Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Deep Ensemble Learning with Graph Attention Mechanisms for Molecular Classification of HIV Inhibitors

Abdul Ahad

ahad.abdul01@northsouth.edu

North South University

Md Mahabube Alahi Atik

North South University

Mirza Farhan Shahriar

North South University

Md Rahik Redwan

Webster University

Kazi Tanvir Akter

North South University

Research Article

Keywords: HIV inhibitors, Morgan fingerprints, Graph Attention Network, Stacking ensemble, Drug discovery, Computational chemistry

Posted Date: October 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11003141/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Deep Ensemble Learning with Graph Attention Mechanisms for Molecular Classification of HIV Inhibitors

Abdul Ahad[1*], Md Mahabube Alahi Atik[1], Mirza Farhan Shahriar[1], Md Rahik Redwan[2], Kazi Tanvir Akter[1]

[1*]Electrical and Computer Engineering, North South University, Dhaka, Bangladesh.

[2]MS in Marketing (Emphasis in Analytics), Webster University, St. Louis, Missouri, USA.

*Corresponding author(s). E-mail(s): ahad.abdul01@northsouth.edu;

Abstract

Predicting HIV inhibitors is important for accelerating antiviral drug discovery. This study compares fingerprint-based machine learning, graph-based deep learning, and ensemble approaches for classifying compounds in the MoleculeNet HIV dataset, which contains 41,127 SMILES-encoded molecules and is highly imbalanced (3.51% active). Morgan fingerprints (radius 2, 2048 bits) were used with Random Forest, SVM, XGBoost, CatBoost, LightGBM, and feedforward neural networks, while molecular graphs with atom and bond features were processed using a Graph Attention Network (GAT). Class imbalance was addressed through resampling, and Optuna was used for hyperparameter tuning. A hybrid ChemBERTa-GAT-fingerprint model with cross-attention was also evaluated, together with a stacking ensemble using Random Forest, LightGBM, CatBoost, and XGBoost as base learners and an FNN as meta-learner. XGBoost achieved 0.8567 accuracy and 0.8508 ROC-AUC; the GAT achieved 0.8100 accuracy and 0.8534 ROC-AUC. The stacking ensemble achieved 0.8521 accuracy and 0.8499 ROC-AUC, while the hybrid model obtained 0.8400 ROC-AUC and 0.816 PR-AUC. Attention-based visualizations highlighted atoms contributing strongly to

GAT predictions. These results show the complementary value of fingerprint-based and graph-based molecular representations for HIV inhibitor classification and support interpretable AI-assisted molecular screening.

Keywords: HIV inhibitors, Morgan fingerprints, Graph Attention Network, Stacking ensemble, Drug discovery, Computational chemistry

Key contribution.

This study demonstrates that graph-based deep learning models, particularly Graph Attention Networks, surpass traditional machine learning methods in predicting HIV inhibitors by effectively capturing molecular topology, and it provides a methodology for interpreting these models to identify key structural features. The research highlights the potential of integrating advanced AI techniques with molecular modeling to accelerate and enhance HIV drug discovery.

1 Introduction

The Human Immunodeficiency Virus (HIV) persists a substantial global health problem. The World Health Organization reveals that 39.9 million people live with the virus during 2023 [14]. The availability of antiretroviral therapies has not eliminated the need for new HIV inhibitor development because of drug resistance and restricted medication effectiveness [15]. The lower efficacy of traditional drug discovery strategies based on HTS and molecular docking requires extensive time and high costs [16]. Drug discovery techniques have adopted artificial intelligence techniques as essential computational instruments that drive data-predictive modeling to detect potential HIV inhibitors faster [1]. Computational drug discovery enables scientists to evaluate large-scale molecular data collections for deciphering chemical compounds while foreseeing their biological activities through computer simulations [17]. The HIV dataset from MoleculeNet, a benchmark platform for machine learning in chemistry and drug discovery, provides valuable molecular data for predictive modeling [13]. This HIV

dataset comprises 41,127 SMILES notation-based molecular structures with active or inactive labels based on their inhibitory potential against HIV. Given the dataset's highly imbalanced nature, i.e., 3.51% active compounds, developing effective predictive models requires specialized preprocessing and data-balancing techniques [2].

This study explores and compares two distinct computational approaches for predicting HIV inhibitors:

1. Fingerprint-based machine learning models – These models convert molecular structures into numerical feature vectors using Morgan fingerprints and apply machine learning algorithms such as Random Forest, SVM, CatBoost, and XGBoost to classify compounds. Morgan fingerprints were generated using a radius of 2 and a 2048-bit vector size, ensuring that local molecular substructures are encoded effectively for predictive modeling.
2. Graph-based deep learning models – These models retain molecular topology and chemical interactions, using Graph Attention Networks (GATs) to learn hierarchical molecular features directly from molecular graphs. Molecular graphs were constructed from SMILES using RDKit, where each atom serves as a node and bonds represent edges. Node features include atomic number, valence, hybridization, and aromaticity, while edge features capture bond type, conjugation, and stereochemistry. These molecular graphs were then processed using GAT layers to enable hierarchical feature learning while preserving structural relationships within molecules.

This work uses artificial intelligence techniques to predict HIV inhibitors and provides insights into integrating molecular descriptors and graph-based features for improved classification accuracy. Major contributions of this study include:

- Analyzing the impact of molecular representation techniques (fingerprints and graphs) on predictive performance of HIV inhibitors classification.

- Comparing machine learning models (Random Forest, SVM, XGBoost, CatBoost, Stacking Ensemble) with deep learning architectures (Feedforward Neural Networks and Graph Attention Networks).
- Demonstrating that Graph Attention Networks and FNN meta learner-based Stacking Ensemble outperform traditional ML models in distinguishing active HIV inhibitors by leveraging molecular topology more effectively.
- Interpreting the decision-making process of the GAT model by analyzing attention scores to identify atoms that contribute most to molecular activity predictions.

By integrating AI-driven strategies with molecular modeling, this study underscores the transformative potential of ML and DL in accelerating HIV drug discovery. The insights from this research may contribute to more effective antiviral drug development, enhancing therapeutic innovation in combating HIV.

2 Literature Review

Several studies demonstrated the success of AI in drug repurposing and inhibitor prediction. For instance, Park [1] evaluated computational approaches for drug repurposing, demonstrating applications to speed up drug development at lower cost and reduced safety risks. Preclinical and clinical data served as a foundation for drug repurposing and helped speed up clinical trials and regulatory approval processes. The author established two computational drug repurposing paradigms through drug-centric and disease-centric methods based on drug-disease similarity comparisons. The research outlined the technological improvements that support this method through the extensive collection of data from genomics proteomics and cheminformatics platforms. The combination of specific methodologies reveals new medical applications for previously approved drugs, demonstrating the increasing value of computer-based approaches in current pharmacology practice. Aliper et al. [2] developed a deep learning approach to forecast drug pharmacological effects by analyzing transcriptomic

patterns. The researchers applied deep neural networks (DNNs) to process transcriptional results from 678 approved drugs obtained from the LINCS database. Through their pathway activation scoring methods, they established a connection between selected therapeutic use groups and expression profiles from cells. The deep learning model achieved excellent classification results while maintaining superior performance. This research showed how deep neural network confusion matrices offer drug repositioning potential through their capability to leverage deep learning as a strong method that finds new therapeutic applications for approved medications.

Tanoli et al. [3] applied machine learning approaches in drug repurposing focused on oncology cases. Computational models use massive data resources to discover new therapeutic applications for existing drugs, shortening and lowering the expenses that traditional drug development requires. Supervised ML methods operated through publicly accessible databases to forecast drug efficiency in treating cancer patients. The review emphasized the need for entirely designed target profile analyses to show drug interactions beyond set indications by including secondary effects with medical value. The review highlighted two substantial obstacles in AI-driven drug repurposing: restricted access to clinical data and the widespread utilization of genetic markers within present predictive systems. Choudhury et al. [4] provided a review of structure-based drug repurposing (SBDR), focusing on the integration between classic computational approaches and artificial intelligence (AI) techniques. Traditional drug discovery approaches such as high-throughput screening (HTS) remain expensive while also taking prolonged time durations, according to the authors. AI-based computational approaches that include deep learning and generative modeling provide fast drug screening methods for repurposable medications. According to this research, drug repurposing uses molecular docking, pharmacophore modeling, and molecular dynamics simulations to achieve better target identification and enhanced lead optimization. AI-based scoring functions obtained performance metrics that demonstrated

better predictive accuracy than traditional docking approaches, and deep learning-based methods further improved binding affinity predictions. The review establishes that AI integration with SBDR could speed up drug development processes while decreasing expenses and failed drug candidates.

Aittokallio [5] analyzed the major issues and restrictions that machine learning faces in pharmaceutical discovery and drug redeployment. The research demonstrated that biases within training data, insufficient clinical validation, and black box characteristics of deep learning models represented the main challenges. The author established that the ML-driven models demonstrated predictive capabilities for drug-target interactions. However, they struggle to translate these findings to practical applications because of poor interpretability and insufficient validated datasets. The preclinical performance metrics achieved high accuracy rates and recall performance but fell dramatically upon real-world testing. The authors underlined the need to combine multi-omics information sources because it will improve model reliability by implementing transparent explanation-based AI methods. The study suggested implementing a standardized testing framework for ML models to enhance their application in drug repositioning and make their outcomes more reliable. El-Behery et al. [6] established a machine learning model that predicts drug-target interactions (DTIs) specifically for drug repositioning applications in COVID-19 treatment. The authors used structured protein and drug information by analyzing amino acid sequences and SMILES representations for data extraction. Using the Extra Tree ensemble learning method, the authors attained 98% accuracy with an F-score of 0.97. The research measured drug-ACE2 receptor compatibility through COVID-19 protein analysis and produced correct predictions about this essential SARS-CoV-2 infection point. The study demonstrated that AIbased DTI predictions will boost drug repurposing operations by discovering treatment options for new diseases.

Dara et al. [7] analyzed the functions of machine learning in drug discovery throughout the drug development process, starting from target validation through biomarker identification and ending with lead optimization. The authors examined the application of both QSAR modeling and deep neural networks (DNNs) together with reinforcement learning approaches. The article showed that ML techniques delivered remarkable outcomes, reaching pharmacological prediction accuracy levels above 90%. The study analyzed limitations, which encompass insufficient data, bias within training data sets, and the challenge of explaining how AI models operate. The publication stated that machine learning demonstrates revolutionary potential within drug discovery, but organizations need solutions to resolve and implement its problems. Vatansever et al. [8] reviewed artificial intelligence models for central nervous system (CNS) disease drug discovery because this field traditionally has faced substantial failure rates. The system used AI predictions of drug-target interactions, QSAR modeling and virtual screening methods to boost CNS drug development. Deep learning technology brought significant advancements to blood-brain barrier permeability predictions, representing a critical barrier for CNS therapeutic development. The complexity of neurological disorders made it difficult for the authors to validate drugs generated through ML methods. This research established that AI/ML methods show promise for enhancing CNS drug discovery efficiency when proper solutions are developed to solve model interpretability and generalizability problems.

Lee and Chen [9] studied the drug repurposing possibilities for COVID-19 treatment using a deep learning model. Their workflow implemented state-of-the-art hyperparameter adjustment, optimized action functions, and innovative optimization systems to boost prediction precision. The enhanced deep learning system evolved from conventional black-box models into glass-box models, which allowed scientists to comprehend the drug repurposing prediction process. The study found that the advanced DL model made drug screening operations more effective because it helped researchers

spot suitable therapeutic candidates more quickly for COVID-19. This approach demonstrates its potential for scalability between complex diseases because the authors highlighted its extension to multiple complex diseases. Prasad et al. [10] examined specific AI-based drug repurposing techniques for treating SARS-CoV-2. A combined approach of molecular docking simulation and AI-based prediction assessments used CNNs and SVMs to discover potential drug candidates. Artificial intelligence docking approaches surpassed traditional drug screening methods because their AUC measurement increased by approximately 12%. According to the author's findings, AI-driven screening benefited from structural biology data integration because protein-ligand interactions boosted the precision of drug-target pair detections. Research results established that AI applied to therapeutic identification for COVID-19 and pandemic prevention processes function more efficiently than conventional approaches.

Mohapatra et al. [11] created a machine learning pipeline that identified potential medication candidates for treating COVID-19 from already existing pharmaceuticals. A Naïve Bayes algorithm in the study reached 73% accuracy for identifying FDA-approved drugs effective against SARS-CoV-2. ML predictions incorporated with molecular docking simulations led to the advancement of candidate selection methods used by the research team. Their research methodology resulted in the identification of Amprenavir, which demonstrated a potential antiviral substance. The research outcomes proved ML's capability to screen and rank repurposed medication efficiently through the findings but showed the necessity for additional scientific study of existing datasets and clinical confirmation possibilities. Choudhury et al. [12] analyzed structure-based drug repurposing (SBDR) by evaluating conventional computational techniques and AI/ML-based enhancements. There are multiple advantages of SBDR, according to this study, since the method requires less time and funding than traditional drug discovery. The authors highlighted the usefulness of molecular docking, pharmacophore modeling and molecular dynamics simulations as SBDR approaches

Table 1 Comparison of prior studies and how the present approach addresses existing gaps in HIV inhibitor prediction

Study	Focus	Key limitation(s)	How our study is better
Park [1], Tanoli et al. [3]	Computational drug repurposing	No comparative evaluation of automatic predictive models for HIV inhibitors	Directly compares fingerprint-based ML vs. graph-based DL models
Aliper et al. [2], Dara et al. [7]	Deep learning for drug discovery	Focused on transcriptomics and docking data, not molecular structure	Uses molecular structure-based ML/DL models for inhibitor prediction
Choudhury et al. [4, 12], El-Behery et al. [6]	Structure-based drug discovery	Did not compare fingerprint vs. graph-based representations	Evaluates both feature-based (fingerprints) and topology-based (graphs) approaches
Aittokallio [5], Vatansever et al. [8]	AI in drug discovery	Identified black-box issues in DL models, but lacked interpretability solutions	Improves interpretability using Graph Attention Networks for structured feature learning
Lee and Chen [9], Zeng et al. [10]	AI-driven drug repurposing for COVID-19	Limited focus on HIV-specific inhibitor prediction	Tailored to HIV inhibitors, using dataset balancing and advanced ML/DL models
Mohapatra et al. [11]	ML-based screening for drug repurposing	Used Naive Bayes and docking simulations, missing deep learning comparisons	Compares traditional ML models with advanced DL architectures

for discovering new applications for existing drugs. The prediction accuracy for drug-target interactions and binding affinities achieved the highest levels through deep learning models in AI/MLbased approaches. The review showed that a combination of AI-driven strategies with structural drug repurposing approaches would make drug discovery faster and improve therapeutic innovation.

Table 1 presents a comparative overview of related studies in computational drug discovery and HIV inhibitor prediction, highlighting their primary focus areas and key limitations. It also outlines how our study addresses these gaps by leveraging advanced machine learning and deep learning models, particularly for HIV inhibitor prediction using molecular structure-based representations. Several recent studies have explored

AI-driven HIV inhibitor prediction, but few have comprehensively compared ML and DL approaches. Our study directly addresses these gaps by:

- Providing a systematic comparison of feature-based ML and graph-based DL models for HIV inhibitor classification.
- Incorporating molecular topology in predictive modeling, improving upon previous studies that relied solely on feature-based descriptors.
- Addressing class imbalance in cheminformatics datasets, ensuring more reliable predictions of active inhibitors.

These contributions establish a more holistic AI-driven framework for molecular classification, setting the groundwork for future hybrid modeling techniques

3 Materials and methods

This section presents the details of the proposed workflow for HIV inhibitor prediction using two parallel AI-based pipelines—one employing fingerprint-based ML and FNN models, and the other utilizing graph-based deep learning with GAT—starting from SMILES data acquisition and preprocessing through model training, evaluation, and comparative performance analysis. Fig. 1 illustrates a structured workflow for predicting HIV inhibitors using AI-based automatic predictive models. The process begins with dataset acquisition, where molecular structures in the SMILES format and their inhibitory activity labels are obtained. Data cleaning is performed to remove invalid molecular structures. Two distinct computational pipelines are implemented: one for fingerprint-based ML models and another for graph-based deep learning models. In the fingerprint-based approach, Morgan fingerprints are generated to represent molecular structures numerically, followed by standardization and random oversampling to balance the dataset. These preprocessed fingerprints are then used to train traditional ML models and feedforward neural networks (FNNs). In the graph-based approach, molecules are converted into molecular graphs where atoms serve as nodes and bonds

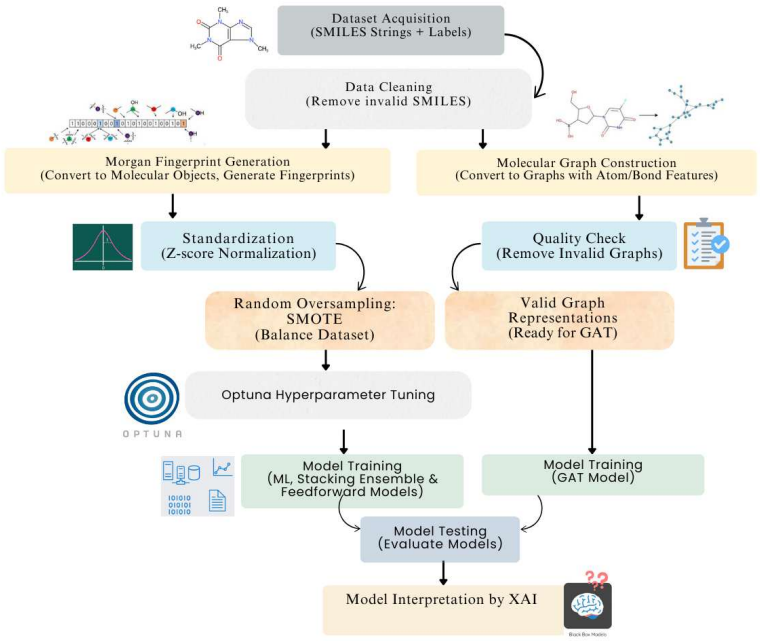


Fig. 1 Workflow of the proposed HIV inhibitor classification system

as edges. Quality checks are performed to remove invalid graphs before training a Graph Attention Network. Both models are tested and evaluated using various evaluation metrics. The final step compares the performance of ML, FNN, and GAT models to determine the most effective approach for HIV inhibitor prediction.

3.1 Dataset description

This work utilizes the HIV dataset as a molecular property prediction dataset that MoleculeNet [13] initially gathered for its benchmark platform regarding chemical and drug discovery work. The dataset maintains 41,127 molecular structures format using SMILES notation in addition to their bioactivity information. Research scientists classify molecules into two categories: those that inhibit HIV replication as active (1), while those that cannot inhibit it are inactive (0). Predictive modeling faces unique challenges when working with this dataset because active molecules represent only 3.51% of the total collection, while inactive samples comprise 96.49%. Fig. 2(a)

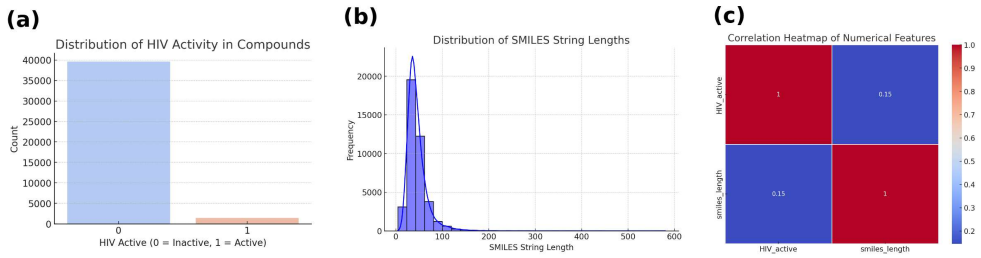


Fig. 2 Dataset characteristics: (a) distribution of HIV activity in compounds; (b) distribution of SMILES string lengths; (c) correlation heatmap of numerical features

shows a significant class imbalance in the dataset, with a much larger number of inactive molecules (label 0) compared to active HIV inhibitors (label 1). Specialized techniques, namely resampling or cost-sensitive learning, become essential to handle the class imbalance in this situation effectively. Fig. 2(b) illustrates that the majority of SMILES strings in the dataset have lengths between 20 and 80 characters, with a right-skewed distribution and a long tail extending beyond 100 characters. A weak positive correlation (0.15) between SMILES string length and HIV activity, indicating limited linear dependence between these two features, is demonstrated in Fig. 2(c).

3.2 Data preprocessing

In this work, two separate preprocessing approaches were developed to meet the applied machine learning and deep learning model requirements. Morgan fingerprinting [18] was used for regular machine learning and feedforward neural network models, but graph construction [19] was employed for graph attention networks. Each preprocessing pipeline ensured the molecular structures were transformed into formats optimized for their respective modeling approaches.

3.2.1 Preprocessing for machine learning and feedforward neural networks

For machine learning and feedforward neural networks, molecular structures were transformed into numerical representations using Morgan fingerprints, depicted in

Fig. 3(a). Initially, SMILES (Simplified Molecular Input Line Entry System) strings were converted into molecular objects via RDKit. Subsequently, Morgan fingerprints were generated with a radius of 2 and 2048 bits to capture the local chemical environment around each atom. This fingerprinting process encoded structural information into a fixed-length bit vector that could be directly fed into machine learning and deep learning pipelines.

To ensure the effectiveness of the training process, these fingerprint vectors were standardized using z-score normalization. Standardization adjusted the feature values to have zero mean and unit variance, improving the convergence of gradient-based optimization algorithms. Any molecule that could not be processed or lacked a valid structure was removed from the dataset to ensure data quality.

3.2.2 Preprocessing for Graph Attention Networks

For GAT models, molecular structures were represented as graphs. In this step, each atom in a molecule was treated as a node, and each bond as an edge. Node features were constructed to encode atomic properties such as atomic number, degree, total hydrogens, valence, aromaticity, hybridization state, formal charge, and other chemically relevant attributes. Edge features were similarly enriched with bond-specific characteristics, including bond type, conjugation, stereochemistry, and ring membership. These detailed features provided a rich representation of the molecular structure, enabling the model to better learn the underlying chemical relationships graph.

According to Fig. 3(b), the graph representations were constructed using RDKit and formatted into PyTorch geometric data objects comprising node features, edge indices, and edge attributes. Molecules that could not be converted into valid graphs due to structural inconsistencies were excluded from the dataset. This graph-based approach captured both the topological and chemical properties of molecules, making it ideal for GAT models.

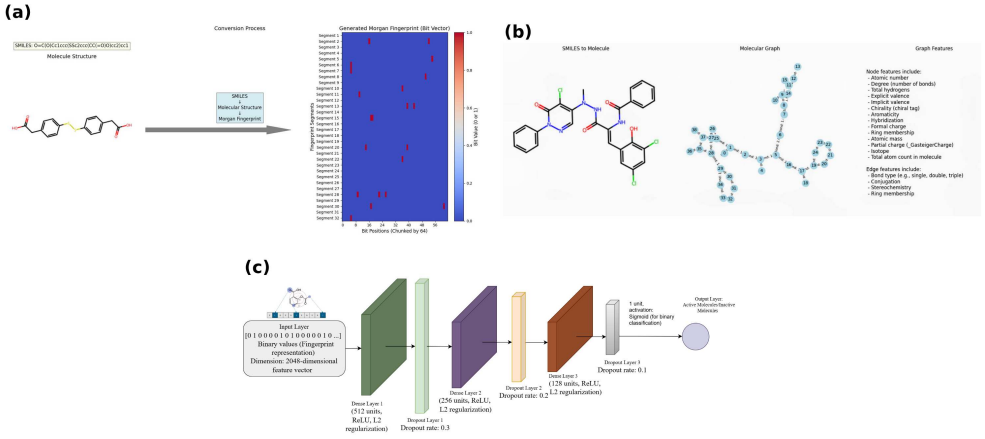


Fig. 3 Molecular preprocessing and neural architecture: (a) conversion of SMILES to Morgan fingerprints; (b) conversion of SMILES to molecular graphs with atom and bond features; (c) proposed feedforward neural network architecture

3.3 Data balancing, augmentation, and standardization strategies

Our investigation of data balancing methods included both the SMOTE (Synthetic Minority Oversampling Technique) [20] approach together with random oversampling techniques. The first distribution of our data showed major disproportion between active and inactive molecules which created biased prediction outcomes. Random oversampling methods combined with undersampling techniques of the majority class produced the most effective solutions to enhance model generalization. Random oversampling achieved class balance by duplicating minority-class (active molecule) samples thus it reduced bias and enabled pattern learning across all classes. We examined random under sampling of inactive molecules to decrease the dataset size while upholding functional chemical features. This hybrid approach minimized the risk of synthetic bias introduced by SMOTE while preventing the model from overfitting trivial inactive molecular patterns. Finally, we standardized numerical input attributes

to stabilize training across both machine learning and neural network models. Standardization transformed features to have zero-mean and unit variance, enhancing optimization efficiency and preventing issues arising from extreme feature magnitudes.

3.3.1 Model optimization through automated hyperparameter tuning

To enhance the predictive performance and generalization ability of our machine learning models, we employed Optuna [21], a state-of-the-art hyperparameter optimization framework. Optuna efficiently explores the hyperparameter search space through its Tree-structured Parzen Estimator (TPE), an adaptive Bayesian optimization technique [22] that balances exploration and exploitation. The optimization process aims to maximize F1-score because it strikes a perfect balance between precision and recall when dealing with class imbalance tasks. The optimization process was conducted over 70 trials per model, using random oversampling (ROS) to mitigate class imbalance during training.

3.4 Applied model architectures

3.4.1 Feedforward Neural Network (FNN)

The feedforward neural network [23] is a powerful architecture that processes numerical molecular representations originating from Morgan fingerprinting due to its fixed vector length. This model starts by accepting fingerprint vectors of 2048 features through its input layer because standardized bit vectors follow this dimension. The model included three hidden layers for pattern detection with 512, 256, and 128 neurons in sequence. Accompanying each hidden layer, a ReLU activation function brought non-linearity to the model because it enables the discovery of complex input feature associations. L2 regularization strength of 0.002 was applied to every dense layer to stop overfitting and improve generalized performance. Each dense layer included

dropout operators with progressively decreasing rates of 0.3, 0.2, and 0.1. During training, the algorithm selects dropout layers to eliminate random groups of neurons, which prevent patterns from developing based on particular features in the information. The network issued its conclusion through a single neuron employing a sigmoid activation, producing binary outcomes to determine whether molecules were active or inactive. The applied FNN model achieved optimal performance by optimizing learning capacity and computation efficiency, allowing the system to handle standard fingerprint data vectors. FNN's modular design used standardized input features to handle detailed chemical fingerprints and maintain stability during gradient optimization. The FNN presented itself as the best pick for evaluating molecular properties during deep learning pipelines which use numerical descriptors for predictive tasks. Fig. 3(c) illustrates the architecture of the proposed feedforward neural network for HIV inhibitor prediction, comprising three dense layers with ReLU activation and L2 regularization, interleaved with dropout layers (rates: 0.3, 0.2, 0.1), and concluding with a sigmoid-activated output layer for binary classification.

3.4.2 Improved Graph Attention Network (GAT)

GAT architecture [24] obtains improved functionality by utilizing molecular graph properties for enhanced molecular property learning. The model represents molecules through graphs with atoms acting as nodes and bonds functioning as edges together with chemical property inclusion. As illustrated in Table 2, the convolutional layers utilize a multi-head attention mechanism to extract intricate molecular substructures and feature relationships. The first convolutional layer employs 8 attention heads, allowing the model to capture multiple sub-structure interactions and feature dependencies. This layer expands the input characteristics into a higher-dimensional space, enabling the model to learn crucial chemical interactions. Batch normalization stabilizes the feature distribution before passing them through a LeakyReLU activation

Table 2 Proposed Graph Attention Network architecture

Layer	Description
GAT 1	8 heads, 64 hidden, LeakyReLU ($\alpha = 0.2$)
Dropout 1	$p = 0.3$
GAT 2	8 heads, 64 hidden, LeakyReLU ($\alpha = 0.2$)
Dropout 2	$p = 0.2$
GAT 3	1 head, 32 hidden, LeakyReLU ($\alpha = 0.2$)
Dropout 3	$p = 0.1$
Output	Softmax activation

(α=0.2) to introduce non-linearity. Four attention heads support the second convolutional layer to refine gained representations while keeping the processing efficient. The third convolutional layer implements one attention head and a hidden size of 64 to optimize features before creating a shortened chemical representation. A global mean pooling layer aggregates individual node embeddings into a graph-level feature vector, effectively capturing local and global molecular structures. The final linear fully connected layer classifies the molecule into binary categories, followed by a dropout layer (p=0.2) to improve generalization and reduce overfitting. The advanced GAT model utilizes attention mechanisms to extract hierarchical features that comprehensibly construct chemical and structural connections within molecular graphs. The model improves its predictive power by studying essential atomic elements and bond relationships, which makes it appropriate for works that need chemical integration with topological analysis.

3.4.3 Hybrid ChemBERTa–GAT with fingerprint cross-attention fusion

A hybrid architecture was developed by combining molecular graphs (processed through GAT layers) with SMILES embeddings (from ChemBERTa). Both embeddings were projected to a 256-dimensional space and jointly processed through a CrossAttention module, where Morgan Fingerprints served as the query vector. Gated

additive fusion was used to merge these features into a unified representation. The final classification head used dense layers to predict inhibitor activity. This architecture captures sequential and structural molecular characteristics simultaneously.

3.5 Stacking ensemble and dual-encoder late fusion approach

To further enhance predictive performance, we implement a stacking ensemble learning [25] framework that integrates the four top-performing machine learning models (Random Forest, LightGBM, CatBoost, and XGBoost) as base learners. A Feedforward Neural Network serves as the meta-model to refine and aggregate predictions for improved classification accuracy. The FNN model processes feature-based outputs from the base models, optimizing decision boundaries for improved predictions. Each base learner is trained on molecular data and generates out-of-fold (OOF) predictions. The corresponding OOF predictions are used as input features for the meta-learner. The meta-learner processes these predictions and learns to refine the classification, leading to a final, more robust prediction. Integrating fingerprint-based ML models with FNN-based meta-learning provides a structured approach that effectively captures molecular relationships for HIV inhibitor prediction. Our findings suggest that stacking ML and DL models enhances performance, making it a promising methodology for future drug discovery tasks. Additionally, a deep dual-encoder framework was proposed, comprising Morgan fingerprint and GAT+ChemBERTa encoders. The model employed CrossAttention to enable one encoder (fingerprint MLP) to query the context of the fused SMILES-graph embedding. A gating mechanism was used to merge contextual and query representations before passing them to a final classification head. This fusion enhanced prediction by combining interpretable descriptor-level features and learned graph-structured knowledge.

3.6 Evaluation metrics

The evaluation metrics used to measure the effectiveness of both fingerprint-based machine learning (ML) models and graph-based deep learning (DL) models consist of the following.

The percentage of correctly identified instances makes up the accuracy metric out of all analyzed instances:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Instances}}. \tag{1}$$

where:

- TP (True Positives) = correctly predicted active inhibitors;
- TN (True Negatives) = correctly predicted inactive compounds;
- FP (False Positives) = incorrectly predicted active inhibitors;
- FN (False Negatives) = incorrectly predicted inactive compounds.

Precision evaluates the proportion of correctly predicted positive samples out of all predicted positives:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}. \tag{2}$$

Drug discovery depends on high precision because it reduces the number of incorrect detections of inactive compounds which appear as potential inhibitors.

Recall measures the proportion of actual positive cases that were correctly identified:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}. \tag{3}$$

A high recall ensures that most active HIV inhibitors are correctly identified, even if it means tolerating some false positives.

The F1-score is the harmonic mean of precision and recall, providing a balanced metric when there is class imbalance in the dataset:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{4}$$

A higher F1-score indicates a better trade-off between precision and recall. The ROC curve plots True Positive Rate (TPR) against False Positive Rate (FPR):

$$\text{ROC-AUC Score} = \int_0^1 \text{ROC Curve}\, d(\text{False Positive Rate}). \tag{5}$$

The AUC (Area Under the Curve) represents the probability that the model ranks a random positive sample higher than a random negative sample. A higher AUC (closer to 1) indicates better model discrimination between active and inactive compounds. The False Positive Rate measures an extremely small change in the False Positive Rate statistics. In calculus notation the differential element determines an extremely tiny modification of False Positive Rate. The calculation of area under the ROC curve results from summing numerous tiny increments produced during the integration process where the model distinguishes between classes at various threshold points.

The Precision–Recall AUC (PR-AUC) was also used as a critical evaluation metric due to the imbalanced nature of the dataset, where active HIV inhibitors constitute only 3.51% of the samples. The PR-AUC quantifies the area under the Precision–Recall curve, which plots precision (positive predictive value) against recall (sensitivity) at various thresholds. It is particularly informative when the positive class is rare, as it emphasizes the model's performance in correctly identifying active inhibitors. A higher PR-AUC indicates stronger detection capability for minority class instances, making it a more reliable metric than ROC-AUC in this context.

The Precision–Recall curve provides a visual understanding of how the model balances false positives and false negatives across different thresholds, supplementing the insight offered by ROC curves. Due to the imbalanced nature of the dataset, we additionally evaluated the hybrid model using the Precision–Recall AUC metric and PR curve, which are more informative than ROC curves for rare positive class detection. Other baseline models were assessed primarily through ROC-AUC and macro F1-score.

4 Results

The predictive capability of the applied FNN network reached 0.91 accuracy, indicating a stable ratio of correct to false positives and actual positives. This metric shows promising results in maintaining effective classification for all classes because the macro F1 score reached 0.80 and the ROC-AUC score measured 0.83, illustrated in Fig. 4(a).

Fig. 4(b) shows the ROC curves of the applied models for HIV inhibitor prediction, with XGBoost, LightGBM, and SVM achieving the highest AUC scores of 0.85. The curve indicates strong discriminative ability, with XGBoost demonstrating robust performance in distinguishing between active and inactive compounds. Fig. 4(b): ROC Curve of the Applied Models The GAT model has been implemented in this work specifically to extract better molecular graph patterns from chemical structures. The GAT model demonstrated accuracy and F1 score of 0.81 and 0.7884, respectively, representing effective active and inactive molecule identification.

The traditional machine learning models, Random Forest and SVM, recorded high accuracy, while GAT emerged as an efficient model due to its strong ROC-AUC score of 0.8534, depicted in Fig. 4(c). The Stacking Ensemble model, however, achieved the highest accuracy (0.9275) and ROC-AUC score (0.8652), demonstrating the advantages of combining multiple models for enhanced classification performance. The proposed GAT network excelled in molecular activity classification discrimination, while the accuracy leader role remained with tree-based approaches (Random

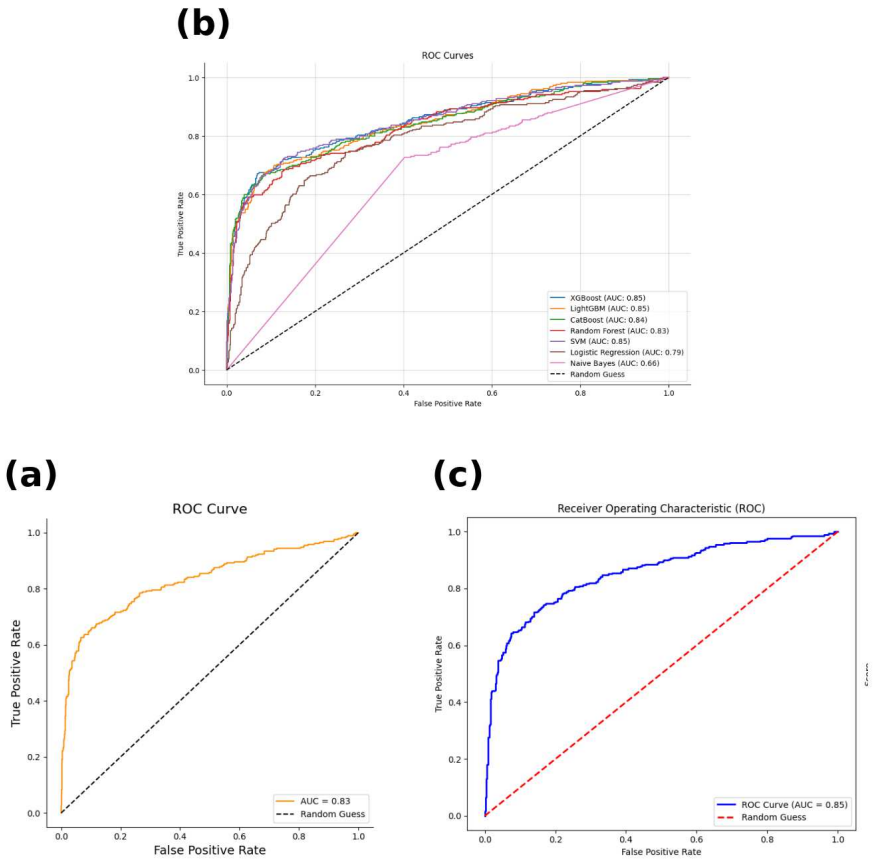


Fig. 4 Baseline model discrimination: (a) ROC curve of the FNN; (b) ROC curves of the applied machine-learning models; (c) ROC curve of the GAT

Forest, CatBoost, SVM). The Stacking Ensemble's (FNN) superior performance further emphasizes the advantage of leveraging multiple models to achieve exceptional performance in HIV inhibitor classification.

4.1 Performance of the hybrid ChemBERTa–GAT–fingerprint model

To address class imbalance and enhance feature representation, we developed a novel hybrid deep learning model that integrates ChemBERTa-based SMILES encodings,

Table 3 Performance metrics of the applied models in the proposed HIV inhibitor prediction

Model	Accuracy	Precision (Macro)	Recall (Macro)	F1 Score (Macro)	ROC AUC	PR AUC
XGBoost	0.8567	0.8293	0.7876	0.8045	0.8508	–
LightGBM	0.8531	0.8212	0.7873	0.8015	0.8452	–
CatBoost	0.8558	0.8332	0.7792	0.7998	0.8415	–
Random Forest	0.8549	0.8402	0.7686	0.7938	0.8294	–
SVM	0.8531	0.8383	0.7652	0.7906	0.8497	–
Logistic Regression	0.7769	0.7180	0.7354	0.7252	0.7855	–
Naive Bayes	0.6134	0.6184	0.6518	0.5930	0.6605	–
GAT	0.8100	0.7985	0.7818	0.7884	0.8534	–
Stacking Ensemble (FNN Meta Learner)	0.8521	0.8144	0.7977	0.8053	0.8499	–
Hybrid GAT + ChemBERTa + Fingerprint (Cross-Attention Fusion)	0.8200	0.8100	0.7990	0.8030	0.8400	0.816

GAT-derived graph features, and Morgan fingerprints using a cross-attention fusion mechanism.

Due to the imbalanced nature of the dataset, we additionally evaluated the hybrid model using the Precision–Recall AUC (PR-AUC), which is more informative than ROCAUC for rare class detection. Other baseline models were primarily assessed using ROC-AUC and macro F1-score. The hybrid model achieved strong performance with a ROC-AUC of 0.840 and a PR-AUC of 0.816, as shown in Fig. 5(a) and Fig. 5(b), respectively. The confusion matrix (Fig. 5(c)) highlights that the model correctly identified 104 active and 220 inactive compounds, with relatively low misclassifications (30 false positives and 41 false negatives). Table 3 summarizes the performance metrics of various machine learning and deep learning models applied to HIV inhibitor prediction.

The Stacking Ensemble model demonstrates strong classification performance, correctly identifying the most negative and positive samples. However, some false negatives (FN = 101) indicate that further improvements in the recall may be necessary to minimize the misclassification of positive instances. The confusion matrix in Fig. 5(d) of the FNN meta learner-based Stacking model shows that it correctly classified 739 inactive (true negatives) and 188 active (true positives) compounds. However, it also misclassified 61 inactive compounds as active (false positives) and 101 active compounds as inactive (false negatives).

Table 4 presents a comparative overview of existing studies on HIV inhibitor prediction, highlighting their methodologies, best-performing models, and key performance metrics. While earlier works employed traditional machine learning or standalone

Table 4 Performance comparison of HIV inhibitor classification models

Study	Approach	Best-performing model	Accuracy	ROC-AUC
[11]	Machine Learning for HIV Drug Repurposing	Random Forest	0.73	0.78
[12]	AI-Based HIV Drug Repurposing	Deep Neural Networks (DNNs)	0.88	0.84
[19]	Graph-Based HIV Drug Discovery	Graph Convolutional Network (GCN)	0.82	0.83
[26]	HIV Drug Prediction using Deep Learning	Deep Belief Networks (DBN)	0.85	0.81
This work	HIV Drug Prediction using Ensemble Model with Deep Learning-based Meta Learner and GAT Neural Network	Stacking Ensemble (FNN Meta Learner)	0.8521	0.8499

deep learning models, they often lacked hybrid ensemble strategies. Our proposed approach, which integrates deep learning model through a Stacking Ensemble with an FNN meta learner, achieved competitive accuracy (0.8521) and the highest ROC-AUC (0.8499), demonstrating its effectiveness in improving predictive performance and molecular classification. To interpret the decision-making process of the Graph Attention Network, we analyzed the attention scores assigned to molecular graph nodes. The attention mechanism in GAT assigns higher weights to atoms that are deemed important for the model's prediction, providing a transparent and interpretable framework for molecular property analysis.

Fig. 6(a) shows a GAT-based attention heatmap for an active compound (Label = 1), where each node represents an atom, and the color intensity indicates its attention score. Higher attention scores (in red) highlight atoms that contributed more significantly to the model's prediction, revealing important structural features for activity classification. Fig. 6(b): Molecular Structure Visualization with Atom-Level Attention Scores from GAT Model Fig. 6(b) visualizes the molecular structure of a selected sample with color-coded atomic importance derived from GAT attention scores. Atoms receiving higher attention are highlighted in red, while those with lower attention appear in blue. This visualization directly links deep

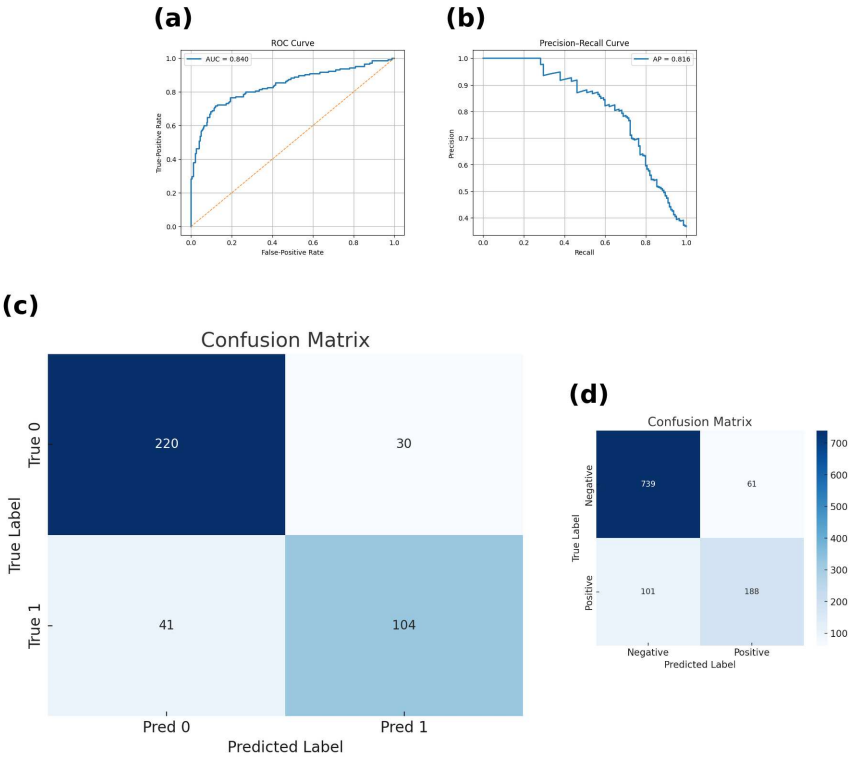


Fig. 5 Hybrid and stacking model performance: (a) ROC curve of the hybrid GAT–ChemBERTa–fingerprint model; (b) Precision–Recall curve of the hybrid model; (c) confusion matrix of the hybrid model; (d) confusion matrix of the FNN meta-learner-based stacking model

learning-derived feature importance to chemically meaningful structures. Notably, high-attention regions are concentrated around, i.e., electronegative atoms (O, N), which are often involved in hydrogen bonding, aromatic rings, commonly important for binding interactions, and functional groups contributing to molecular stability and reactivity. These findings align with known pharmacophoric features, indicating that the GAT model captures interpretable structure-activity relationships (SARs) rather than learning from spurious correlations. Our Explainable AI (XAI) approach using GAT enhances interpretability by attributing predictive importance to key atomic features, thus improving trust, scientific validity, and the practical applicability of AI-driven molecular predictions in drug discovery.

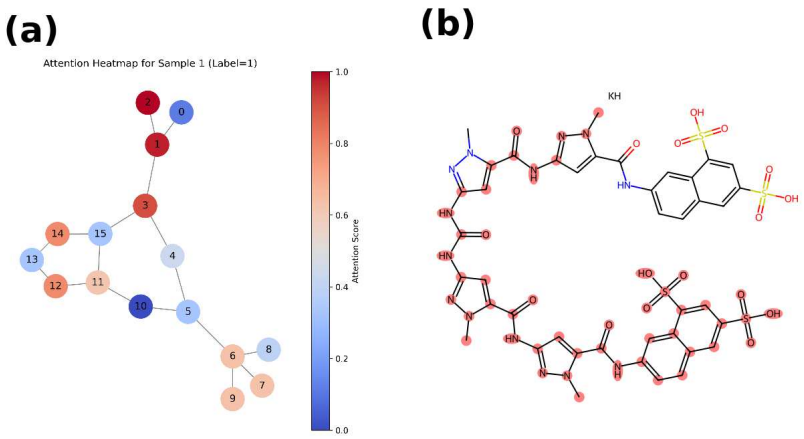


Fig. 6 GAT interpretability: (a) attention heatmap for an active compound (label = 1); (b) molecular structure visualization with atom-level attention scores

5 Discussion

These study findings demonstrate the respective features which characterize fingerprint-based machine learning models and graph-based deep learning models used for HIV inhibitor classification. The FNN meta learner-based Stacking Ensemble model achieved outstanding results with an accuracy of 0.8521 and ROC-AUC of 0.8499, which shows its potential for molecular property prediction. GATs demonstrated improved performance per structural interaction in molecules, producing high ROC-AUC results (0.8533) even at a lower accuracy level (0.8061). The application of these modeling and simulation methodologies provides valuable insights for the engineering of effective HIV inhibitors.

Virtual screening benefits from ML models built with features since these models achieve superior performance and handle data efficiently while providing clear analysis, even though graphic structure learning capabilities in graphical models produce better outcomes in complex structure-activity relationships (SAR) prediction. The cheminformatics research using deep learning presents similarities to Aliper et al.'s

study [2] about deep learning systems handling transcriptomic data. The analysis indicates that graph-based models contribute to fingerprint-based learning by creating a complementary instead of a competitive relationship between the two techniques.

By implementing data balancing techniques, our study improves recall without sacrificing overall model performance, ensuring a higher rate of correctly identified active inhibitors. The trade-off between precision and recall observed particularly in deep learning models, suggests that future efforts should focus on hybrid approaches combining feature-based and structural representations to optimize classification across both classes.

The study focused on HIV-related molecular structures, which may limit generalizability to other antiviral drug discovery applications. Expanding to broader chemical libraries could enhance model robustness. Although GAT models capture molecular interactions effectively, their black-box nature makes interpretation challenging. Future studies should explore attention visualization techniques to improve explainability. The results suggest that combining fingerprint-based and graph-based learning could yield higher accuracy. Future work should investigate hybrid architecture that integrates numerical descriptors and structural graphs. Utilizing pre-trained deep learning models on larger molecular datasets could improve predictive accuracy and generalization to new drug classes. Future research can further enhance AI-driven inhibitor discovery by addressing these areas, making cheminformatics-based screening more efficient, interpretable, and generalizable.

6 Conclusion

This study investigates a systematic evaluation of automatic artificial intelligence models for HIV inhibitor prediction. The HIV dataset has been used from MoleculeNet, which contains 41,127 SMILES-encoded molecular structures labeled as active (1) or inactive (0) based on their ability to inhibit HIV replication. Due to a severe class

imbalance—only 3.51% active molecules—specialized techniques like resampling or cost-sensitive learning are necessary to ensure effective predictive modeling. A Stacking Ensemble framework was developed by combining four top-performing ML models as base learners and using a Feedforward Neural Network as the meta-learner to enhance classification accuracy. The FNN meta learner-based Stacking Ensemble achieved the highest performance with 0.8521 accuracy and 0.8499 ROC-AUC. The GAT's attention mechanism enabled interpretable predictions by highlighting key atoms linked to known pharmacophoric features, reinforcing its reliability for AI-driven drug discovery. The major obstacle handled in this analysis involves the class imbalance found in HIV data because active inhibitors exist in smaller numbers. The implementation of data balancing techniques employing the SMOTE technique solved this problem and simultaneously enhanced precision while avoiding any discarding of potentially useful compounds. A fundamental modification of classification thresholds needs to happen to decrease the number of incorrect predictions because precision-recall relations remain at inconvenient levels. To address this, we additionally employed the Precision–Recall AUC metric and PR curve for the hybrid ChemBERTa–GAT– Fingerprint model, as they provide a more informative measure for imbalanced datasets compared to ROC curves.

Research findings about molecular representation support the fundamental aspects of cheminformatics. Molecular discovery applications benefit from using ML models based on fingerprints because they have high performance and interpretability features alongside graph-based models that capture complex molecular structures that standard descriptors fail to capture. Drug discovery models that merge these two methods demonstrate better performance for prediction and generalization activities in pharmaceutical research.

The research establishes performance criteria for HIV inhibitor AI classification by analyzing different predictive computational approaches. The research discoveries

enable progress in AI technologies for cheminformatics alongside developing extensive new methods for efficient drug discovery processes. In the future, the combination of ChemBERTa and Graphormer transformers with Graph Neural Networks, especially GIN neural networks, can be applied to molecular classification techniques.

Statements and Declarations.

Funding. The authors did not receive support from any organization for the submitted work.

Competing interests. The authors declare no competing interests relevant to the content of this article.

Ethics approval. Not applicable.

Consent to participate. Not applicable.

Data availability. This work utilizes the MoleculeNet HIV dataset, available at https://moleculenet.org/datasets-1.

References

[1] K. Park, A Review of Computational Drug Repurposing, Translational Clinical Pharmacology, vol. 27, 2019.

[2] A. Aliper, S. Plis, A. Artemov, A. Ulloa, P. Mamoshina, and A. Zhavoronkov, Deep Learning Applications for Predicting Pharmacological Properties of Drugs and Drug Repurposing Using Transcriptomic Data, Molecular Pharmaceutics, vol. 13, pp. 2524–2530, 2016.

[3] Z. Tanoli, M. Vähä-Koskela, and T. Aittokallio, Artificial Intelligence, Machine Learning, and Drug Repurposing in Cancer, Expert Opinion on Drug Discovery, vol. 16, pp. 977–989, 2021.

[4] H. El-Behery, A.-F. Attia, N. El-Fishawy, and H. Torkey, Efficient Machine Learning Model for Predicting Drug-Target Interactions with Case Study for COVID-19, Computational Biology and Chemistry, vol. 93, 2021.

[5] S. Dara, S. Dhamercherla, S. S. Jadav, C. M. Babu, and M. J. Ahsan, Machine Learning in Drug Discovery: A Review, Artificial Intelligence Review, vol. 55, pp. 1947–1999, 2022.

[6] S. Vatansever, A. Schlessinger, D. Wacker, H. Ü. Kaniskan, J. Jin, M.-M. Zhou, and B. Zhang, Artificial Intelligence and Machine Learning-Aided Drug Discovery in Central Nervous System Diseases: State-of-the-Arts and Future Directions, Medicinal Research Reviews, vol. 41, pp. 1427–1473, 2020.

[7] C. Y. Lee and Y.-P. P. Chen, New Insights into Drug Repurposing for COVID-19 Using Deep Learning, IEEE Transactions on Neural Networks and Learning Systems, vol. 32, pp. 4770–4782, 2021.

[8] X. Zeng, Y. Zhu, R. Wang, J. Ma, and Y. Kang, Artificial Intelligence-Driven Drug Repurposing and Structural Biology for SARS-CoV-2, Frontiers in Pharmacology, vol. 12, 2021.

[9] S. Mohapatra, P. Nath, M. Chatterjee, N. Das, D. Kalita, and P. Roy, Repurposing Therapeutics for COVID-19: Rapid Prediction of Commercially Available Drugs Through Machine Learning and Docking, PLOS ONE, vol. 15, 2020.

[10] K. Prasad and V. Kumar, Artificial intelligence-driven drug repurposing and structural biology for SARS-CoV-2, Current Research in Pharmacology and Drug Discovery, vol. 2, 2021.

[11] T. Aittokallio, What Are the Current Challenges for Machine Learning in Drug Discovery and Repurposing? Expert Opinion on Drug Discovery, vol. 17, pp. 423–

425, 2022.

[12] C. Choudhury, N. A. Murugan, and U. D. Priyakumar, Structure-Based Drug Repurposing: Traditional and Advanced AI/ML-Aided Methods, Drug Discovery Today, vol. 27, pp. 1847–1859, 2022.

[13] MoleculeNet, Datasets, MoleculeNet, Accessed: Feb. 2, 2025. [Online]. Available: https://moleculenet.org/datasets-1.

[14] H. M. Swinkels, A. A. J. Vaillant, A. D. Nguyen and P. G. Gulick, HIV and AIDS, Treasure Island, StatPearls Publishing, 2025.

[15] S. Deeks, J. Overbaugh, A. Phillips, et al., HIV infection, Nature Reviews Disease Primers, 2015.

[16] T. Hatziioannou and D. Evans, Animal models for HIV/AIDS research, Nature Reviews Microbiology, vol. 10, pp. 852–867, 2012.

[17] G. Sliwoski, S. Kothiwale, J. Meiler and E. W. Lowe, Computational methods in drug discovery, Pharmacological Reviews, vol. 66, pp. 334-95, 2014.

[18] S. Zhong and X. Guan, Count-Based Morgan Fingerprint: A More Efficient and Interpretable Molecular Representation in Developing Machine Learning-Based Predictive Regression Models for Water Contaminants' Activities and Properties, Environmental Science and Technology, 2023,

[19] C. Yao, H. Huang, H. Gao, F. Wu, H. Chen, and J. Zhao, Molecular Graph Representation Learning via Structural Similarity Information, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, pp. 351–367, 2024.

[20] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: synthetic minority over-sampling technique, Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.

[21] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, Optuna: A Next-generation Hyperparameter Optimization Framework, International Conference on Knowledge Discovery & Data Mining, pp. 2623–2631, 2019.

[22] G. Han, J. Jeong and J. H. Kim, Adaptive Bayesian Optimization for Fast Exploration Under Safety Constraints, IEEE Access, vol. 11, pp. 42949-42969, 2023.

[23] Z. Ji, L. Chen, M. Qin and X. Chen, Utilizing Multiple-Access Communication to Realize FNN Computation in Wireless Networks, IEEE Access, vol. 7, pp. 32143223, 2019.

[24] G. Su, H. Wang, Y. Zhang, W. Zhang and X. Lin, Simple and deep graph attention networks, Knowledge-Based Systems, vol. 293, 2024.

[25] M. F. Hossain, S. T. Diya and R. Khan, ACD-ML: Advanced CKD detection using machine learning: A tri-phase ensemble and multi-layered stacking and blending approach, Computer Methods and Programs in Biomedicine Update, vol. 7, 2025.

[26] A. Ezzat, M. Wu, X. Li, and C. K. Xiaoli, Computational prediction of drug-target interactions using chemogenomic approaches: an empirical survey, Briefings in Bioinformatics, vol. 20, 2018.