Generative AI Framework for Expediting the Search for New Therapeutics using Generative Adversarial Networks

Meeradevi, Vineet Sadarangani, Darshan Bankapure, Hrishikesh Haritas, and Prathmesh Sayal

***Abstract*—Drug Discovery and Repurposing (DDR) is one of the methodologies to find new uses of existing drugs. The proposed study uses novel artificial intelligence (AI) and generative adversarial network (GAN) approaches to expedite DDR. The proposed framework uses the advantages of the GAN model to generate new drug-like molecules and analyze their potential applications; drug repurposing aims to compare the proposed novel method with the traditional approach. The framework acts as a key assistive tool for various experts from the medical industry in research which helps to accelerate discovery of new therapeutic selections. The proposed work also uses Graph Neural Network (GNN) architecture for molecular structure. The dataset is curated, which encompasses the chemical structures of existing drugs, their biological properties and associated diseases, obtained from public repositories DrugBank, PubChem, and Gene Expression Omnibus (GEO). The study detects drugs with psychoactive substances and repurposing includes generating molecules for specific disease targets. The model performance is evaluated for its fitness and performs analysis on generated molecules with a confidence level of 85% for COVID and 78% for cancer disease drugs.**

***Index Terms*—Drug discovery and repurposing, generative adversarial networks, artificial intelligence, graph neural networks, deep learning.**

I. Introduction

DRUG Discovery and Repurposing (DDR) is a technique used to analyze and identify molecules in existing drugs and determine whether they can be used for new treatments. DDR received a large positive response during the COVID-19 pandemic. The proposed study will help researchers in the medical industry to accelerate the discovery process and find solutions using existing drugs for new therapeutic care.

The study explores AI and GAN specifically utilizing Wasserstein Generative Adversarial Networks with Gradient Penalty (WGAN-GP) [1] for drug discovery and the pre-existing GDRnet model [2] for repurposing, which is a specialized deep learning model. The proposed study aims to expedite DDR using AI techniques and identify new therapeutic uses for different disease indicators. The proposed work utilizes WGAN-GP to generate new drug-like molecules and analyze their applications, which helps medical experts in their research. The GDRnet model is used in predicting new therapeutic uses for existing drugs and enhancing the repurposing process, while GAN is used for drug discovery. The dataset is curated to encompass the chemical structures of known drugs and their disease associations from public repositories such as DrugBank, PubChem, and Gene Expression Omnibus (GEO).

Discovering and developing novel drugs is one of the most time-consuming and resource-heavy tasks across science. Traditionally, developing a single drug can span more than a decade and requires billions of dollars in investment. However, a significant proportion of candidate molecules still fail during clinical trials [3]. Recent advances in Artificial Intelligence (AI), particularly in deep learning, have opened new avenues for tackling each of these challenges with greater efficiency and accuracy. In molecular generation, GANs have shown considerable promise in learning complex chemical distributions to produce novel, synthetically feasible compounds. Similarly, Graph Neural Networks (GNNs) have proven effective in modeling the relational structure of drug-disease and drug-target interaction networks, making them well suited for repurposing predictions. However, existing studies largely address drug discovery and drug repurposing in isolation, treating them as independent problems rather than complementary stages of a unified workflow. We propose an integrated AI-based framework that combines drug discovery and drug repurposing within an end-to-end pipeline. Specifically, this framework invokes WGAN-GP for generating potentially therapeutic drug-like molecules, and GDRnet for predicting new therapeutic associations for existing drugs by leveraging relational properties with structurally similar drugs. The dataset utilized is produced by curating several public repositories DrugBank, PubChem, and GEO which include chemical structures, biological activity profiles, and disease associations. Additionally, the framework incorporates a drug detection component to identify psychoactive substances within chemical datasets, providing a more comprehensive analysis.

The key contributions of this article are as follows:

- The study offers a practical workflow that assists researchers in the pharmaceutical industry in fast-tracking

Manuscript received April 15, 2026; revised July xx, 20xx. Date of publication August xx, 20xx. Date of current version August xx, 20xx. This paper was presented in part at the Conference on Software, Telecommunications and Computer Networks (SoftCOM) 2019.

Meeradevi is with the Department of Artificial Intelligence and Machine Learning, Ramaiah Institute of Technology, Bengaluru, India (e-mail: meera_ak@msrit.edu).

Vineet Sadarangani, Darshan Bankapure, and Hrishikesh Haritas are with the Department of Artificial Intelligence and Machine Learning, Ramaiah Institute of Technology, Bengaluru, India (e-mail: vineet.sadarangani@gmail.com; bankapuredarshan@gmail.com; hrishikeshkh@gmail.com).

Prathmesh Sayal is with the Department of Computer Science in Cyber Security, Ramaiah Institute of Technology, Bengaluru, India (e-mail: 1ms23cy049@msrit.edu).

Digital Object Identifier 10.24138/jcomss.v1XiY.ZZZ

nascent drug discovery and the repurposing of existing drugs for treatment of rare diseases.
- Preprocessed drug data is leveraged to aid DDR, and GAN is used to generate new drug-like molecules.
- Algorithm implementation for detection of psychoactive substances.

The article is organized as follows. Section II presents a literature survey on algorithms and techniques used in drug discovery. Section III covers materials and methods, focusing on architecture, data metrics, preprocessing, and proposed algorithms. Section IV presents results and discussion including performance analysis and visualization, and Section V concludes the paper.

II. Literature Survey

In [4] the authors provide a comprehensive overview of how artificial intelligence techniques are being applied in drug repurposing to accelerate the identification of new therapeutic indications for existing drugs. The study reviews machine learning, deep learning, and network-based approaches used to analyze large-scale biomedical data such as genomics, proteomics, chemical structures, and electronic health records. The paper highlights how AI models improve prediction accuracy, reduce experimental costs, and shorten drug development timelines. The authors conclude that AI-driven drug repurposing has significant potential to transform modern drug discovery, especially in addressing complex diseases and improving personalized medicine.

In [5] the research focuses on existing approved drugs and finding new therapeutic uses for them, which is the main purpose of DDR. Repurposing the same drug is less time-consuming, less risky, and cheaper, and has shown a higher success rate. The availability of computational methods and datasets increased interest in this research during the COVID-19 pandemic. The study concludes that drug repurposing holds strong potential for accelerating new therapeutic development, mainly for rare and neglected diseases.

In [6] the paper applies a DL approach for drug repurposing using a GNN model on a biomedical dataset. The model predicts unknown drug-disease associations. Using GNN, the authors develop the model BEHOR, an extension of their earlier work REDIRECTION built using the DISNET biomedical knowledge graph. The study achieves accuracy of 0.96 AUROC and 0.95 AUPRC.

In [3] the authors highlight the importance of DDR. Identifying and analyzing new clinical uses for existing drugs can help speed up the drug discovery process and address global health challenges by using well-established safety profiles. Using multidisciplinary approaches in this field expands the opportunities in DDR. The authors discuss challenges such as target identification, ethical considerations, and regulatory bodies involved. Advances in computational methods help accelerate therapeutic development.

In [7] the paper addresses the diverse needs of DDR and offers key insights for future initiatives globally. The authors conducted an extensive literature survey for drug repurposing to identify cost-effective and timely strategies in new drug discovery. The work examines how existing drugs can significantly reduce time and cost compared to traditional approaches. The review covers repurposing approaches such as *in silico* screening, network pharmacology, and IPR rights. The authors conclude that there is a growing demand for DDR in the medical industry despite limitations. The study uses computational algorithms such as QSAR, molecular docking, and ML models SVM and CNN, along with gene expression for signature matching using CMap, which gives predictive accuracy of 75 -80%.

In [8] the authors discuss the role of computational strategies in reshaping modern drug discovery, with a strong focus on *in silico* methods such as molecular docking, virtual screening, molecular dynamics simulations, and machine learning models. The study emphasizes how these techniques enable rapid evaluation of large chemical libraries and improve understanding of drug-target interactions. The paper also highlights the integration of AI with traditional computational chemistry to enhance prediction reliability and reduce experimental dependency. The authors conclude that computational approaches are becoming essential tools for accelerating drug development and improving efficiency in both novel drug discovery and drug repurposing.

In [9] the authors review various computational strategies in transforming drug discovery, integrating approaches such as structure-based and data-driven methods across the drug development pipeline. The authors use algorithms for molecular docking and simulations, QSAR models, and SVM, CNN, GNN for drug interactions. The study achieves 90% accuracy on ML models. Model validation remains challenging due to data bias, interpretability issues, and clinical testing requirements.

In [10] the authors outline how using existing drugs provides opportunities in identifying new clinical indicators with reduced cost and time without hindering performance. The study aims for robust clinical development planning with precise dose rationale. The study also aims to strengthen integrated evidence for repurposing outcomes and presents the UCL Repurposing Therapeutic Innovation Network as a collaboration bringing expertise from multiple disciplines.

In [11] the authors propose DrugRepo, a novel framework integrating genomic information with chemical structure to predict new drug associations. The study employs fusion methods comprising molecular features with genomic features such as chemical descriptors and similarity metrics for drug-gene and disease-gene interactions. ML models such as SVM and random forest achieve prediction accuracy of 85- 90% on ROC-AUC metrics. The study concludes that combining chemical and genomic features significantly improves model robustness.

In [12] the authors propose DDR as an optimal alternative to traditional methods of drug discovery. The study aims to identify new uses of existing drugs to reduce time and cost, increase success rate, and ensure patient safety with good accuracy.

In [13] the study focuses on developing a drug discovery iterative AI framework that helps in optimizing molecular structure. Drugs focusing on a single target may lose efficacy

due to adaptability issues. The aim is to develop multi-target therapeutics that are more optimal. Reinforcement learning (RL) makes dynamic decisions in molecular design, allowing models to maximize reward. In [14] the study highlights the use of AI in drug discovery. AI is revolutionizing healthcare and many industries by refining drug design, optimizing molecular structure, and tailoring therapeutics based on patient profiles.

In [15] the authors discuss DrugGPT, which generates protein sequences from small molecules. DrugGPT is a transformer-based model showing high performance for pharmaceutical research in drug discovery. Furthermore, DrugGen utilizes large language models and reinforcement learning to enhance drug discovery [16].

Generative AI (GAI) technologies in drug discovery are reducing cost and time by leveraging GAN, reinforcement learning, and autoencoders to generate new molecules with desired properties, reducing clinical trial failure rates, and helping researchers repurpose existing medications [17]. In [18], a survey shows how pharmaceutical companies are introducing new treatments and enhancing the DDR process using AI to bring more optimal and effective therapeutics to market.

However, several gaps remain. Most studies address drug discovery or drug repurposing in isolation, not treating them as complementary stages of the drug development pipeline. GNN-based repurposing models such as BEHOR and DrugRepo, while accurate, operate without generative components, meaning newly synthesized candidate molecules cannot directly feed into repurposing predictions. Similarly, GAN-based discovery frameworks do not account for the relational aspects of drug-disease mapping. The identification of psychoactive substances within the same pipeline is also not fully investigated in the literature. Finally, most works validate on a single disease, making cross-disease generalizability difficult to assess. These gaps motivate an integrated framework that unifies molecular generation, relational repurposing prediction, and drug detection under a common architecture, evaluated across multiple disease contexts.

III. Materials and Methods

The DDR process involves detecting drug candidates through computational and experimental methods. Model performance is evaluated using efficacy metrics and progressing clinical trials.

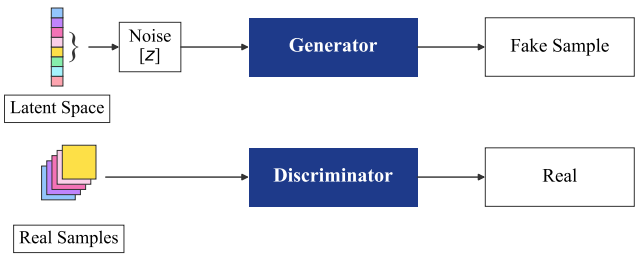

Fig. 1. Neural Networks in GAN Architecture.

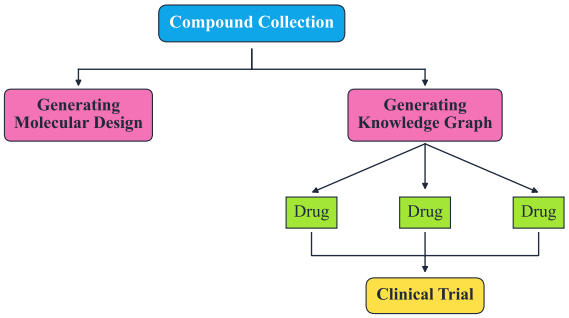

Fig. 2. Software Process for Drug Discovery and Repurposing.

In the proposed work, AI algorithms and bioinformatics tools are used in recognizing drug-protein interactions.

Leveraging GDRnet provides a powerful approach in finding new therapeutic uses for existing drugs, helping fast-track by avoiding many initial stages of drug development, interaction modeling, and clinical testing. The compound collection of the drug is taken to design molecule compositions, and a knowledge graph is generated in parallel using GNN for drug analysis. Clinical testing is then performed with various trials, as shown in Fig. 2.

This model is composed of an Encoder-Decoder that converts SMILES representations (molecule descriptions) into latent space vectors. These vectors are then used as real data in the training of a WGAN-GP network comprising a Generator and Discriminator, as shown in Fig. 1. The model uses a GNN to analyze interconnected data. The GAN-based approach has also been used for skin cancer to improve disease detection accuracy; models such as VGG16 and MobileNetV2 have shown 97% accuracy on malignant sample datasets [19].

The process involves loading pre-processed data, preparing input structures through functions like `get_disease_batches` and `get_drug_name_desc_dict`, and using the predict function to generate predictions. The results provide predictions on the likelihood of existing drugs being effective for treating specific diseases.

The database used to train and test the WGAN-GP contains filtered properties from a large set of molecular features. To incorporate larger molecules (up to 20 heavy atoms) into the training data, a subset of the ZINC rings dataset was utilized. SMILES data was downloaded from ZINC and filtered to include molecules with 10–20 heavy atoms containing only Carbon (C), Nitrogen (N), and Oxygen (O). The first 10,000 molecules from this set replaced the first 10,000 entries in the original training data (which consisted of smaller molecules from QM9 with up to 9 heavy atoms, corresponding to 10–20 total atoms including hydrogen).

Table II is used for training and testing the GDRnet model for Drug Repurposing. The first column contains the Compound ID, followed by the Drug Name, its approval Status, and finally its Molecular Weight (MW).

Table I is used to construct candidate molecular structures in the generator, which are then evaluated by the discriminator

TABLE I. Database for Drug Discovery

SMILES	pCHEMBL	MW (g/mol)	LogP
NC(=O)c1cccc(OC2CC3CCC(C2)N3C2(c3ccccc3)CC2)c1	4.62	362.47	3.85
O=C(Nc1cccc(NC(=O)c2cccc(N3CCOCC3)c2)c1)c1cccc(OC2CCN(CC3CC3)CC2)c1	9.3	554.69	5.28
CSCCC(NC(=O)C(Cc1ccccc1)NC(=O)CNC(=O)C(N)Cc1ccc(O)cc1)C(N)=O	6.68	572.69	−1.05
CCC(=O)NCCc1ccc2[nH]c3c(c2c1)CC1(O)C2Cc4ccc(O)c5c4C1(CCN2CC1CC1)C3O5	8.8	513.64	3.64
O=C1C2CCCC1(c1cccc(O)c1)CCN2CCc1ccccc1	5.42	335.45	3.7
Coc1ccc2c3c10C1C(=O)CCC4(O)C(C2)N(CC2CC2)CCC314	6.89	355.43	1.83

TABLE II. Database for Drug Repurposing

Compound ID	Drug Name	Status	MW (g/mol)
DB00001	Lepirudin	approved	6963.425
DB00002	Cetuximab	approved	145781.6
DB00003	Dornase alfa	approved	29253.9
DB00004	Denileukin diftitox	approved, investigational	57647.3
DB00005	Etanercept	approved, investigational	51234.9
DB00006	Bivalirudin	approved, investigational	2180.3
DB00007	Leuprolide	approved, investigational	1209.4
DB00008	Peginterferon alfa-2a	approved	60000
DB00009	Alteplase	approved	59042.3
DB00010	Sermorelin	approved, withdrawn	3357.882
DB00011	Interferon alfa-n1	approved, investigational	19241.1
DB00012	Darbepoetin alfa	approved, investigational	18396.1

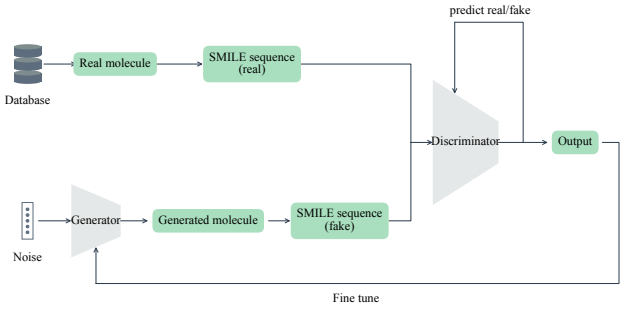

Fig. 3. Flowchart for Drug Discovery.

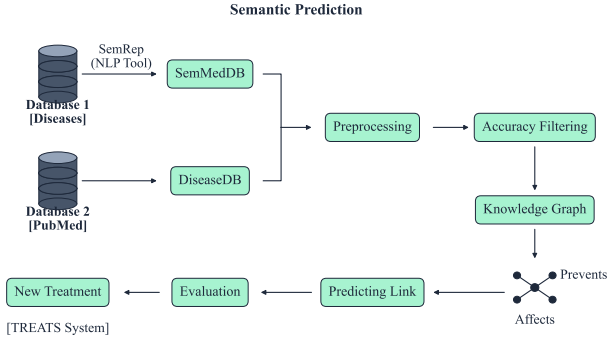

Fig. 4. Knowledge Graph Construction Workflow for Drug Repurposing.

to assess their structural feasibility and authenticity, as shown in Fig. 3. The data from the database should be verified and assured. Data compression techniques such as PCA and Huffman encoding can be used to secure molecular structures.

A. Data Curation and Preprocessing

The experiment uses two separate datasets gathered to meet the specific needs of the WGAN-GP and GDRnet machine learning models. Fig. 4 shows the data curation process, which involves extracting information from various public databases and filtering and standardizing the raw data to make it usable with deep learning systems. This process optionally incorporates literature-based relation extraction tools to augment the knowledge graph alongside structured databases.

The drug discovery module used molecular data originating from both the QM9 and ZINC databases. WGAN-GP training required filtering this raw data to retain molecules with 10–20 heavy atoms from ZINC, and small molecules with up to 9 heavy atoms (which correspond to 10–20 total atoms including hydrogen) from QM9, containing exclusively Carbon (C), Nitrogen (N), and Oxygen (O). The filtering process generated a training set of 100,000 molecules: 90,000 from QM9 and 10,000 from ZINC.

The drug repurposing module used a dataset from DrugBank, PubChem, and GEO to create a relational database of known drug-disease associations, including chemical structures, biological activities, and disease targets. The training of GDRnet required both physicochemical properties and biological interaction scores to be synchronized for optimal results.

B. Algorithms

Below is the explanation of the working of the GDRnet Model Architecture for Drug Repurposing.

The GDRnet algorithm uses a neural network module, initializes the decoder, computes interaction scores, and combines and decodes inputs. The data is loaded and preprocessed to extract detailed drug features. The proposed model is executed after feeding input features and adjacency matrices, helping in accurate model prediction with drug efficacy for specific diseases. The model outputs predictions in a structured format mapping drug names to predicted probabilities of efficacy.

Notation:

- G: Generator neural network mapping noise z to synthetic samples
- D: Discriminator (critic) neural network scoring sample authenticity
- θ_G, θ_D: Learnable parameters of generator and discriminator
- m: Batch size
- n_{critic}: Number of discriminator updates per generator update (typically 5)
- λ: Gradient penalty coefficient (typically 10)
- α: Learning rate
- β_1, β_2: Adam optimizer momentum parameters
- $\|\cdot\|_2$: Euclidean (L2) norm
- ∇_x: Gradient operator with respect to x

Algorithm 1 Training Algorithm for GAN with Gradient Penalty

Initialize generator G and discriminator D neural network parameters θ_G and θ_D
Initialize Adam optimizers: optimizer_G $\leftarrow$ Adam(θ_G, α, β_1, β_2) and optimizer_D $\leftarrow$ Adam(θ_D, α, β_1, β_2)
for epoch $= 1$ to T **do**
  for $i = 1$ to n_{critic} **do**
    Sample minibatch $\{x^{(1)}, \ldots, x^{(m)}\}$ from real data distribution P_{data}
    Sample minibatch $\{z^{(1)}, \ldots, z^{(m)}\}$ from noise prior $p(z)$, typically $z \sim \mathcal{N}(0, I)$
    Generate fake samples: $\tilde{x}^{(j)} \leftarrow G(z^{(j)})$ for $j = 1, \ldots, m$
    Compute discriminator scores on real data: $D_{\text{real}} \leftarrow \frac{1}{m}\sum_j D(x^{(j)})$
    Compute discriminator scores on fake data: $D_{\text{fake}} \leftarrow \frac{1}{m}\sum_j D(\tilde{x}^{(j)})$
    Compute Wasserstein loss: $\mathcal{L}_W \leftarrow D_{\text{real}} - D_{\text{fake}}$
    Sample random coefficients $\varepsilon^{(j)} \sim \text{Uniform}[0, 1]$ for $j = 1, \ldots, m$
    Compute interpolated samples: $\hat{x}^{(j)} \leftarrow \varepsilon^{(j)} x^{(j)} + (1 - \varepsilon^{(j)})\tilde{x}^{(j)}$
    Compute gradients of discriminator at interpolated points: $g^{(j)} \leftarrow \nabla_{\hat{x}} D(\hat{x}^{(j)})$
    Compute gradient penalty: $\mathcal{L}_{GP} \leftarrow \frac{1}{m}\sum_j (\|g^{(j)}\|_2 - 1)^2$
    Compute total discriminator loss: $\mathcal{L}_D \leftarrow -\mathcal{L}_W + \lambda \cdot \mathcal{L}_{GP}$
    Update discriminator: $\theta_D \leftarrow$ Adam($\theta_D, \nabla_{\theta_D}\mathcal{L}_D$, optimizer_D)
  end for
  Sample minibatch $\{z^{(1)}, \ldots, z^{(m)}\}$ from noise prior $p(z)$
  Generate fake samples: $\tilde{x}^{(j)} \leftarrow G(z^{(j)})$ for $j = 1, \ldots, m$
  Compute generator loss: $\mathcal{L}_G \leftarrow -\frac{1}{m}\sum_j D(\tilde{x}^{(j)})$
  Update generator: $\theta_G \leftarrow$ Adam($\theta_G, \nabla_{\theta_G}\mathcal{L}_G$, optimizer_G)
end for
return G, D

Algorithm 2 Drug Repurposing Prediction

Load node features X and node mapping M
Construct relational adjacency tensors A_r and A_{2r}
Define GDRnet with graph encoder and interaction decoder
Load pretrained weights W into GDRnet
Load drug metadata and standardize input tensors
// Prediction Procedure
Generate disease–drug batches from disease list D
for each batch **do**
  Compute disease–drug interaction scores using GDRnet
  Map predicted scores to corresponding drug identifiers
end for
Aggregate predictions across batches
Rank drugs by predicted probability
return top-29 ranked drugs

Notation

- X: Node feature matrix (diseases and drugs)
- M: Mapping between node indices and entity IDs
- A_r, A_{2r}: Relation-specific adjacency matrices
- GDRnet: Graph neural network for disease–drug interaction prediction
- W: Pretrained model parameters
- D: Set of query diseases
- Batch: Subset of disease–drug node pairs
- Score: Predicted interaction probability

C. Information about the Implementation of Modules

The Drug Discovery module combines several techniques: an autoencoder (specifically an Encoder-Decoder using RNNs), a special GAN called WGAN-GP, a property predictor, and an optimization step.

The autoencoder learns a new way to represent molecules (context vectors) that are easier for the system to work with. The encoder converts any molecule description (SMILES string) into this new format. This allows WGAN-GP to be trained on this new data format, enabling it to generate new molecules that fit the learned patterns. The proposed study not only generates molecules for drug discovery but also generates molecules with specific properties to increase optimization. The work uses the feedbackGAN model [20] to steer the molecules gradually to generate candidates with desired properties.

The proposed study uses a predictor to find and generate promising molecules selected based on many features such as binding affinity, surface area, etc. The selected molecules are fed back to the system to generate better candidates. The proposed work first learns how to represent molecules and then applies this knowledge base to generate new molecules with required properties for drug discovery. The key components are chemical properties of molecules and adjacency matrices to represent entity relationships and nodes for mapping unique identifiers, leveraging GPU acceleration.

IV. Results and Discussions

The generated molecules are analyzed in detail, including their optimization metrics, structural diversity, and key chemical properties, as shown in Fig. 5.

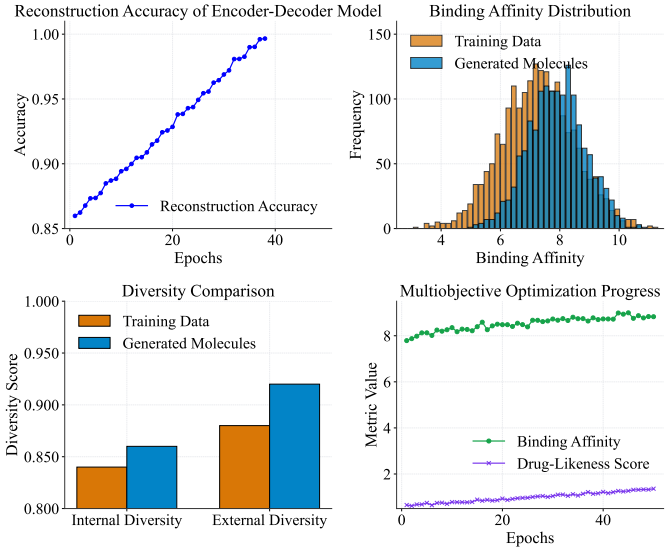

Fig. 5. Visualization of evaluation metrics for the generated molecular candidates, showing reconstruction accuracy, binding affinity distribution, diversity comparison, and multiobjective optimization progress.

Fig. 6 presents a sample output array from the WGAN-GP model. This array contains encoded values that can be decoded and mapped to the molecules present in the database.

The corresponding SMILES representation and the chemical structure can then be reconstructed.

A. Performance Analysis

Initially, a comprehensive dataset of drugs and their efficacy against various symptoms is compiled. For a given new disease characterized by three to four symptoms, the model identifies drugs known to alleviate these symptoms. The dataset is used to train the GANs, enabling them to learn the underlying feature representations of effective drugs. The features include molecular structures, physicochemical properties, and biological activity profiles. The model obtains an array of numbers after performing the prediction task, which is concentrated near the value of 7.42, as shown in Fig. 6. The presence of repeated values in this output array (e.g., 6.959321) is due to the generator producing duplicate SMILES structures during the decoding of latent vectors, which is a known characteristic of sequence-based molecular generators and slight mode collapse, resulting in identical predicted property values for those duplicate entries. Fig. 7 displays the properties of the predicted molecule for in-depth insights.

```
(array([7.4220266, 7.099023 , 7.040101 , 6.7202454,
    6.382882 , 6.959321 ,
       7.0241838, 6.959321 , 6.959321 , 6.959321 ,
            6.7202454,
       6.382882 , 7.040101 , 6.7202454, 7.099023 ,
            7.4220266, 6.9215317,
       6.959321 , 6.9002047, 6.959321 , 6.959321 ,
            6.959321 , 6.9002047,
       6.7202454, 6.382882 , 7.1046286, 7.040101 ,
            7.040101 , 7.4220266,
       7.040101 , 7.040101 , 7.1046286, 6.9215317,
            6.959321 , 6.9002047,
       6.9215317, 6.959321 , 7.0241838, 6.9215317,
            6.959321 , 7.0241838,
       6.959321 , 6.959321 , 6.959321 , 6.7202454,
            7.4220266,
       7.040101 , 6.9002047, 7.040101 , 7.040101 ,
            6.382882 , 6.7202454,
       7.040101 , 7.040101 , 7.1046286, 7.040101 ,
            7.1046286,
       6.9215317, 7.040101 , 7.040101 , 6.9002047,
            6.9215307, 6.9593215,
       7.0241838], dtype=float32), [0, 1, 2, 3, 4, 5, 6,
            7, 8, 9, 10, 11, ...
```

Fig. 6. Array of Predictions from WGAN-GP.

Fig. 8 represents the repurposed drugs along with the confidence levels of the predictions. It predicts the most effective drugs for the disease.

The histogram in Fig. 9 reveals a characteristic bell-shaped distribution centered around a predicted value of approximately 7.0, with the majority of generated molecules clustering between 6.8 and 7.2 on the property scale. This demonstrates important characteristics of the generated molecular ensemble.

First, the molecules exhibit homogeneity in terms of the specific property being measured, which could represent attributes such as lipophilicity, molecular weight, binding affinity, or synthetic accessibility scores depending on the property predictor employed. Second, the relatively low frequency of outliers at the distribution tails (values below 6.4 or above 7.4) demonstrates that the generator has learned appropriate constraints, avoiding the production of molecules with extreme or potentially problematic property values that would be unsuitable for pharmaceutical applications.

```
Molecular Weight: 554.6910000000003
LogP: 5.2810000000000006
Number of Hydrogen Donors: 2
Number of Hydrogen Acceptors: 6
Number of Rotatable Bonds: 9
Topological Polar Surface Area: 83.14
```

Fig. 7. Properties of the Predicted Molecule.

```
[('Riboflavin', 3.043088674545288),
 ('Oxygen', 2.903512954711914),
 ('Arginine', 2.73553690840224),
 ('Latanoprost', 2.5230348110198975),
 ('Pyrophosphoric acid', 2.519444465637207),
 ('Travoprost', 2.5017123222351074),
 ('Unoprostone', 2.49930047988916),
 ('Clonidine', 2.462891101837158),
 ('Verteporfin', 2.4051170349121094),
 ('Alprostadil', 2.3639168739318848),
 ('Sodium chloride', 2.3634204864501953),
 ('Atenolol', 2.3394763469696045),
 ('Choline', 2.328782558441162),
 ('Ramipril', 2.2718259968753967),
 ('Epinephrine', 2.261606112003326),
 ('Ranitidine', 2.259394407272339),
 ('Echothiophate', 2.251870632171631),
 ('Penicillamine', 2.2474560737609863),
 ('Captopril', 2.2362401485443115),
 ('Aflibercept', 2.236328601837158),
 ('Valsartan', 2.2300167083740234),
 ('Metoprolol', 2.227164411895752),
 ('Misoprostol', 2.212934970855713),
 ('Technetium Tc-99m pertechnetate', 2.2116193771362305),
 ('Levobunolol', 2.2069042656555176),
 ('Pilocarpine', 2.1915645599365234),
 ('Dipivefrine', 2.183328151702881),
 ('Carboprost tromethamine', 2.179192066192627),
 ('Calcium acetate', 2.1767497062683105)]
```

Fig. 8. Repurposed Drugs and Predicted Confidence Levels.

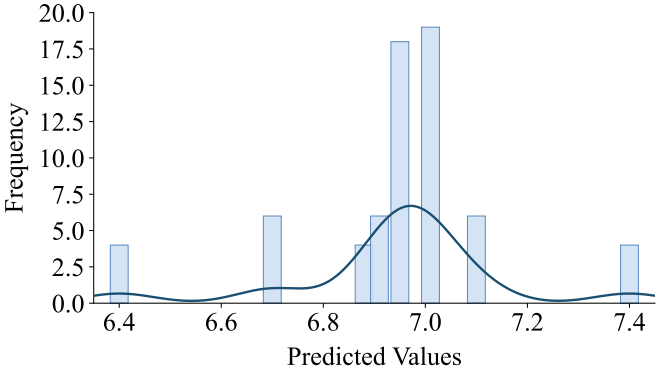

Fig. 9. Histogram of Frequency of the Predicted Values.

The scatter plot shown in Fig. 10 shows the dataset values spread broadly between 4 and 11. Predicted values show alignment in the range 4–11. The green highlighted circles indicate that the generated molecule value is consistent with the dataset.

The vertical spread of data points at any given index reveals the diversity of property values present in the training dataset, with some molecules exhibiting lower values around 4.0 and others reaching up to 11.0. The fact that the generated molecule's predicted value lies within the densest cluster suggests that the model has learned to prioritize the generation of molecules with mainstream pharmaceutical properties

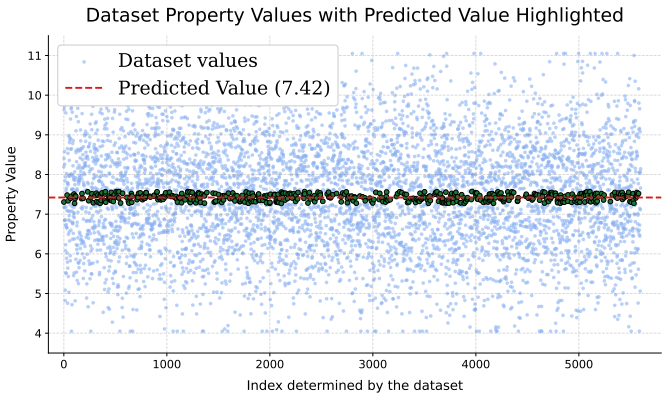

Fig. 10. Scatter Plot Comparing the Predicted Property Value.

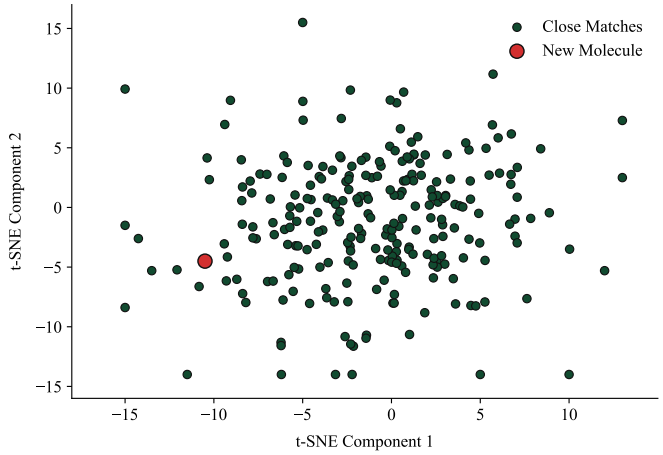

Fig. 11. t-SNE Plot Showing a Dense Cluster of Close Matches.

rather than edge cases. The horizontal index axis represents the sequential ordering of molecules in the dataset, and the relatively uniform distribution of property values across all indices confirms that the training data does not suffer from significant clustering or ordering biases.

Fig. 11 shows the t-distributed Stochastic Neighbor Embedding (t-SNE) visualization of the chemical space, revealing the structural relationships between known molecules and a newly generated candidate. The t-SNE algorithm reduces the high-dimensional molecular descriptor space into a two-dimensional representation while preserving local neighborhood structures, enabling visual assessment of molecular similarity patterns.

The plot demonstrates a dense central cluster of green data points representing close structural matches from the training dataset, with the distribution extending across both t-SNE components ranging from approximately -15 to $+15$.

The newly generated molecule, highlighted in red, is positioned within the dense cluster region at approximately $(-10, -5)$ in the t-SNE coordinate space. This placement is chemically significant as it indicates that the generated molecule shares substantial structural similarity with multiple known compounds in the training set.

Fig. 12 shows the average confidence level of the drug for various diseases: approximately 55% for diabetes, 85% for COVID, and 78% for cancer. The proposed model is efficient

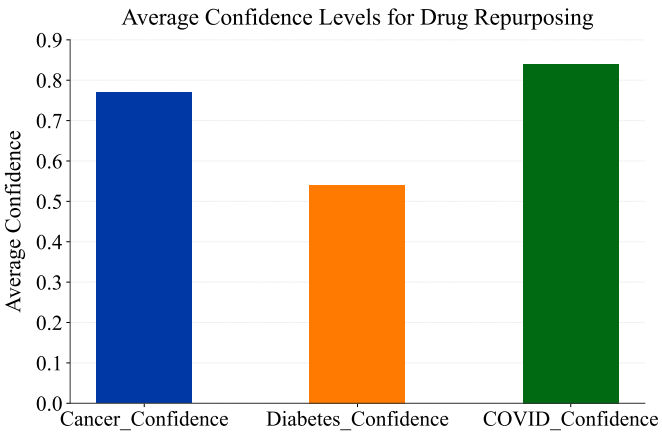

Fig. 12. Average Confidence Levels for Drug Repurposing.

in identifying the potential new therapeutic uses for existing drugs, with higher confidence levels for COVID and cancer drugs in predicting drug-disease associations.

The existing approach uses VGG16 or MobileNetV2 models, which show an accuracy of 97% in image-based disease detection and are limited to diagnostic tasks without using any drug molecules or drug properties. The proposed approach integrates GNN with WGAN-GP, which generates novel SMILES sequences in controlled physicochemical properties, bridging the gap between detection and molecular design. The dataset used in the proposed approach is a hybrid dataset with 100,000 molecules. The model has the capability to handle latent space vectors when compared to traditional SMILES-only models. The GDRnet in the proposed approach, used for drug repurposing, differentiates from existing methods by applying encoder-decoder architectures that compute interaction scores rather than just similarity matching.

V. Conclusion and Future Work

This study demonstrates that integrating generative and relational deep learning frameworks can significantly accelerate the drug discovery and repurposing pipeline. By combining WGAN-GP for molecular generation with GDRnet for relational drug-target predictions, the proposed framework successfully identifies novel drug-like candidates and aligns them with prospective disease targets. Experimental evaluation shows prediction confidence levels of 85% and 78% for COVID and cancer, respectively, verifying the utility of the hybrid approach. While computational complexity and biological validation remain key challenges, this integrated pipeline establishes a robust foundation for automated, multi-objective therapeutic design.

Future work will focus on incorporating additional clinical trial data, patient records, and genomic profiles from public and proprietary sources to enable personalized treatment modeling and tailor drug discovery efforts. Furthermore, more advanced data cleaning and augmentation methods can be implemented to ensure high-quality, comprehensive data capturing a wide range of biological and chemical properties. Additionally, the framework will be extended to handle multi-target therapeutics, and in vitro testing protocols will be

integrated to experimentally validate the bioactivity of the generated candidates.

ACKNOWLEDGMENT

The authors would like to thank the open-source repositories DrugBank, PubChem, and Gene Expression Omnibus for providing the data used in this study.

REFERENCES

[1] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved Training of Wasserstein GANs," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017, doi: 10.5555/3157382.3157640.

[2] C. Doshi and A. Chepuri, "A computational approach to drug repurposing using graph neural networks," *Computers in Biology and Medicine*, vol. 147, Art. no. 105728, 2022, doi: 10.1016/j.compbiomed.2022.105728.

[3] S. Pushpakom *et al.*, "Drug repurposing: progress, challenges and recommendations," *Nature Reviews Drug Discovery*, vol. 18, no. 1, pp. 41–58, 2019, doi: 10.1038/nrd.2018.168.

[4] Z. Wan, X. Sun, Y. Li, T. Chu, X. Hao, Y. Cao, and P. Zhang, "Applications of Artificial Intelligence in Drug Repurposing," *Advanced Science*, vol. 12, no. 14, Art. no. 2411325, 2025, doi: 10.1002/advs.202411325.

[5] N. Krishnamurthy, A. A. Grimshaw, S. A. Axson, S. H. Choe, and J. E. Miller, "Drug Repurposing: A Systematic Review on Root Causes, Barriers and Facilitators," *BMC Health Services Research*, vol. 22, no. 1, Art. no. 970, 2022, doi: 10.1186/s12913-022-08365-z.

[6] A. Ayuso-Muñoz, L. Prieto-Santamaría, E. Ugarte-Carro, E. Serrano, and A. Rodriguez-González, "Uncovering Hidden Therapeutic Indications Through Drug Repurposing with Graph Neural Networks and Heterogeneous Data," *Artificial Intelligence in Medicine*, vol. 145, Art. no. 102687, 2023, doi: 10.1016/j.artmed.2023.102687.

[7] A. H. Al Khzem and S. M. Wali, "Drug Repurposing as an Effective Drug Discovery Strategy: A Critical Review," *Drug Design, Development and Therapy*, vol. 19, pp. 12019–12034, 2025, doi: 10.2147/DDDT.S423156.

[8] M. Tutone and A. M. Almerico, "Computational Strategies Reshaping Modern Drug Discovery," *Molecules*, vol. 31, no. 2, Art. no. 200, 2026, doi: 10.3390/molecules31020200.

[9] S. Khan, J. Agnihotri, S. Patil, and N. Khan, "Drug Repurposing: A Futuristic Approach in Drug Discovery," *Journal of Pharmaceutical and Biological Sciences*, vol. 11, no. 1, pp. 66–69, 2023, doi: 10.9790/3008-1101016669.

[10] A. Recino, M. L. D. Rayner, J. L. Rohn, and O. Della Pasqua, "Therapeutic Innovation in Drug Repurposing: Challenges and Opportunities," *Drug Discovery Today*, vol. 30, no. 7, Art. no. 104390, 2025, doi: 10.1016/j.drudis.2024.104390.

[11] Y. Wang, J. Aldahdooh, Y. Hu, H. Yang, M. V. Koskela, J. Tang, and Z. Tanoli, "DrugRepo: A Novel Approach to Repurposing Drugs Based on Chemical and Genomic Features," *Scientific Reports*, vol. 12, no. 1, Art. no. 21116, 2022, doi: 10.1038/s41598-022-21116-0.

[12] M. C. De Rosa, R. Purohit, and A. T. García-Sosa, "Drug Repurposing: A Nexus of Innovation, Science, and Potential," *Scientific Reports*, vol. 13, no. 1, Art. no. 17887, 2023, doi: 10.1038/s41598-023-17887-x.

[13] S. I. Kang, J. H. Shin, B. M. Wu, and H. S. Choi, "Deep Generative AI for Multi-Target Therapeutic Design: Toward Self-Improving Drug Discovery Framework," *International Journal of Molecular Sciences*, vol. 26, no. 23, Art. no. 11443, 2025, doi: 10.3390/ijms262311443.

[14] S. Kant, Deepika, and S. Roy, "Artificial Intelligence in Drug Discovery and Development: Transforming Challenges into Opportunities," *Discover Pharmaceutical Science*, vol. 1, Art. no. 7, 2025, doi: 10.1007/s44337-024-00007-y.

[15] Q. Yin *et al.*, "DrugGPT: A generative modeling pipeline for molecule design," *bioRxiv*, 2023, doi: 10.1101/2023.06.29.546131.

[16] M. Sheikholeslami, N. Mazrouei, Y. Gheisari *et al.*, "DrugGen Enhances Drug Discovery with Large Language Models and Reinforcement Learning," *Scientific Reports*, vol. 15, Art. no. 13445, 2025, doi: 10.1038/s41598-025-13445-x.

[17] B. Potineni, "Generative AI in Drug Discovery: Accelerating the Search for New Therapeutics," *World Journal of Advanced Engineering Technology and Sciences*, vol. 15, no. 1, pp. 1784–1794, 2025, doi: 10.30574/wjaets.2025.15.1.0025.

[18] F. Rajaei, C. Minoccheri, E. Wittrup, R. C. Wilson, B. D. Athey, G. S. Omenn, and K. Najarian, "AI-Based Computational Methods in Early Drug Discovery and Post Market Drug Assessment: A Survey," *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 22, no. 1, pp. 97–115, 2025, doi: 10.1109/TCBB.2024.3355001.

[19] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017, doi: 10.1038/nature21056.

[20] A. Gupta and J. Zou, "Feedback GAN for DNA optimizes protein functions," *Nature Machine Intelligence*, vol. 1, pp. 105–111, 2019, doi: 10.1038/s42256-019-0017-4.

Meeradevi works as an Associate Professor in the Department of Artificial Intelligence and Machine Learning, Ramaiah Institute of Technology, Bangalore. She has 15 years of teaching and research experience. Her research specialization includes Artificial Intelligence, Machine Learning, Networks, Security, and Deep Learning.

Vineet Sadarangani is currently working as an apprentice at Trellix (erstwhile McAfee Enterprise and FireEye) as a Software Development Engineer in Test, and completed his undergraduate studies at Ramaiah Institute of Technology with distinction. He has accumulated over two years of industry experience through internships at Unisys, Samsung R&D, and Larsen & Toubro. His research interests lie in Artificial Intelligence, with a particular focus on Computer Vision and Deep Learning. He has been recognized with a Certificate of Excellence from Samsung and has won an innovation competition organized by Unisys.

Darshan Bankapure is a final-year student pursuing a Bachelor's degree in Artificial Intelligence and Machine Learning at Ramaiah Institute of Technology, Bangalore. His academic journey is centered around research and innovation. At OneNine AI, he developed no-code data transformation solutions. His work at Unisys involved full-stack development, while his research at Samsung R&D focused on a per-pixel deflicker approach for low-light video quality. At Larsen & Toubro Defense, he applied predictive analytics to propulsion systems. His publications appear in Taylor & Francis and IEEE journals.

Hrishikesh Haritas is a Project Associate at the Department of Computational and Data Sciences, IISc, Bangalore. He has previously worked on computer vision research projects with Samsung and explainable deep learning projects with Unisys. He holds a B.E. degree from Ramaiah Institute of Technology, India, and examines fintech graph analytics in large-scale payment systems.

Prathmesh Sayal is currently pursuing a Bachelor of Engineering in Computer Science with a specialization in Cybersecurity from Ramaiah Institute of Technology. He actively interned at ISRO, contributing to projects that leverage programming expertise in Python, TensorFlow, and C++. He has previously collaborated on cybersecurity and data-driven initiatives at Hewlett Packard Enterprise and is currently working as a summer intern at Fidelity Investments.