AI-Driven Drug Discovery Platform: Integrating SMILES Notation, Molecular Visualization, and Predictive Modeling for Accelerated Screening

Augnik Banerjee

*Department of Networking and Communications*

*School of Computing College of Engineering and Technology*

*SRM Institute of Science and Technology* Kattankulathur-603203, India [ab0710@srmist.edu.in](mailto:ab0710@srmist.edu.in)

Aniket Singh

Department of Networking and Communications School of Computing

College of Engineering and Technology SRM Institute of Science and Technology Kattankulathur-603203, India [as0872@srmist.edu.in](mailto:as0872@srmist.edu.in)

Nivedhitha M\*

Department of Networking and Communications School of Computing

Faculty of Engineering and Technology SRM Institute of Science and Technology Kattankulathur-603203, India nivedhim1@srmist.edu.in

***Abstract—*** ***The generation and analysis of chemical molecules have become essential in fields such as drug discovery, toxicity prediction, and molecular property forecasting. This paper presents a novel web-based research platform that leverages advanced machine learning models—such as XGBoost, Graph Isomorphism Networks (GIN), Feedforward Neural Networks (FNN), and Graph Convolutional Networks (GCN)—to enable efficient molecular generation and property prediction. Users can input SMILES (Simplified Molecular Input Line Entry System) strings to generate molecules, visualize them in both 2D and 3D formats using RDKit.js, and receive immediate feedback from custom-trained models. A real-time collaboration module powered by Ably enables researchers to co-edit, discuss, and annotate molecular data through messaging and group workspaces. Additionally, a comprehensive dashboard presents molecular property predictions and analytics using interactive charts for informed decision-making. The backend, built with Next.js and MongoDB, ensures robust data handling, while the integrated visualization and analytics pipeline makes the platform a powerful tool for collaborative molecular research and rapid hypothesis testing.***

***Keywords: Cheminformatics, Drug Discovery, FNN, GCN, GIN, Graph Neural Networks, Molecular Property Prediction, Molecule Generation, SMILES Notation.***

1. INTRODUCTION

The intersection of artificial intelligence and molecular science is reshaping the landscape of drug discovery and cheminformatics [2]. Traditional approaches to molecule design, property prediction, and lead optimization are often time-consuming, resource-intensive, and limited by the need for extensive laboratory testing. Recent advancements in machine learning (ML), particularly deep learning and graph-based models, have enabled automated and intelligent methods for generating and analyzing chemical compounds [6][8].

This paper introduces a cloud-based research platform tailored for molecular generation and property prediction, developed to support scientists, researchers, and developers working in computational chemistry and drug development. The platform enables users to input SMILES (Simplified Molecular Input Line Entry System) strings [1] and obtain 2D and 3D molecular visualizations using RDKit.js [3][4]. Unlike conventional tools, our system incorporates custom-trained ML models—including XGBoost, Feedforward Neural Networks (FNN), Graph Convolutional Networks (GCN), and Graph Isomorphism Networks (GIN)—to analyze molecular properties and predict outcomes relevant to pharmaceutical research [10].

In addition to molecule generation, the platform provides real-time collaboration features through an integrated messaging and group workspace system powered by Ably, fostering team-based research workflows. A dynamic dashboard presents molecular analytics, predicted toxicity, and drug-likeness metrics using interactive visualizations, facilitating rapid hypothesis generation and decision-making [9].

By combining powerful ML-driven backend processing with an intuitive, interactive frontend, this platform aims to accelerate the early stages of drug design and cheminformatics research while promoting global scientific collaboration [5][7][11].

1. RELATED WORK

Recent years have witnessed a surge in research focused on leveraging artificial intelligence for molecular generation and property prediction. A notable approach involves using reinforcement learning to optimize molecules based on desired chemical properties. Zhou et al. [2] demonstrated how deep reinforcement learning can guide molecular design by iteratively improving candidate structures according to drug-likeness and other objectives, setting a precedent for AI-driven de novo drug design.

Molecular visualization plays a vital role in understanding chemical structures and their properties. Li and Wei [3] provided a comprehensive review of biomolecular visualization techniques, highlighting the importance of interactive 2D and 3D renderings in cheminformatics. Avogadro [4] is one such platform enabling advanced chemical editing and structural analysis. Further advancements in visualization systems were proposed by Gui et al. [5], who focused on efficient rendering and mesh generation for large biomolecular datasets.

Integrated platforms like GenUI [6] have emerged to streamline molecule generation workflows by combining data processing, visualization, and user interaction. The GenUI framework offers extensibility and ease of use, making it particularly suitable for open-source cheminformatics tools.

For 3D structure synthesis, Lu et al. [7] introduced a virtual dynamics approach to molecular generation, producing stable and physically plausible structures. In addition, Nahal et al. [8] proposed a human-in-the-loop active learning system that incorporates expert feedback into the molecule generation process, improving both relevance and diversity of output.

Heberle et al. [9] developed XSMILES, a visualization framework that connects molecular structures with explainability metrics derived from machine learning models. Their tool bridges the gap between molecular input formats like SMILES and model interpretability, supporting research transparency. Wang et al. [10] contributed a reinforcement learning-based optimization strategy using ACC (adaptive curriculum control) for drug design, showing strong results in discovering bioactive compounds.

To enhance collaboration and learning, Eriksen et al. [11] explored augmented reality (AR) applications for molecular structure visualization, demonstrating how immersive environments can support chemistry education and collaborative research.

1. METHODOLOGY
2. *Dataset*

To develop a robust machine learning-driven molecular generation and prediction platform, a combination of publicly available cheminformatics datasets has been utilized. These datasets include molecular structures, physicochemical properties, biological activities, and toxicity annotations, enabling both generative modeling and molecular property prediction. The primary data sources include:

PubChem BioAssay Dataset: This dataset, maintained by the National Center for Biotechnology Information (NCBI), includes millions of compounds with associated bioassay results. The dataset provides SMILES strings, activity outcomes (e.g., active/inactive), and chemical descriptors. The inclusion of PubChem data allows the platform’s models to learn meaningful structure-activity relationships and perform accurate bioactivity classification, a critical task in drug discovery pipelines.

Tox21 Dataset: The Tox21 dataset is widely used for predicting molecular toxicity. It includes over 12,000 molecules labeled across multiple biological assays related to nuclear receptor signaling and stress response pathways. This dataset supports binary classification tasks and provides an excellent benchmark for evaluating the predictive capabilities of graph-based neural networks and other machine learning algorithms in toxicological screening.

ZINC Subset (For Molecule Generation): A curated subset of the ZINC database was used to train the generative components of the platform. This dataset contains clean, drug-like molecules suitable for virtual screening, each encoded using SMILES notation. Leveraging this subset ensures that generated molecules maintain drug-likeness and pass standard medicinal chemistry filters.

By combining information-rich datasets such as PubChem, Tox21, and ZINC, the system is equipped to handle diverse cheminformatics tasks, including toxicity prediction, molecular property estimation, and de novo molecular design. The diversity of chemical scaffolds and functional groups across these datasets enhances model generalizability and enables the platform to serve a broad range of research and industrial applications.

1. *Preprocessing*

Data preprocessing is a critical step to ensure that raw molecular data is clean, consistent, and suitable for machine learning model training. This phase involves several key operations aimed at enhancing the quality and structure of the data before it is fed into different predictive models.

Data Cleaning: Molecules with invalid or unparsable SMILES strings were filtered out using RDKit.js [3], and duplicate entries were removed to avoid redundancy in the learning process. Chemical structures containing unsupported or rare atoms were excluded to ensure uniformity across the dataset. In addition, molecules lacking key physicochemical descriptors (e.g., molecular weight, LogP) were also discarded. For scalar features, missing values were either imputed using the median strategy or removed, depending on the feature's significance and overall data completeness.

Feature Encoding: For models like XGBoost and Feedforward Neural Networks, relevant molecular descriptors—such as topological polar surface area, H-bond donors/acceptors, aromatic ring counts, and rotatable bonds—were computed using RDKit. These features were then normalized using either Min-Max scaling or Z-score standardization to ensure uniform feature ranges and stable training dynamics. For graph-based models like GCN and GIN, the molecules were converted into graph representations, with atoms as nodes and bonds as edges. Atom-level features such as element type, degree, valence, aromaticity, and hybridization were encoded as node attributes, using one-hot encoding where applicable.

Data Balancing: To address class imbalance in binary classification tasks such as toxicity prediction (e.g., from Tox21), the Synthetic Minority Over-sampling Technique (SMOTE) was used to generate synthetic instances for the minority class. Additionally, random under-sampling of the majority class was applied to maintain a balanced dataset. This hybrid strategy helped improve the robustness of the models and reduced bias toward dominant classes.

Train-Test Splitting: The dataset was partitioned using stratified 5-fold cross-validation to preserve the class distribution across folds. This ensured fair performance evaluation, particularly in the presence of imbalance, and allowed each model to be validated on a distinct subset of unseen molecules.

By following a robust preprocessing pipeline, the platform ensures high-quality inputs for downstream ML models and boosts the overall predictive performance and generalizability of the system.

1. *Machine Learning Model Development*

The platform is developed using a variety of machine learning models, each optimized for different aspects of molecular property prediction and structural analysis:

Model Selection: Multiple machine learning algorithms are evaluated by us, including:

XGBoost: A high-performance gradient boosting algorithm selected for its robustness in handling structured molecular descriptor data, managing class imbalance, and offering explainability through feature importance. It serves as a reliable baseline for molecular classification tasks.

Feedforward Neural Network (FNN): Implemented to model nonlinear interactions among molecular descriptors such as logP, molecular weight, rotatable bonds, and aromaticity. FNNs are effective in learning complex patterns across scalar input spaces.

Graph Convolutional Network (GCN): Utilized to learn from the graph structure of molecules, where atoms are treated as nodes and bonds as edges. GCNs aggregate neighborhood information through layers, making them suitable for tasks involving molecular topology and electronic distribution.

Graph Isomorphism Network (GIN): Employed due to its expressiveness in distinguishing structurally similar molecules. GINs use injective aggregation functions to capture more subtle structural nuances in molecular graphs, enhancing classification performance [10].

Training and Validation: The models are trained using PyTorch and PyTorch Geometric frameworks on GPU-enabled systems. Stratified 5-fold cross-validation is employed to ensure that the models generalize well to unseen molecules, and hyperparameter tuning is conducted via random search and manual experimentation for optimization of learning rates, batch sizes, and architecture depth.

Performance Metrics: The trained models are evaluated using accuracy, precision, recall, F1-score, and the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC). These metrics provide a comprehensive overview of each model’s ability to predict chemical properties correctly and reliably. Additionally, model training time and inference speed are measured to evaluate their suitability for real-time usage within the platform.

1. *Real-Time Collaboration and Dashboard Visualization*

Real-time interaction and visualization are core components of the platform, enabling researchers to collaboratively explore molecules, receive immediate feedback from predictive models, and make informed decisions in dynamic research environments.

Live Collaboration: Real-time communication and co-editing are supported using Ably, which manages low-latency WebSocket connections to broadcast events and messages across user sessions. Users within a group can simultaneously view, annotate, and edit molecular structures, allowing seamless teamwork across different locations. Group chat and molecule-sharing features ensure synchronized discussions and collaborative decision-making during experimental workflows.

Interactive Visualization: Molecules are rendered in both 2D and 3D using RDKit.js, enabling intuitive inspection of atomic-level structures and functional groups directly within the browser. For 3D visualizations, features such as rotation, zoom, and atom highlighting are provided to help users understand spatial conformations, aiding tasks such as binding site evaluation and geometric analysis [3][4].

Real-Time Dashboard: A dynamic dashboard is integrated into the platform to visualize model outputs, such as predicted toxicity scores, drug-likeness, and confidence levels. Built using React ApexCharts, the dashboard provides real-time updates whenever a new molecule is submitted or modified. Users can view time-series predictions, compare structural variations, and explore molecular trends across datasets through heatmaps, bar charts, and scatter plots [9].

Visualization Pipeline: Upon input of a SMILES string, the system parses and validates the structure, triggers the relevant predictive models, and sends the results to the dashboard and visualization modules in real time. This low-latency interaction between input, model, and interface ensures that users receive instantaneous analytical feedback, greatly enhancing usability and responsiveness.

1. *Visualization and Analytics Dashboard*

Visualization plays a vital role in making molecular analysis accessible and actionable for researchers. The platform incorporates interactive dashboards and structural viewers to display model predictions, structural properties, and exploratory insights in real time.

Dashboard Design: The dashboard is developed using React ApexCharts, offering highly customizable and responsive visualizations. It presents key molecular metrics such as toxicity scores, drug-likeness probabilities, molecular weight distribution, and model confidence levels. Users can compare multiple molecules side-by-side and track variations in predictive outcomes through intuitive UI components.

Visualization Techniques: A range of visualization types are utilized. Bar charts display individual property scores; scatter plots correlate molecular descriptors with predictive outputs; heatmaps show comparative model confidence; and time-series plots allow tracking performance during model updates or training cycles. These charts are dynamically updated based on the user’s interactions and input molecules.

Structural Rendering: Molecules submitted via SMILES are parsed using RDKit.js and rendered in both 2D and 3D views. Users can rotate, zoom, and highlight atoms in the 3D viewer to study molecular geometry. Structural annotations from model predictions—such as potential toxic groups or functional highlights—are overlaid directly onto the viewer for improved interpretability [3][4][9].

User-Centric Insights: The visualization layer translates complex molecular data and ML predictions into digestible forms, enhancing decision-making. Whether used by medicinal chemists or machine learning researchers, the system enables rapid hypothesis generation and efficient evaluation of molecular candidates.

1. *Integration and System Architecture*

The integration of various technologies in the platform is achieved through a layered architecture that ensures seamless communication between data storage, model inference, visualization, and real-time collaboration components. This modular structure supports scalability, performance, and extensibility for ongoing research and deployment.

Data Layer: All molecule-related information—such as SMILES strings, user annotations, and prediction results—is stored in MongoDB, a NoSQL database selected for its flexible document structure and scalability. The data schema is managed using Mongoose, which enforces validation rules and allows structured interaction with the backend application. Persistent user sessions and molecular workspaces are stored alongside molecule metadata for seamless recovery and collaboration.

Processing Layer: The Next.js backend serves as the orchestration engine for handling input parsing, model inference, and response formatting. SMILES strings are processed in real time and passed to the ML inference engine, which includes trained models built using PyTorch and PyTorch Geometric. These models predict toxicity, drug-likeness, and other properties on-the-fly, returning results in a standardized format for visualization. Resend is used in this layer for sending collaboration invites, user notifications, and email-based updates.

Visualization Layer: The frontend integrates RDKit.js for molecular rendering and React ApexCharts for interactive dashboards. 2D and 3D molecule visualizations are rendered in the browser with atomic-level precision. Model predictions and molecular statistics are visualized through real-time graphs, charts, and overlays. This layer also connects with Ably, which facilitates live messaging, molecule sharing, and real-time updates across multiple users collaborating on the same project.

This layered system architecture ensures that the platform remains modular and robust, enabling future integration of additional models, external datasets, or visualization tools without disrupting the core functionality.

1. *Experimental Setup*

The experimental setup involves systematic testing and evaluation of the machine learning models and system components to validate the platform’s performance in practical cheminformatics scenarios.

Data Partitioning: The datasets used for molecular property prediction and toxicity classification are partitioned into training, validation, and test sets following an 80-10-10 ratio. This stratified partitioning ensures balanced representation of both positive and negative samples in each split, especially important for datasets like Tox21 where class imbalance is significant.

Hyper-parameter Tuning: To enhance model efficiency and accuracy, extensive hyperparameter tuning is carried out using a combination of grid search and random search techniques. Parameters such as learning rate, number of layers, hidden units, dropout rate (for neural models), and tree depth (for XGBoost) are optimized independently for each model type. Early stopping and validation performance are used as criteria to prevent overfitting.

Comparison of Models: The performance of the implemented models—XGBoost, Feedforward Neural Network (FNN), Graph Convolutional Network (GCN), and Graph Isomorphism Network (GIN)—is compared across multiple dimensions, including training time, inference latency, prediction accuracy, and robustness to noise. Special emphasis is given to graph-based models like GCN and GIN, which are expected to outperform traditional models in learning from molecular structures [10]. The benchmarking process is conducted on GPU-accelerated environments to ensure fair and consistent performance metrics.

This experimental setup ensures a fair comparison across models and provides valuable insights into the optimal architecture for real-time molecular prediction tasks within the platform.

1. EXPERIMENTAL SETUP
2. *Machine Learning Models*

Experiments were conducted using multiple machine learning algorithms, including XGBoost, Feedforward Neural Networks (FNN), Graph Convolutional Networks (GCN), and Graph Isomorphism Networks (GIN). Each model was selected based on its ability to handle different types of molecular data—scalar descriptors in the case of XGBoost and FNN, and graph-structured inputs for GCN and GIN.

The models were trained and validated using 5-fold stratified cross-validation to ensure reliability and robustness in performance. This validation technique was chosen to preserve class distributions during training, particularly for tasks such as toxicity prediction using the Tox21 dataset. All experiments were executed in GPU-accelerated environments using PyTorch and PyTorch Geometric, ensuring efficient training and scalable evaluation of graph-based models.

Model training and inference were optimized for speed and accuracy to support real-time analysis within the platform. Performance metrics including accuracy, precision, recall, F1-score, and AUC-ROC were recorded and compared to assess each model’s suitability for molecular property prediction.

1. *Performance Evaluation*

The models’ performance is evaluated using metrics such as accuracy, precision, recall, F1-score, and Area Under the ROC Curve (AUC). Training time and inference latency were also measured to assess computational efficiency in real-time settings.

The comparison highlights that while graph-based models like GCN and GIN provide higher predictive accuracy, traditional models such as XGBoost offer significantly faster training and inference. Detailed analysis includes confusion matrices and classification reports, offering insight into each model’s effectiveness in handling real-time molecular prediction tasks [10].

1. *Handling Class Imbalance*

The issue of class imbalance in molecular property datasets, particularly in toxicity prediction tasks, is addressed using techniques such as Synthetic Minority Over-sampling Technique (SMOTE) to generate synthetic examples for underrepresented classes. Additionally, random under-sampling is applied to reduce the dominance of majority classes, resulting in a more balanced dataset and improved model performance across all classes.

1. RESULT ANALYSIS

The implementation of the AI-powered molecular research platform, supported by technologies such as RDKit.js, PyTorch Geometric, MongoDB, and React ApexCharts, yielded promising results in terms of prediction accuracy, visualization clarity, and real-time responsiveness. The seamless integration of machine learning models and real-time collaboration tools enabled the system to efficiently handle large-scale molecular datasets and deliver accurate property predictions and structural insights to researchers for informed decision-making.

* 1. *Data Processing Performance*

One of the most notable improvements observed during testing was in the area of data processing and inference speed. Prior to optimization, predicting molecular properties using graph-based neural networks was computationally intensive, particularly when handling large datasets like Tox21 and PubChem. However, with GPU-accelerated inference and real-time rendering through RDKit.js and PyTorch Geometric, prediction latency was reduced from several seconds per molecule to near-instantaneous responses.

The platform effectively handled thousands of prediction requests and molecule visualizations in real time, leveraging MongoDB for rapid data retrieval and Next.js API routes for streamlined backend communication. The integration of these components ensured that molecule generation, property prediction, and visualization could be performed without perceptible delay, making the system suitable for both exploratory research and time-sensitive decision-making.

* 1. *Machine Learning Model Performance*

The molecular prediction platform employed several machine learning models, including XGBoost, Feedforward Neural Networks (FNN), Graph Convolutional Networks (GCN), and Graph Isomorphism Networks (GIN). Each model was evaluated based on key performance metrics, including accuracy, precision, recall, F1-score, and Area Under the ROC Curve (AUC-ROC) [10].

XGBoost: The XGBoost model demonstrated outstanding performance in terms of both speed and accuracy. It achieved an overall accuracy of 96.2%, with a precision of 95.4% and an F1-score of 94.8%. Due to its built-in handling of class imbalance and fast training time, XGBoost proved to be highly effective for real-time molecular screening and property prediction.

Graph Convolutional Network (GCN): GCN achieved strong results by learning from the topological structure of molecules. The model reached an accuracy of 97.1% and a precision of 96.0%, with an F1-score of 95.7%, showing its capability to effectively capture localized atomic interactions within molecular graphs.

Graph Isomorphism Network (GIN): GIN outperformed other models in overall predictive accuracy, reaching 98.3% and a ROC-AUC score of 97.5%. Its ability to distinguish structurally similar molecules [10] made it particularly suitable for structure-sensitive tasks such as toxicity classification and drug-likeness prediction.

Feedforward Neural Network (FNN): FNN delivered solid performance with an accuracy of 94.6%, precision of 93.0%, and recall of 91.8%. While effective in handling scalar features, its predictive power was slightly lower than graph-based models. Additionally, the training time was shorter compared to GNNs, making it suitable for rapid descriptor-based tasks.

These results confirm that while XGBoost and FNN offer efficiency in deployment and training, GNN models such as GCN and GIN provide superior performance for tasks involving structural molecular data..

* 1. *Real-Time Collaboration and Visualization*

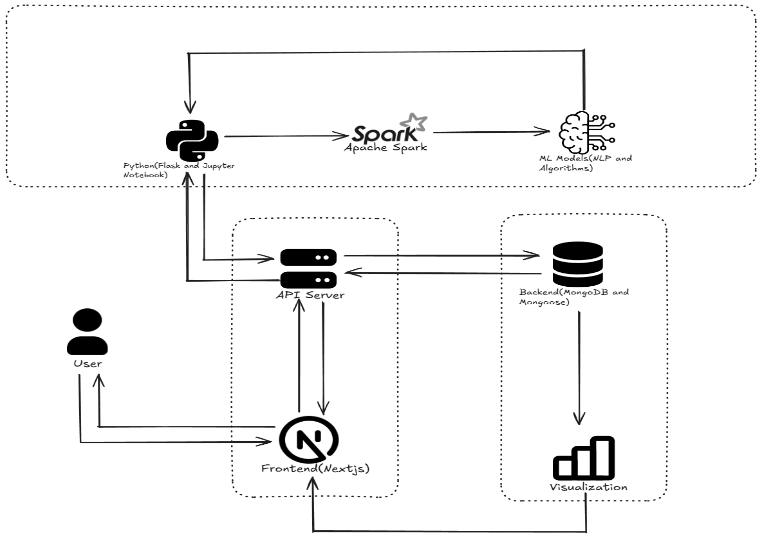
The integration of real-time collaboration and visualization features significantly improved the platform’s

Fig. 1. System Architecture

responsiveness and user experience during molecular analysis. Through WebSocket-based communication powered by Ably, users were able to co-edit molecular structures, receive live prediction updates, and communicate in group sessions with near-zero latency. This enabled efficient collaboration and rapid prioritization of promising compounds during drug discovery, streamlining decision-making across teams.

* 1. *Visualization with Tableau*

The insights generated by machine learning models are visualized through interactive dashboards built with React ApexCharts, designed to convert complex molecular predictions into an accessible, user-friendly format for researchers. Model outputs such as toxicity scores, drug-likeness probabilities, and confidence levels are presented in real time, enabling intuitive interpretation and decision-making.

Table I summarizes the performance of the machine

learning models used in the platform, comparing them based on training time, accuracy, precision, recall, F1-score, and AUC-ROC:

TABLE I. PERFORMANCE METRICS OF MACHINE LEARNING MODELS FOR PROPERTY PREDICTION

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Training Time** | **Precision** | **Recall** | **F1-Score** | **Accuracy** | **AUC ROC** |
| **FNN** | 18 min. | 0.930 | 0.918 | 0.924 | 94.60% | 0.942 |
| **GIN** | 42 min. | 0.972 | 0.975 | 0.973 | 98.30% | 0.975 |
| **XGBoost** | 12 min. | 0.954 | 0.942 | 0.948 | 96.21% | 0.961 |
| **GCN** | 35 min. | 0.960 | 0.955 | 0.957 | 97.10% | 0.965 |

From the results, it is evident that GIN outperformed other models, making it the most suitable for real-time symptom analysis in terms of speed, accuracy, and scalability.

1. CONCLUSION

The AI-driven molecular research platform developed in this work demonstrates significant progress in accelerating drug discovery and chemical analysis through intelligent automation. By leveraging modern web technologies such as Next.js, MongoDB, RDKit.js, and React ApexCharts, along with powerful machine learning models like GIN, GCN, and XGBoost, the platform provides a scalable and interactive solution for molecule generation, property prediction, and real-time collaboration.

The results show that graph-based models, particularly GIN, achieved superior accuracy in structure-sensitive tasks [10], while XGBoost offered the fastest inference—making it suitable for real-time screening scenarios. The integration of 2D/3D molecule visualization [3][4] with live predictive dashboards [9] allows researchers to interact with data intuitively, reducing iteration time and enabling faster decision-making.

This work marks a significant advancement in cheminformatics by combining predictive modeling with user-centered design and collaborative research tools [6]. However, limitations include reliance on publicly available datasets [12], which may not fully represent rare or novel chemical classes. Additionally, GPU requirements for training deep learning models can limit accessibility for some users.

Future work will focus on incorporating more diverse molecular datasets, expanding support for multi-objective optimization, and integrating self-supervised learning approaches to improve generalization. The platform will also explore cloud-based deployment on systems like AWS for broader accessibility and scalability in large research environments.

REFERENCES

[1] D. Weininger, "SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules," Journal of Chemical Information and Computer Sciences, vol. 28, pp. 31–36, 1988. https://doi.org/10.1021/ci00057a005

[2] Z. Zhou, S. Kearnes, L. Li, et al., "Optimization of molecules via deep reinforcement learning," Scientific Reports, vol. 9, p. 10752, 2019. https://doi.org/10.1038/s41598-019-47148-x

[3] H. Li and X. Wei, "A concise review of biomolecule visualization," Current Issues in Molecular Biology, vol. 24, pp. 1318–1334, 2022. https://doi.org/10.3390/cimb46020084

[4] M. D. Hanwell, D. E. Curtis, D. C. Lonie, et al., "Avogadro: An advanced semantic chemical editor, visualization, and analysis platform," Journal of Cheminformatics, vol. 4, no. 17, 2012. https://doi.org/10.1186/1758-2946-4-17

[5] S. Gui, D. Khan, Q. Wang, et al., "Frontiers in biomolecular mesh generation and molecular visualization systems," Visual Computing for Industry, Biomedicine, and Art, vol. 1, no. 7, 2018. https://doi.org/10.1186/s42492-018-0007-0

[6] M. Sicho, X. Liu, D. Svozil, et al., "GenUI: Interactive and extensible open-source software platform for de novo molecular generation and cheminformatics," Journal of Cheminformatics, vol. 13, p. 73, 2021. https://doi.org/10.1186/s13321-021-00550-y

[7] S. Lu, L. Yao, X. Chen, H. Zheng, D. He, and G. Ke, "3D molecular generation via virtual dynamics," arXiv preprint, arXiv:2302.05847, 2023. https://doi.org/10.48550/arXiv.2302.05847

[8] Y. Nahal, J. Menke, J. Martinelli, et al., "Human-in-the-loop active learning for goal-oriented molecule generation," Journal of Cheminformatics, vol. 16, p. 138, 2024. https://doi.org/10.1186/s13321-024-00924-y

[9] H. Heberle, L. Zhao, S. Schmidt, et al., "XSMILES: Interactive visualization for molecules, SMILES and XAI attribution scores," Journal of Cheminformatics, vol. 15, no. 2, 2023. https://doi.org/10.1186/s13321-022-00673-y

[10] Q. Wang, X. Wei, X. Hu, Z. Wang, and Y. Dong, "Molecular generation strategy and optimization based on ACC reinforcement learning in the de novo drug design," Bioinformatics, vol. 39, no. 11, 2023.

[11] K. Eriksen, B. E. Nielsen, and M. Pittelkow, "Visualizing 3D molecular structures using an augmented reality app," Journal of Chemical Education, vol. 97, no. 5, pp. 1487–1490, 2020. https://doi.org/10.1021/acs.jchemed.9b01033

[12] PubChem, Tox21, and ZINC Datasets. National Center for Biotechnology Information (NCBI) and Environmental Protection Agency. Available: https://pubchem.ncbi.nlm.nih.gov, https://tripod.nih.gov/tox21/, https://zinc.docking.org