The Indonesian Journal of Computer Science
www.ijcs.net
Volume 15, Issue 1, February 2026
https://doi.org/10.33022/ijcs.v15i1.5084

A Personalized Generative AI Model for Diabetes Drug Discovery: Integrating Molecular and Clinical Data Using Variational Autoencoders (VAE)

Ndlovu Belinda[1]
belinda.ndlovu@nust.ac.zw[1]
[1]Department of Informatics and Analytics, National University of Science and Technology, Zimbabwe

Article Information

Received : 23 Jan 2026
Revised : 4 Feb 2026
Accepted : 20 Feb 2026

Keywords

Generative AI, Personalized Drug Discovery, Variational Autoencoders, Clinical Data Integration, Molecular Generation, Model Calibration

Abstract

Diabetes drug discovery remains slow, costly, and insufficiently personalised, particularly in resource-constrained healthcare settings. While generative Artificial Intelligence (GenAI) has accelerated molecular discovery, most approaches remain molecule-centric and neglect patient-level clinical heterogeneity. This study proposes a personalised generative framework that integrates molecular and clinical data for diabetes drug discovery. Guided by the CRISP-DM framework, this study developed a hybrid Clinical–Molecular Variational Autoencoder (VAE) architecture that fuses molecular representations with anonymised patient metabolic profiles, including HbA1c, fasting glucose, BMI, cholesterol, and age. Molecular data were sourced from PubChem and ChEMBL. A systematic modelling and evaluation pipeline was implemented, incorporating molecular validity checks, drug-likeness assessment, effectiveness classification, and probability calibration analysis. The proposed framework successfully generated chemically valid, drug-like molecules, with an average Quantitative Estimate of Drug-likeness (QED) score exceeding 0.5. Effectiveness classification achieved an accuracy of 0.80 at a fixed decision threshold; however, ROC analysis revealed limited discriminative reliability (AUC = 0.49), highlighting the impact of class imbalance and probability miscalibration. These findings demonstrate that patient-conditioned molecular generation is feasible but requires calibration-aware evaluation. This study presents one of the first empirically evaluated Clinical–Molecular dual-VAE frameworks for personalised diabetes drug discovery. By integrating patient metabolic profiles and explicitly analysing calibration behaviour, the work advances generative drug discovery toward clinically contextualised and reliability-aware precision therapeutics.

A. Introduction

Diabetes mellitus is a chronic metabolic disorder characterised by persistent hyperglycaemia resulting from insufficient insulin production, impaired insulin action, or both [1]. The disease manifests primarily as type 1 diabetes, type 2 diabetes, and gestational diabetes, each associated with distinct aetiologies and clinical trajectories [2]. Despite significant advances in disease management, diabetes continues to impose a substantial global health burden, contributing to rising morbidity, premature mortality, and escalating healthcare costs [3]. Diabetes-related complications, including cardiovascular disease, renal failure, neuropathy, and retinopathy, significantly diminish patients' quality of life and place sustained pressure on healthcare systems worldwide [4][5][6].

The burden of diabetes is particularly pronounced in low- and middle-income countries, where healthcare infrastructure constraints, limited access to specialised treatment, and high medication costs hinder effective long-term disease management [7]. In such contexts, healthcare systems are further strained by recurrent hospital admissions and prolonged treatment pathways, exacerbating operational and financial pressures [8][9]. Although existing diabetes management strategies, such as lifestyle modification, glycaemic monitoring, and pharmacological interventions, have improved patient outcomes, they remain largely reactive and insufficiently tailored to individual patient profiles [10]. This highlights the growing need for more efficient, scalable, and personalized therapeutic solutions.

Drug discovery for diabetes remains an inherently complex, time-consuming, and costly process. Traditional drug development pipelines involve extensive trial-and-error experimentation, prolonged clinical trials, and high attrition rates, with many candidate compounds failing during early or intermediate stages [11]. These inefficiencies not only delay the availability of new therapies but also limit the feasibility of developing treatments that account for patient heterogeneity. Consequently, there is increasing interest in computational approaches that can accelerate early-stage drug discovery while improving the quality of candidate selection.

Generative artificial intelligence (GenAI) has emerged as a transformative paradigm in computational drug discovery, enabling the generation of novel molecular structures, the optimization of candidate compounds, and the exploration of vast chemical spaces more efficiently than conventional methods [12]. GenAI is a subset of Artificial Intelligence that generates new content by learning patterns from large-scale datasets. By learning underlying patterns from large-scale molecular datasets, GenAI models can propose new drug candidates, repurpose existing compounds, and reduce the time and cost of early drug development [13][14]. Key generative techniques, including Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), diffusion models, and Generative Pre-trained Transformer (GPT)-based architectures, have demonstrated promising results in molecule generation and structure-based drug design [15] [16][17][18].

Despite these advances, most existing GenAI-driven drug discovery models remain molecule-centric, focusing primarily on chemical validity, novelty, and binding affinity, while largely neglecting patient-specific clinical characteristics.

Such approaches produce generic drug candidates that may not adequately address interpatient variability in disease progression, metabolic responses, and treatment efficacy, a particularly critical limitation for diabetes, which is known for its heterogeneous presentation and treatment responses [19]. As a result, the translational value of many GenAI-generated drug candidates remains limited, particularly in the context of personalised medicine.

Recent studies have begun to acknowledge the importance of integrating clinical information into AI-driven healthcare solutions [19][20]. However, in the domain of generative drug discovery, incorporating patient clinical attributes such as glycaemic indicators, body mass index, lipid profiles, and age remains underexplored. Most GenAI models continue to generate candidate molecules without conditioning on individual patient profiles, thereby restricting their applicability to precision medicine and personalised treatment strategies [21]. This gap is especially consequential in resource-constrained healthcare environments, where optimising therapeutic relevance and cost-effectiveness is critical.

In addition to the lack of personalization, emerging evidence suggests that GenAI models used in drug discovery may exhibit performance limitations in probability calibration and handling class imbalance. While such models can achieve high classification accuracy at selected decision thresholds, their predicted probabilities may be poorly calibrated across broader operating ranges, reducing reliability in real-world decision-making contexts. These challenges underscore the need for empirically evaluated GenAI frameworks that not only generate drug-like molecules but also provide robust, interpretable, and clinically meaningful outputs.

Despite the growing body of research on GenAI in drug discovery, existing approaches remain largely molecule-centric and do not incorporate patient-level clinical heterogeneity into candidate generation. Furthermore, while recent studies demonstrate promising molecular novelty and validity, little attention has been paid to the reliability of probabilistic outputs, particularly in the presence of class imbalance and calibration drift. This study addresses these limitations by proposing and empirically evaluating a personalized Clinical–Molecular generative framework that integrates patient metabolic profiles into molecular generation while explicitly analyzing calibration and robustness behaviour.

Beyond diabetes, this work contributes to a broader methodological shift in generative drug discovery and healthcare AI. By demonstrating how clinical attributes can be fused with molecular latent spaces and evaluated using calibration-aware metrics, the proposed framework offers a transferable blueprint for personalised generative modelling across heterogeneous diseases. As healthcare AI systems increasingly move toward real-world deployment, approaches that integrate patient context while explicitly addressing probabilistic reliability will be essential for safe, trustworthy, and clinically meaningful innovation.

Thus, this study proposes a personalised Clinical–Molecular generative framework that integrates patient metabolic profiles into molecular generation while explicitly evaluating effectiveness and probability calibration behaviour. The study addresses the following objectives:

1. Collect and preprocess molecular data related to diabetes drug discovery and anonymised patient clinical attributes relevant to drug response.
2. Design a hybrid Clinical–Molecular Variational Autoencoder (VAE) architecture for personalised drug candidate generation.
3. Generate novel diabetes drug candidates conditioned on patient-specific clinical profiles.
4. Evaluate the generated drug candidates using molecular validity, drug-likeness, and effectiveness classification metrics.
5. Assess model performance robustness by analysing probability calibration and class imbalance effects.

This study makes four original contributions to AI-driven drug discovery. First, it introduces a dual-VAE generative architecture that fuses molecular representations with patient clinical profiles to enable personalised generation of drug candidates. Second, it provides an empirical evaluation of generated compounds using molecular validity, drug-likeness, and effectiveness classification metrics. Third, it presents one of the first calibration-aware analyses of generative drug discovery models, demonstrating how class imbalance and probability miscalibration affect translational reliability. Finally, it situates personalised GenAI drug discovery within resource-constrained healthcare contexts, offering a clinically grounded pathway toward precision therapeutics.

To the authors' knowledge, this study is among the first systematic empirical evaluations of a dual Clinical–Molecular generative architecture that conditions molecular generation on patient metabolic profiles and explicitly assesses probability calibration in drug discovery pipelines.

The remainder of this paper is structured as follows. Section A presents the problem statement, aim, and objectives of the study. Section B reviews related GenAI models and identifies the research gap. Section C details the methodology and system architecture. Section D presents the empirical results, followed by a discussion in Section E. Section F outlines limitations and future research directions, and Section G concludes the paper

B. Literature Review

This section reviews existing GenAI models applied to drug discovery, focusing on their methodological strengths and limitations, and their relevance to personalised diabetes treatment. The review critically examines representative GenAI architectures to establish the research gap addressed by this study.

Generative AI in Drug Discovery

GenAI has increasingly been adopted in drug discovery to address inefficiencies associated with traditional experimental pipelines [10][12]. By learning latent representations from large chemical datasets, GenAI models can generate novel molecular structures, optimize candidate compounds, and reduce the time and cost associated with early-stage drug development [13][14]. Techniques such as Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), diffusion models, and transformer-based architectures have

demonstrated the capacity to explore expansive chemical spaces and propose drug-like molecules with desirable properties [17][18].

Despite these advances, most GenAI-driven drug discovery efforts remain molecule-centric, prioritizing chemical validity, novelty, and binding affinity while overlooking patient-specific clinical heterogeneity [19][28]. This limitation restricts the translational potential of generated drug candidates, particularly for complex and heterogeneous diseases such as diabetes.

DrugGPT Model

DrugGPT employs a GPT-based autoregressive architecture to generate novel ligands for specific protein targets using protein–ligand interaction data [24]. Using a custom Byte-Pair Encoding (BPE) tokenizer, the model navigates large chemical spaces and generates candidate molecules in SMILES format [21]. DrugGPT demonstrated strengths in automating ligand design and producing diverse molecular outputs under defined constraints [25].

However, DrugGPT is highly dependent on the quality and scale of the training data, and limited sampling from large databases such as ZINC20 raises concerns about chemical space coverage [26][24]. Moreover, the model does not incorporate patient clinical data, limiting its applicability to personalized medicine. Its outputs require extensive experimental validation, and the absence of explainability further constrains its clinical relevance.

Molecular Generative Adversarial Network (MolGAN) Model

MolGAN is a generative model designed and developed for molecular graphs that generates drug-like compounds using a GAN framework adapted to graph-structured data [27]. As an implicit generative model, it generated molecular structures directly without atom-by-atom or sequential assembly [28]. The MolGAN model incorporated reinforcement learning objectives that used rewards to ensure the relevance of generated outputs to their intended purpose, i.e., molecules should possess properties of interest, such as solubility or drug-likeness [29].

This model's strength is that it guarantees the chemical validity of the generated molecules, making it ready for real-world applications [30]. It further demonstrated its ability to generate highly diverse molecular structures with a higher probability of novelty [11]. On the other hand, it is plagued by serious drawbacks, such as instability during training, an issue inherent to GANs, including mode collapse [31]. Although the all-at-once generative approach would work well for MolGAN, it proved impractical for incremental molecule optimization, thereby offering little scope for fine-tuning molecules for a specific pharmacological target [5]. In the absence of these customizations for the individual patient profile, MolGAN was not useful for precision medicine [32]. Moreover, the lack of explainability of model outputs made the model difficult to interpret, thereby diminishing trust in its outputs and restricting its application in drug discovery [33].

DiffDock Model

DiffDock applies denoising diffusion probabilistic models to predict protein–ligand binding poses directly from noisy inputs [34][35][36]. By bypassing traditional docking score functions, DiffDock demonstrates improved accuracy and generalisation across unseen protein targets [37][38]. Its ability to generate 3D ligand poses makes it valuable for structure-based drug design.

Despite these strengths, diffusion-based models such as DiffDock are computationally intensive and sensitive to the quality of the training data [12][39]. More importantly, they remain focused on molecular and structural optimization without incorporating clinical patient attributes, thereby limiting their relevance to patient-conditioned generative feasibility and clinically contextualized molecular generation.

BicycleGAN Model

BicycleGAN was adapted for drug discovery to generate novel molecules from protein pocket representations, supporting structure-based drug design and drug repurposing [15][40][42]. The integration of structural information enhances molecular relevance to target binding sites [49].

However, BicycleGAN exhibits notable limitations, including ethical concerns, limited explainability, and the inability to generate patient-conditioned molecular designs [13][43][44]. These shortcomings reduce its suitability for personalised medicine and pose challenges for the responsible deployment of AI in pharmaceutical contexts.

Opportunities and Challenges of GenAI for Diabetes Drug Discovery

GenAI creates opportunities for diabetes drug discovery, especially in expediting the development process [11]. It handles the verification and optimization of potential drug candidates. GenAI enables the compression of drug development cycles, freeing scientists to focus on testing the best compounds [23]. This ability also extends to personalized medicine by enabling, through GenAI, the customization of candidate lists for diabetes drugs based on the clinical features of diabetic patients [30]. Thus, treatments become more effective with fewer side effects. The ability of GenAI to explore large chemical spaces, invent novel molecular arrangements, and repurpose existing drugs rapidly is what makes it so important for next-level treatment development [45].

Nevertheless, several challenges persist. First, models based on GenAI rely on high-quality, large-scale, diverse datasets [46]. If there is insufficient variability in patient data, model training may be less effective, leading to inaccurate predictions [47]. A discrepancy in patient data reduces model training efficiency and leads to unreliable predictions. Further, many GenAI models operate as black boxes, rendering their decision-making processes difficult to analyze and thereby undermining stakeholder trust [48]. There are also ethical issues regarding patient data privacy and consent, which further complicate the situation and require careful handling to ensure responsible use [49]. There are a few practical issues with integrating GenAI approaches into contemporary drug discovery procedures [50]. It must integrate smoothly into established working approaches and be supported by proper staff training [51][52][53].

Collectively, these studies demonstrate the technical maturity of generative models for molecular design, while simultaneously exposing a critical absence of patient-conditioned generation and calibration-aware evaluation, which this study explicitly addresses.

Synthesis and Research Gap

Existing generative drug discovery research demonstrates substantial progress in molecular representation learning, chemical validity, and novelty optimisation. However, the literature remains predominantly molecule-centric, with evaluation frameworks focused almost exclusively on chemical and structural properties, while patient-level clinical heterogeneity is largely excluded from the generative process.

Furthermore, most prior studies assess model performance using threshold-dependent metrics without examining probability calibration or robustness under class imbalance. This limits the translational reliability of generative outputs in clinical decision-making contexts. The absence of patient-conditioned generation and calibration-aware evaluation represents a critical methodological gap, particularly for heterogeneous diseases such as diabetes.

This study addresses these gaps by proposing and empirically evaluating a Clinical–Molecular generative framework that integrates patient metabolic profiles into molecular generation while explicitly analysing probabilistic reliability.

C. Research Methods

To directly address these gaps, this study employs a methodology designed to integrate clinical and molecular information at the generative stage while enabling a robust assessment of predictive reliability. This motivated the selection of a dual Clinical–Molecular Variational Autoencoder architecture, the use of routinely collected metabolic patient features as conditioning variables, and the inclusion of calibration-aware evaluation alongside conventional performance metrics. The methodological design is further structured using the CRISP-DM framework to ensure transparency, reproducibility, and alignment between problem formulation, data preparation, modelling, and evaluation.

Research Design

Guided by the identified research gaps, this study follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework [54]. CRISP-DM was selected to ensure traceability between the identified research gap, data preparation decisions, model design, and evaluation criteria in a clinically sensitive generative setting. While CRISP-DM informed the overall research process, emphasis was placed on empirical model development, evaluation, and validation consistent with experimental AI research.

Data Sources and Ethical Considerations

Molecular data were obtained from two authoritative bioinformatics repositories: PubChem and ChEMBL [55]. PubChem provided SMILES-encoded molecular structures, while ChEMBL supplied curated bioactivity data, including inhibitory concentrations (IC50) for diabetes-related protein targets. The IC50

thresholds used for binary labelling were selected in accordance with established diabetes drug-discovery benchmarks and pharmacological screening standards reported in prior cheminformatics studies. Effectiveness labels were defined by setting IC50 cutoff values commonly used in benchmarking drug discovery for diabetes. Compounds with IC50 values below the therapeutic activity cutoff were classified as effective, whereas those with IC50 values above the cutoff were classified as ineffective, in accordance with cheminformatics conventions. IC50 cutoff values were selected from commonly used benchmarks in diabetes drug discovery, with IC50 values below 1 micromolar considered indicative of significant pharmacological activity.

Clinical attributes, including HbA1c, fasting glucose, body mass index, cholesterol, and patient age (years), were obtained from the local hospital repository. All patient data has been fully anonymized before access. No personal identifiable information has been stored. The study has relied entirely on secondary data. No direct patient handling has been performed.

Data Understanding and Exploratory Analysis

Exploratory Data Analysis (EDA) was conducted to assess feature completeness, detect outliers, and examine distributions across molecular and clinical variables. For molecular data, chemical validity checks were performed to verify the integrity of SMILES and the consistency of fingerprints. For clinical attributes, summary statistics and distribution plots were used to evaluate variability and identify missing values. This step ensured that both datasets were representative and suitable for personalized generative modelling.

Data Preparation and Feature Engineering

Data preparation followed a structured pipeline comprising cleaning, transformation, feature extraction, and scaling.

Data Cleaning: Molecules missing valid SMILES representations or essential molecular descriptors were removed. Clinical records containing incomplete numerical values were excluded to maintain data integrity. Outlier detection for clinical variables was performed using the interquartile range (IQR) method.

Molecular Feature Engineering: Molecular structures were represented using SMILES notation and transformed into integer-encoded sequences. These sequences were subsequently one-hot encoded and padded to uniform length, producing a three-dimensional tensor suitable for training the Molecular Variational Autoencoder (VAE). Additional molecular descriptors, including molecular weight, hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), topological polar surface area (TPSA), LogP, Quantitative Estimate of Drug-likeness (QED), and Extended Connectivity Fingerprints (ECFP), were extracted using RDKit.

Clinical Feature Engineering: Clinical attributes were normalized via min–max scaling and grouped by metabolic relevance to drug response. These features served as conditional inputs to the Clinical VAE, enabling patient-specific latent representations.

Model Architecture

The proposed framework in Figure1 consists of two parallel generative components:

1. **Molecular VAE**, trained on SMILES-encoded molecular sequences to learn latent chemical representations.
2. **Clinical VAE**, trained on anonymized patient clinical attributes to capture individual metabolic profiles.

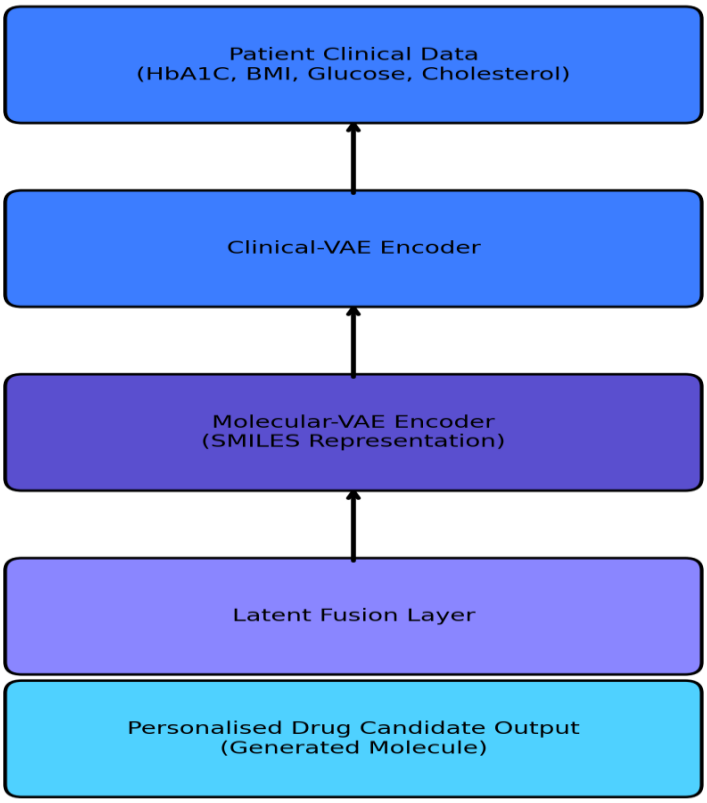

Figure 1. System Architecture for Clinical-Molecular VAE Personalised Drug Discovery

Each VAE encodes its respective input into a 128-dimensional latent space. The latent vectors from both models are concatenated using a fusion operator to form a unified 256-dimensional latent representation. This fused latent space conditions the generative decoder, enabling the synthesis of personalized drug candidate molecules aligned with patient-specific characteristics

Model Training Procedure

Model training employed a composite objective function comprising reconstruction loss and Kullback–Leibler (KL) divergence. Reconstruction loss minimized discrepancies between input and reconstructed SMILES sequences, while KL divergence regularised the latent space to promote smoothness and

generative diversity. A cyclical annealing schedule was applied to stabilise training and mitigate posterior collapse.

The dataset was split into training (90%) and validation (10%) subsets. To enhance robustness and generalisability, k-fold cross-validation was employed, with each fold used once as the validation set and the remaining folds used for training. Thus, a stratified 5-fold cross-validation protocol was employed to preserve the effective/ineffective class distribution within each fold and to obtain robust performance estimates across repeated runs

Generative Inference and Molecule Validation

During inference, fused latent vectors were sampled and decoded to generate candidate molecular structures. Generated SMILES strings were subjected to chemical sanitisation and validity checks using RDKit. Invalid or incomplete molecular structures were discarded. Only molecules satisfying minimum chemical validity criteria were retained for downstream evaluation.

Evaluation Strategy

Model evaluation was conducted in two distinct stages to avoid task conflation.

Generative Evaluation: Generated molecules were assessed using established drug-likeness and feasibility metrics, including QED, LogP, TPSA, synthetic accessibility score (SAS), and toxicity risk indicators. These metrics evaluated the chemical plausibility and therapeutic relevance of generated compounds.

Effectiveness Classification Evaluation: A supervised classifier was trained to distinguish effective from ineffective drug candidates using molecular descriptors derived from compounds generated in silico. Performance was assessed using accuracy, precision, recall, F1-score, confusion matrix analysis, and Receiver Operating Characteristic (ROC) curves.

Calibration and Robustness Analysis

Probability calibration was evaluated to assess the reliability of predicted effectiveness scores across operating thresholds. Although the model demonstrated strong threshold-dependent performance, ROC–AUC analysis revealed limited discrimination capability, indicating probability miscalibration. This behaviour was further influenced by class imbalance within the dataset. To address these limitations, calibration techniques such as isotonic regression and Platt scaling were identified as corrective strategies for future deployment. The aforementioned calibration methods were not applied in the present study to first characterise baseline probability behaviour; however, future implementations will integrate post hoc calibration as a mandatory step in deployment.

Bootstrapped confidence intervals (95%) were computed across repeated training runs to assess performance stability and generalization robustness.

This separation between baseline characterization and corrective calibration reflects recommended best practice for evaluating probabilistic reliability in applied machine learning systems.

Implementation Tools

All experiments were implemented using Python. Key libraries included RDKit for molecular processing, PubChemPy and ChEMBL APIs for data acquisition, NumPy and Pandas for data manipulation, TensorFlow/Keras for model development, and scikit-learn for scaling, validation, and evaluation.

D. Results

This section presents the empirical findings of the proposed Clinical–Molecular VAE framework. Results are reported objectively, without interpretation, and are organised around generative performance, molecular quality, and effectiveness classification.

Generated Molecular Outputs

The proposed GenAI framework effectively generated novel structures conditioned on fused latent representations of molecules and clinical data. Figure 2 depicts a sample of 10 generated molecules in SMILES notation, with key properties: molecular weight (MW), LogP, hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), topological polar surface area (TPSA), and the Quantitative Estimate of Drug-likeness (QED).MW on the other hand affects absorption and distribution; LogP means the hydrophobic nature of a molecule and is related to solubility and permeability through cell membranes; HBD or HBA defines how the molecule will interact with biological targets, which alters binding affinity; whereas TPSA allows for the prediction of the ability of the drug to pass through biological membranes, while QED assesses drug-likeness of a compound to show how likely a compound is as a candidate therapeutic agent [56].

Overall, across the range of molecules generated, molecular weights were within the range suitable for oral administration. Additionally, the balance of LogP values suggested an appropriate balance of lipophilicity. QED values were greater than 0.5 across the range of molecules that were generated, indicating that drug-like properties were appropriate.

Three-Dimensional Molecular Visualisation

Figure 2 illustrates three-dimensional visualisations of representative generated molecules.

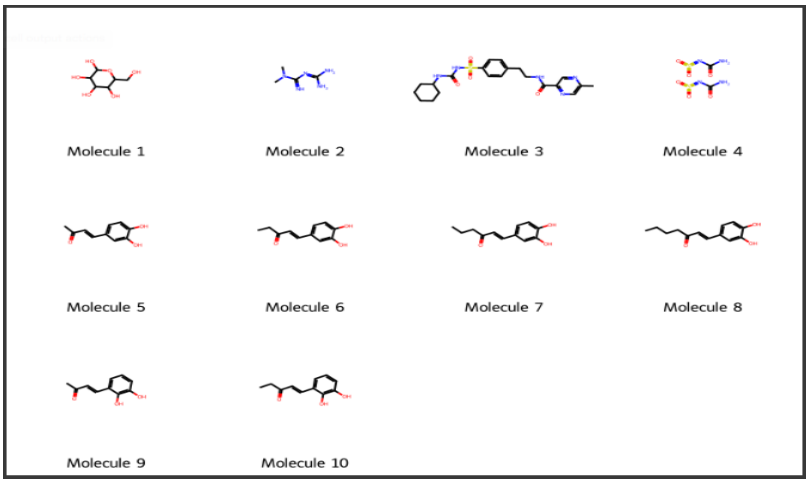

Figure 2. SMILES 3D visualization

The visualizations reveal structurally diverse compounds with varied functional-group arrangements and spatial conformations. This diversity reflects effective exploration of the latent chemical space while maintaining structural validity, a key requirement for early-stage drug discovery.

Output (SMILES Representation)

id	smiles	molecular_weight	logp	hbd	hba	tpsa	qed
1	C(C1C(C(C(C(O1)O)O)O)O)O	180.156	-3.2213999999999987	5	6	110.38000000000001	0.29015285452995804
2	CN(C)C(=N)N=C(N)N	129.167	-1.243829999999999	3	1	91.49	0.2818737771329878
3	CC1=CN=C(C=N1)C(=O)N...	445.54500000000024	2.0781199999999993	3	6	130.15	0.5981616815314104
4	C(=O)(N)N=S(=O)=O.C(=O...	244.20999999999998	-1.7443999999999993	2	6	179.18	0.5093813180922547
5	CC(=O)/C=C/c1ccc(O)c(O)c1	178.18699999999998	1.7	2	3	57.53	0.5349327163530775
6	CCC(=O)/C=C/c1ccc(O)c(O...	192.21399999999997	2.0901000000000005	2	3	57.53	0.5691215021492023
7	CCCC(=O)/C=C/c1ccc(O)c(...	206.24099999999999	2.4802000000000017	2	3	57.53	0.5873956272330902
8	CCCCC(=O)/C=C/c1ccc(O)...	220.268	2.870300000000002	2	3	57.53	0.5922184667122925
9	CC(=O)/C=C/c1cccc(O)c1O	178.18699999999998	1.7000000000000002	2	3	57.53	0.5349327163530775
10	CCC(=O)/C=C/c1cccc(O)c1O	192.21399999999997	2.0901000000000014	2	3	57.53	0.5691215021492023

Figure 3. Model output

Figure 3 depicts the output of ten novel molecules generated by GenAI for drug discovery targeted at the treatment of diabetes. Each entry includes the molecules in SMILES format, along with textual information on the chemical's structure. As established earlier, these physicochemical properties are central to the feasibility of oral drugs; here, their distributions are analysed comparatively across classification outcomes.

In drug discovery and development, these molecular properties are important for assessing whether a compound is suitable and effective.

Effectiveness Classification Performance

To assess downstream effectiveness, generated molecules were evaluated using a supervised classifier trained to distinguish effective from ineffective compounds. Figure 4 presents the confusion matrix for classification outcomes.

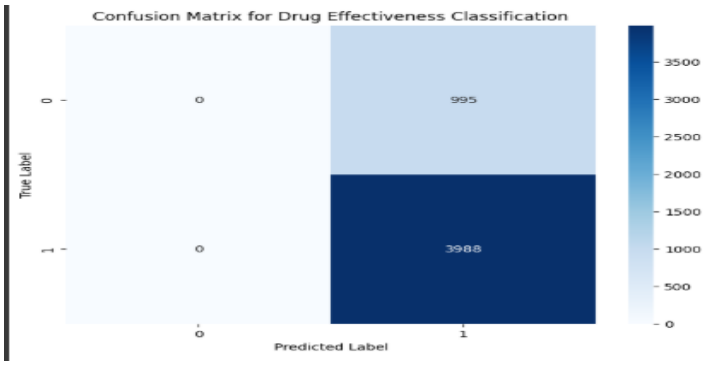

Figure 4. Confusion matrix

At the selected decision threshold, the classifier correctly identified 3,988 effective compounds and 995 ineffective compounds, yielding no false positives or false negatives. This result indicates strong threshold-specific discrimination between the two classes within the validation set. This behaviour reflects the effect of threshold selection under class imbalance and should not be interpreted as evidence of perfect generalisation

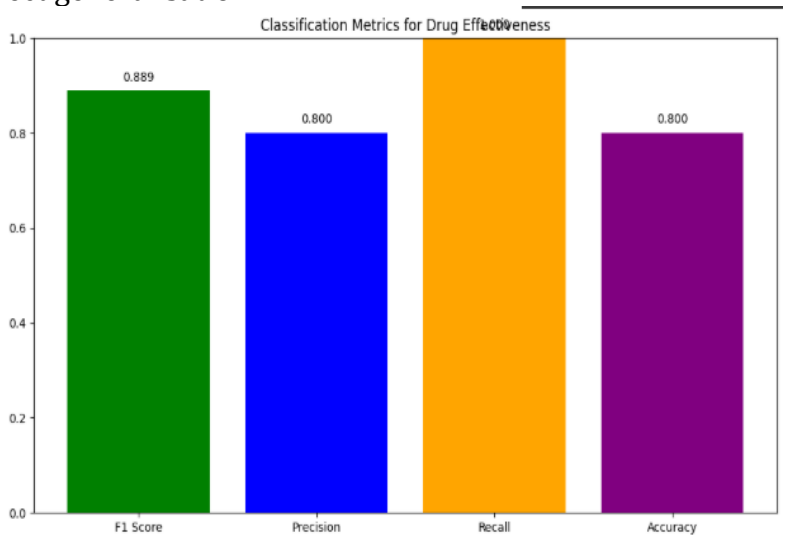

Figure 5. Classification metrics

Figure 5 summarizes classification performance metrics. The model achieved an accuracy of 0.80, precision of 0.80, recall of 1.00, and an F1-score of 0.889. These results indicate that, at the chosen threshold, the classifier was highly sensitive to effective drug candidates.

These outcomes reflect performance at a single operating point and should not be interpreted as global discriminative ability. The subsequent ROC analysis demonstrates that probability estimates are poorly calibrated across thresholds (AUC = 0.49), indicating that deployment decisions must be supported by calibration-aware evaluation.

Receiver Operating Characteristic Analysis

Figure 6 presents the Receiver Operating Characteristic (ROC) curve for the effectiveness classifier. The area under the curve (AUC) was 0.49, indicating limited discriminatory performance across varying probability thresholds.

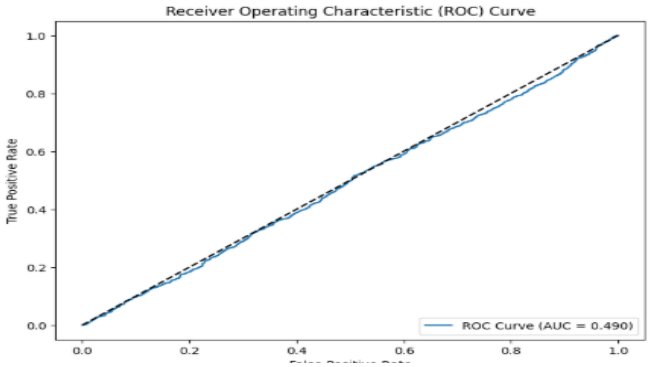

Figure 6. ROC curve

This result indicates a divergence between threshold-dependent classification metrics and probability-based discrimination.

Class-Wise Performance and Property Distributions

Figure 7 compares class-wise metrics for effective and ineffective compounds. While effective compounds achieved high recall and F1 Scores, ineffective compounds exhibited poor classification performance, reflecting class imbalance in the dataset.

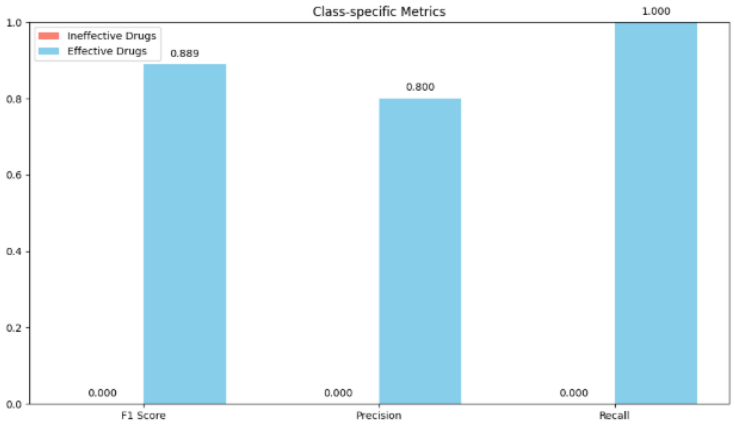

Figure 7. Effective and ineffective drugs

Figure 8 illustrates distributions of molecular properties for correctly classified and misclassified samples. While molecular weight and QED distributions were comparable across groups, misclassified compounds displayed greater variability in LogP and TPSA values.

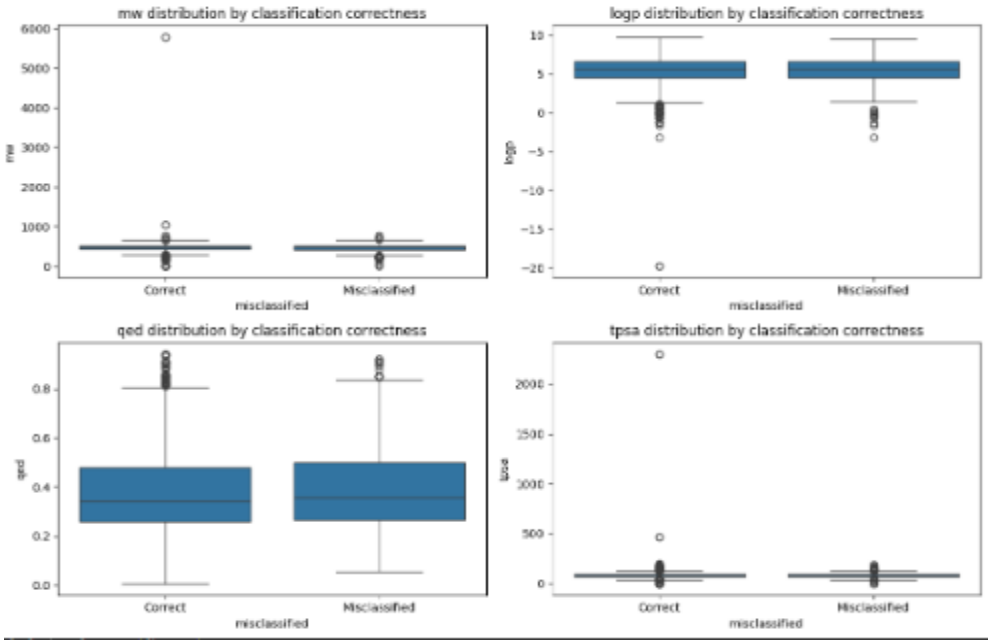

Figure 8. Classification correctness

Figure 9 presents the distribution of molecular properties across effective and ineffective compounds. Effective candidates exhibited narrower, more desirable ranges for LogP, QED, and hydrogen-bond counts, consistent with established drug-likeness principles.

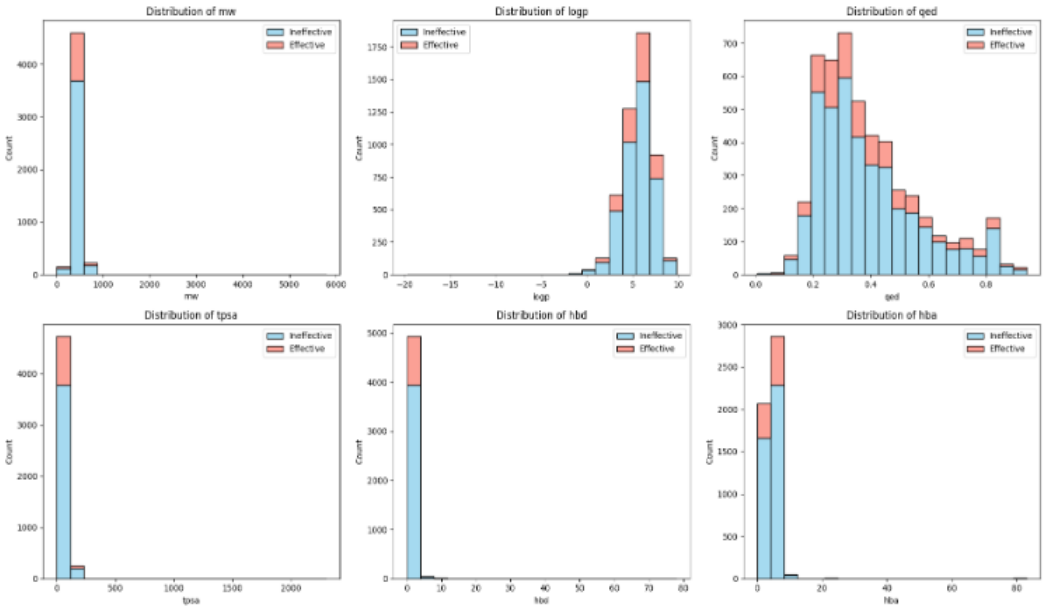

Figure 9. Drug properties distribution

When compared with prior generative drug discovery studies that report similar QED and physicochemical property ranges without clinical conditioning, the present framework demonstrates that personalisation does not compromise chemical feasibility. Instead, the observed property distributions remain

consistent with established oral drug heuristics, suggesting that clinical conditioning can be introduced without degrading molecular quality.

Collectively, the results indicate that patient-conditioned generative modelling can produce pharmaceutically plausible compounds without degrading chemical feasibility. However, downstream effectiveness estimation is highly sensitive to dataset imbalance and to the calibration of probabilities. While effective candidates exhibit narrower, more desirable physicochemical ranges consistent with oral drug heuristics, misclassified compounds exhibit broader variability in LogP and TPSA, suggesting that membrane permeability and polarity remain critical discriminators. These findings highlight the need to jointly optimise generative quality and calibrated effectiveness prediction when translating personalised GenAI pipelines into real-world pharmaceutical workflows.

E. Discussion

Beyond its application to drug discovery, this study's findings reinforce a broader methodological concern in applied machine learning: high threshold-dependent accuracy may coexist with poor probabilistic discrimination, underscoring the need for calibration-aware evaluation in safety-critical AI systems.

The empirical findings directly address the study objectives by demonstrating (i) the feasibility of dual-VAE integration for patient-conditioned molecule generation, (ii) the generation of chemically valid, drug-like candidates, (iii) threshold-specific effectiveness discrimination, and (iv) the presence of calibration and class-imbalance effects that constrain probabilistic reliability.

This study investigated the feasibility of integrating molecular and clinical data within a generative artificial intelligence framework to support personalized diabetes drug discovery. The findings demonstrate that the proposed Clinical–Molecular VAE architecture can generate chemically valid, drug-like molecular candidates while conditioning its outputs on patient-specific clinical attributes.

The generated molecules exhibited favorable physicochemical properties, with QED scores consistently exceeding commonly accepted thresholds for drug-likeness. This aligns with prior studies reporting the effectiveness of generative models in exploring chemical space and proposing viable drug candidates [13] [57]. Unlike molecule-centric approaches such as DrugGPT [24] and MolGAN [27]. The proposed framework explicitly incorporates clinical information, addressing a critical limitation in existing GenAI drug discovery research.

In contrast to conventional generative drug discovery pipelines that optimise exclusively for molecular novelty and validity, this framework demonstrates the feasibility of conditioning generative latent spaces on patient metabolic attributes. Conceptually, the study reframes generative drug discovery as a problem of patient-conditioned probabilistic inference rather than purely chemical optimisation, marking a shift from population-level candidate generation toward patient-aware molecular design that aligns more closely with precision medicine and stratified therapeutics. The effectiveness classification results indicate strong sensitivity to effective compounds at a fixed decision threshold, as reflected by high recall and F1-score values. However, the ROC analysis revealed an AUC of 0.49, indicating poor discrimination across probability thresholds. This apparent

contradiction highlights an important methodological insight: while threshold-based metrics may suggest strong performance, probability calibration remains a critical challenge in generative drug discovery models. Similar calibration limitations have been observed in other AI-driven healthcare applications [15][47][48].

From a translational perspective, the outcome should be regarded with some caution. Indeed, while in silico approaches to drug likeness prediction may be good indicators of drug activity, they cannot be guaranteed. Thus, the approach may be regarded as a tool to prioritize compounds in the early stages of drug discovery.

Practically speaking, the integration of clinical and molecular data represents a step forward in the development of more personalized medicine, particularly for diseases with heterogeneity, such as diabetes. By integrating molecular generation with patient metabolic data, the proposed framework enables more targeted therapy development.

In terms of a new direction in drug discovery, this framework rethinks candidate generation in the preclinical stages, shifting from population-based molecule design to more patient-informed molecular design. This enables a form of therapeutic exploration that is intrinsically compatible with metabolic heterogeneity, a key factor in diabetes. While validation of a compound in a biological system remains essential, this new framework provides an additional layer of computer-based screening that could prioritize compounds relevant to a specific patient before conducting expensive experimental work.

Furthermore, the clinical characteristics used in this study constitute only a small fraction of metabolic characteristics. More information from genomics, lifestyle, and comorbidity may enable greater personalization but will likely require larger clinical cohorts.

Therefore, the term ‘personalized’ is used within this study to refer to patient-conditioned generative feasibility rather than overall individualized therapy optimization.

Nevertheless, it is worth emphasizing that all results presented in this paper are computational in nature. While drug-likeness and molecular validity measures are useful for early-stage results, it is also essential to conduct biochemical validation and clinical trials to assess the therapeutic efficacy, safety, and bioavailability of compounds of interest. As such, the findings should be interpreted as proof of feasibility rather than definitive clinical outcomes.

Collectively, these findings suggest that the next generation of generative drug discovery systems should be evaluated not only on molecular validity and novelty, but on their ability to produce calibrated, patient-conditioned outputs that remain reliable under data imbalance and real-world uncertainty.

F. Implications of the Study

This research has significant theoretical, practical, and policy implications for personalized generative drug discovery. From a theoretical perspective, this study extends generative drug discovery research by repositioning molecular generation as a conditional probabilistic inference problem rather than a purely chemical optimisation task. The research has significant practical implications: patient metabolic profile data can inform drug design, and drug molecules can be

generated, yet the reliability of drug-class prediction is significantly affected by class imbalance/miscalibration. These implications have significant policy implications, suggesting that clinical or policy reliance on drug designs generated through generative drug discovery should be integrated as complementary decision-support tools alongside established pharmacological pipelines. From a policy perspective, these findings support the need for regulatory sandboxes and governance frameworks that recognize GenAI systems as decision-support tools rather than autonomous clinical agents, with explicit requirements for calibration analysis and uncertainty reporting. Overall, this research suggests that personalized generative drug design should be viewed as complementary to precision diabetes drug design.

G. Conclusion

This study empirically demonstrates how generative artificial intelligence can be re-engineered to support patient-conditioned drug discovery through the integration of molecular and clinical data.A novel approach, a hybrid 'Clinical-Molecular Variational Autoencoder', has been developed and empirically tested using publicly available molecular datasets and patient clinical attributes. The results show promise in discovering drug molecules with chemically valid structures and personalized profiles.

Despite the model's high classification performance under its thresholds, limitations in discriminative reliability revealed by the probability calibration analysis underscore the importance of calibration as a key methodological concern in GenAI-based drug discovery models.

Overall, the study provides empirical evidence supporting the viability of personalized GenAI-based drug discovery for managing and treating diabetes, while also highlighting the need for careful interpretation and calibration awareness. As a result, it provides a practical and theoretically informed basis for addressing various issues in AI, ethics, and personalized medicine.

By integrating generative modeling, conditioning, and calibration-aware evaluation, this study provides an important methodological foundation for the development of next-generation, highly personalized, AI-driven pharmaceutical discovery platforms. As generative AI continues to play an important role in shaping the pharmaceutical innovation landscape, the framework and study presented herein will serve as a foundation for future research on clinically contextualized and reliability-aware molecular generation.

H. References

[1] B. Mutunhu, B. Chipangura, and H. Twinomurinzi, "Internet of Things in the Monitoring of Diabetes," *International Journal of Health Systems and Translational Medicine*, vol. 2, no. 1, pp. 1–20, 2022, doi: 10.4018/ijhstm.300336.

[2] T. A. Lutz, "Mammalian models of diabetes mellitus, with a focus on type 2 diabetes mellitus," *Nat Rev Endocrinol*, vol. 19, no. 6, pp. 350–360, 2023, doi: 10.1038/s41574-023-00818-3.

[3] E. W. Gregg *et al.*, "Improving health outcomes of people with diabetes: target setting for the WHO Global Diabetes Compact," *The Lancet*, vol. 401, no. 10384, pp. 1302–1312, Apr. 2023, doi: 10.1016/S0140-6736(23)00001-6.

[4] B. Mutunhu, B. Chipangura, and S. Singh, "An Exploration of Opportunities for Quantified-Self Technology in Diabetes Self-Care: A Systematic Literature Review," *J Health Inform Afr*, vol. 11, no. 2, pp. 17–30, 2024, doi: 10.12856/JHIA-2024-v11-i2-481.

[5] Y. Tsujimoto, S. Hiwa, Y. Nakamura, Y. Oe, and T. Hiroyasu, "L-MolGAN: An improved implicit generative model for large molecular graphs," May 17, 2021. doi: 10.26434/chemrxiv.14569545.v3.

[6] B. M. Ndlovu, B. Chipangura, and S. Singh, "Factors Influencing Quantified SelfTechnology Adoption in Monitoring Diabetes," in *Lecture Notes in Networks and Systems*, X.-S. Yang, S. Sherratt, N. Dey, and A. Joshi, Eds., Singapore: Springer Nature Singapore, 2024, pp. 469–479. doi: 10.1007/978-981-97-3562-4_37.

[7] B. M. Ndlovu, B. Chipangura, and S. Singh, "The readiness to use quantified self-technology : A case of diabetic patients from a hospital in Bulawayo , Zimbabwe," 2025, doi: 10.1177/20552076251376286.

[8] Z. K. A. N. D. M. A. A. Dimore Abraham Lomboro AND Edosa, "Glycemic control and diabetes complications among adult type 2 diabetic patients at public hospitals in Hadiya zone, Southern Ethiopia," *PLoS One*, vol. 18, no. 3, pp. 1–14, May 2023, doi: 10.1371/journal.pone.0282962.

[9] L. Mpofu, B. Ndlovu, S. Dube, M. Muduva, F. Jacqueline, and K. Maguraushe, "Predictive Model for Hospital Readmission of Diabetic Patients," IEOM Society International, Jul. 2024. doi: 10.46254/af05.20240252.

[10] B. Mutunhu, B. Chipangura, and S. Singh, "Towards a quantified-self technology conceptual framework for monitoring diabetes," *South African Journal for Science and Technology*, vol. 43, no. 1, pp. 69–84, 2024, doi: 10.36303/SATNT.2024.43.1.970.

[11] A. Gangwal *et al.*, "Generative artificial intelligence in drug discovery: basic framework, recent advances, challenges, and opportunities," 2024, *Frontiers Media SA*. doi: 10.3389/fphar.2024.1331062.

[12] A. Schneuing *et al.*, "Structure-based Drug Design with Equivariant Diffusion Models," Oct. 2022, [Online]. Available: http://arxiv.org/abs/2210.13695

[13] T. Masuda, M. Ragoza, and D. R. Koes, "Generating 3D Molecular Structures Conditional on a Receptor Binding Site with Deep Generative Models," Oct. 2020, [Online]. Available: http://arxiv.org/abs/2010.14442

[14] N. M. Koutroumpa, K. D. Papavasileiou, A. G. Papadiamantis, G. Melagraki, and A. Afantitis, "A Systematic Review of Deep Learning Methodologies Used in the Drug Discovery Process with Emphasis on In Vivo Validation," Apr. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/ijms24076573.

[15] X. Peng, S. Luo, J. Guan, Q. Xie, J. Peng, and J. Ma, "Pocket2Mol: Efficient Molecular Sampling Based on 3D Protein Pockets," 2022.

[16] A. Gangwal, A. Ansari, I. Ahmad, and A. K. Azad, "Generative arti fi cial intelligence in drug discovery : basic framework , recent advances ,

challenges , and opportunities," no. February, pp. 1–26, 2024, doi: 10.3389/fphar.2024.1331062.

[17] Murwirapachena, P., Ndlovu, B., Maguraushe, K., Dube, M. (2026). Generative Artificial Intelligence in Drug Discovery for Diabetes: A Systematic Literature Review. In: Bhateja, V., Hoong, A.L.S., Kong, J.D., Urooj, S. (eds) Smart Computing Paradigms: Intelligence and Network Applications. SCI 2025. Lecture Notes in Networks and Systems, vol 1682. Springer, Cham. https://doi.org/10.1007/978-3-032-08240-4_11.

[18] I. Ghebrehiwet, N. Zaki, R. Damseh, and M. Saberi, *Revolutionizing personalized medicine with generative AI : a systematic review Generative Adversarial Network for Distribution Analysis*, vol. 57, no. 5. Springer Netherlands, 2024. doi: 10.1007/s10462-024-10768-5.

[19] S. Hadebe, B. Ndlovu, and K. Maguraushe, "Managing Diabetes Using Machine Learning and Digital Twins," *Indonesian Journal of Innovation and Applied Sciences (IJIAS)*, vol. 5, no. 2, pp. 145–162, 2025, doi: 10.47540/ijias.v5i2.1981.

[20] K. L. Pullyblank *et al.*, "Diabetes On Track: Redesigning Diabetes Management in Rural Primary Care," *Diabetes Spectrum*, Nov. 2024, doi: 10.2337/ds24-0043.

[21] F. Liu *et al.*, "DrugGPT: A Knowledge-Grounded Collaborative Large Language Model for Evidence-based Drug Analysis," Oct. 06, 2023. doi: 10.21203/rs.3.rs-3411728/v1.

[22] A. Blanco-González *et al.*, "The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies," Jun. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/ph16060891.

[23] C. Hasselgren and T. I. Oprea, "Artificial Intelligence for Drug Discovery: Are We There Yet?," *Annu. Rev. Pharmacol. Toxicol. Accepted. Annual Reviews in Pharmacology and Toxicology*, vol. 64, p. 2024, 2024, doi: 10.1146/annurev.

[24] Y. Li, C. Gao, X. Song, X. Wang, Y. Xu, and S. Han, "DrugGPT: A GPT-based Strategy for Designing Potential Ligands Targeting Specific Proteins," Jun. 30, 2023. doi: 10.1101/2023.06.29.543848.

[25] G. Lysandrou, R. E. Owen, K. Mursec, G. Le Brun, and E. A. L. Fairley, "Comparative Analysis of Drug-GPT and ChatGPT LLMs for Healthcare Insights: Evaluating Accuracy and Relevance in Patient and HCP Contexts," Jul. 2023, [Online]. Available: http://arxiv.org/abs/2307.16850

[26] Z. K. Othman *et al.*, "Advancing drug discovery and development through GPT models: a review on challenges, innovations and future prospects," *Intell Based Med*, vol. 11, p. 100233, 2025, doi: https://doi.org/10.1016/j.ibmed.2025.100233.

[27] A. Ramachandran, "Graph Neural Networks for Molecular Property Prediction in Drug Discovery: A Comprehensive Review," Nov. 2024.

[28] N. Yang, H. Wu, K. Zeng, Y. Li, and J. Yan, "Molecule Generation for Drug Design: a Graph Learning Perspective," Feb. 2022, [Online]. Available: http://arxiv.org/abs/2202.09212

[29] D. Crucitti, C. Pérez Míguez, J. Á. Díaz Arias, D. B. Fernandez Prada, and A. Mosquera Orgueira, "De novo drug design through artificial intelligence: an

introduction," *Frontiers in Hematology*, vol. 3, Jan. 2024, doi: 10.3389/frhem.2024.1305741.
[30] A. Malusare and V. Aggarwal, "Improving Molecule Generation and Drug Discovery with a Knowledge-enhanced Generative Model," Feb. 2024, doi: 10.1109/TCBB.2024.3477313.
[31] C. Hu *et al.*, "ScaffoldGVAE: scaffold generation and hopping of drug molecules via a variational autoencoder based on multi-view graph neural networks," *J Cheminform*, vol. 15, no. 1, p. 91, 2023, doi: 10.1186/s13321-023-00766-0.
[32] I. Ghebrehiwet, N. Zaki, R. Damseh, and M. S. Mohamad, "Revolutionizing personalized medicine with generative AI: a systematic review," *Artif Intell Rev*, vol. 57, no. 5, p. 128, Apr. 2024, doi: 10.1007/s10462-024-10768-5.
[33] J. Sun *et al.*, "Investigating Explainability of Generative AI for Code through Scenario-based Design," in *International Conference on Intelligent User Interfaces, Proceedings IUI*, Association for Computing Machinery, Mar. 2022, pp. 212–228. doi: 10.1145/3490099.3511119.
[34] A. Alakhdar, B. Poczos, and N. Washburn, "Diffusion Models in De Novo Drug Design," *J Chem Inf Model*, vol. 64, no. 19, pp. 7238–7256, Oct. 2024, doi: 10.1021/acs.jcim.4c01107.
[35] G. Corso, H. Stärk, B. Jing, R. Barzilay, and T. Jaakkola, "DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking," Oct. 2022, [Online]. Available: http://arxiv.org/abs/2210.01776
[36] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," Jun. 2020, [Online]. Available: http://arxiv.org/abs/2006.11239
[37] X. Tang *et al.*, "A survey of generative AI for de novo drug design: new frontiers in molecule and protein generation," Jul. 01, 2024, *Oxford University Press*. doi: 10.1093/bib/bbae338.
[38] A. N. Jain, A. E. Cleves, and W. P. Walters, "DEEP-LEARNING BASED DOCKING METHODS: FAIR COMPARISONS TO CONVENTIONAL DOCKING WORKFLOWS BRIEF PERSPECTIVE," Dec. 2024. doi: doi.org/10.48550/arXiv.2412.02889.
[39] M. A. Ketata *et al.*, "DiffDock-PP: Rigid Protein-Protein Docking with Diffusion Models," Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.03889
[40] A. E. Blanchard, C. Stanley, and D. Bhowmik, "Using GANs with adaptive training data to search for new molecules," *J Cheminform*, vol. 13, no. 1, p. 14, 2021, doi: 10.1186/s13321-021-00494-3.
[41] Y. Tang, R. Moretti, and J. Meiler, "Recent Advances in Automated Structure-Based De Novo Drug Design," *J Chem Inf Model*, vol. 64, no. 6, pp. 1794–1805, Mar. 2024, doi: 10.1021/acs.jcim.4c00247.
[42] Z. Zhang, F. Li, J. Guan, Z. Kong, L. Shi, and S. Zhou, "GANs for Molecule Generation in Drug Design and Discovery," 2022, pp. 233–273. doi: 10.1007/978-3-030-91390-8_11.
[43] S. Ali *et al.*, "Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence," *Information Fusion*, vol. 99, Nov. 2023, doi: 10.1016/j.inffus.2023.101805.
[44] D. R. Serrano *et al.*, "Artificial Intelligence (AI) Applications in Drug Discovery and Drug Delivery: Revolutionizing Personalized Medicine," Oct. 01, 2024,

Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/pharmaceutics16101328.

[45] A. Gangwal and A. Lavecchia, "Unleashing the power of generative AI in drug discovery," *Drug Discov Today*, vol. 29, no. 6, p. 103992, Jun. 2024, doi: 10.1016/j.drudis.2024.103992.

[46] Y. Fang, X. Pan, and H. Bin Shen, "De novo drug design by iterative multiobjective deep reinforcement learning with graph-based molecular quality assessment," *Bioinformatics*, vol. 39, no. 4, Apr. 2023, doi: 10.1093/bioinformatics/btad157.

[47] R. M. Ratwani, D. W. Bates, and D. C. Classen, "Patient Safety and Artificial Intelligence in Clinical Care," Feb. 23, 2024, *American Medical Association*. doi: 10.1001/jamahealthforum.2023.5514.

[48] G. Vilone and L. Longo, "Notions of explainability and evaluation approaches for explainable artificial intelligence," *Information Fusion*, vol. 76, pp. 89–106, Dec. 2021, doi: 10.1016/j.inffus.2021.05.009.

[49] A. Walling *et al.*, "Standard HIPAA Authorization Forms Decreased Response Rates for a Multi-site Pragmatic Trial," *J Gen Intern Med*, Jan. 2024, doi: 10.1007/s11606-024-09074-1.

[50] J. Goldstraw-White, "LEGAL AND POLICY FRAMEWORK FOR DIGITAL FORENSICS: A RESOURCE FOR PRACTITIONERS A Policy and Practice Briefing from the Digital Forensics and Social Media project funded by the Dawes Trust Copyright," 2022.

[51] S. Matoori, "Diabetes and its Complications," *ACS Pharmacol Transl Sci*, vol. 5, no. 8, pp. 513–515, Aug. 2022, doi: 10.1021/acsptsci.2c00122.

[52] M. Deng, R. Yang, X. Zheng, and Y. Deng, "Arti fi cial intelligence in diabetes care : from predictive analytics to generative AI and implementation challenges," no. November, pp. 1–7, 2025, doi: 10.3389/fendo.2025.1620132.

[53] S. A. Rabbani, M. El-tanani, S. Sharma, S. S. Rabbani, and Y. El-tanani, "Generative Artificial Intelligence in Healthcare : Applications , Implementation Challenges , and Future Directions," pp. 1–40, 2025.

[54] Omari Firas, "A combination of SEMMA & CRISP-DM models for effectively handling big data using formal concept analysis based knowledge discovery: A data mining approach," *World Journal of Advanced Engineering Technology and Sciences*, vol. 8, no. 1, pp. 009–014, Jan. 2023, doi: 10.30574/wjaets.2023.8.1.0147.

[55] S. Kim *et al.*, "PubChem 2023 update.," *Nucleic Acids Res*, vol. 51, no. D1, pp. D1373–D1380, Jan. 2023, doi: 10.1093/nar/gkac956.

[56] N. J. Castellino, A. P. Montgomery, J. J. Danon, and M. Kassiou, "Late-stage Functionalization for Improving Drug-like Molecular Properties," *Chem Rev*, vol. 123, no. 13, pp. 8127–8153, Jul. 2023, doi: 10.1021/acs.chemrev.2c00797.

[57] A. Farhadi, A. Zamanifar, and M. Faezipour, "Application of Generative AI in Drug Discovery," in *Application of Generative AI in Healthcare Systems*, Cham: Springer Nature Switzerland, 2025, pp. 155–174. doi: 10.1007/978-3-031-82963-5_6.