diagnostics, ensuring the reliability and accuracy of the generated genomes. Leveraging insights from Chen et al. (2020)[5], our approach further extends to population-scale genomic data augmentation, employing conditional GANs to enhance the diver- sity and representation of the generated genomes across diverse population groups. By synthesizing these diverse research directions, we aim to develop a robust framework for creating artificial human genomes that capture essential genomic features, including allele frequencies, population structures, and evolutionary dynamics.

Moreover, our methodology incorporates expertise from Tallec et al. (2023)[1], focusing on the development of advanced GNN architectures tailored specifically for generating artificial human genomes. This involves fine-tuning the parameters of GNNs to accurately capture the intricate features present in real genomic data, ensuring that the generated genomes exhibit biologically plausible characteristics.

Additionally, our approach aligns with the visionary perspective outlined in Waqas et al. (2023)[6], which emphasizes the transformative potential of generative AI and foundation models in revolution- izing digital pathology and genomic research. By synthesizing insights from these diverse sources, our methodology aims to push the boundaries of genomic data synthesis, paving the way for innovative applications in functional genomics, evolutionary biology, and personalized medicine.

4 Conclusion

The papers reviewed provide a comprehensive overview of the potential of generative neural networks (GNNs) in creating artificial human genomes and revolutionizing genomic diagnostics. While the theoretical framework and methodology outlined in these papers show promise, concrete empirical re- sults from experimental studies are necessary to validate the effectiveness and utility of this approach. The integration of GNNs into genomic research and diagnostics holds great potential for advancing our understanding of genetic mechanisms, improving disease diagnostics, and facilitating personalized medicine. However, further research and empirical validation are needed to fully harness the capabil- ities of GNNs in this domain. Additionally, collaborative efforts between researchers, clinicians, and industry partners will be crucial for translating these advancements into real-world applications that benefit patients and healthcare systems. Overall, the exploration of GNNs in genomics represents a promising avenue for innovation and discovery in the field of precision medicine.

References

[1] Francesco Montinaro Tallec, Cyril Furtlehner, Luca Pagani, and Flora Jay. Creating artificial human genomes using generative neural networks 2. *Neural Networks*, 2:3.

[2] Burak Yelmen, Aur´elien Decelle, Linda Ongaro, Davide Marnetto, Corentin Tallec, Francesco Montinaro, Cyril Furtlehner, Luca Pagani, and Flora Jay. Creating artificial human genomes using generative neural networks. *PLoS genetics*, 17(2):e1009303, 2021.

[3] Burak Yelmen and Flora Jay. An overview of deep generative models in functional and evolutionary genomics. *Annual Review of Biomedical Data Science*, 6:173–189, 2023.

[4] Raquel Dias and Ali Torkamani. Artificial intelligence in clinical and genomic diagnostics. *Genome medicine*, 11(1):70, 2019.

[5] Junjie Chen, Mohammad Erfan Mowlaei, and Xinghua Shi. Population-scale genomic data aug- mentation based on conditional generative adversarial networks. In *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pages 1–6, 2020.

[6] Asim Waqas, Marilyn M Bui, Eric F Glassy, Issam El Naqa, Piotr Borkowski, Andrew A Borkowski, and Ghulam Rasool. Revolutionizing digital pathology with the power of generative artificial intelligence and foundation models. *Laboratory Investigation*, page 100255, 2023.