Intelligent Systems Research and Applications Journal

www.israj.org
eISSN: 3120 - 0826

Explainable Machine Learning-Guided Virtual Screening and Molecular Docking for Identification of Novel FYN Kinase Inhibitors

Ahmet Turan Demir[1, *]

[1] Department of Biomaterials and Tissue Engineering, Institute of Graduate Studies, Tokat Gaziosmanpaşa University, 60250 Tokat, Türkiye

ARTICLE INFO

Article history:
Received 2 May 2026
Received in revised form 23 June 2026
Accepted 1 July 2026
Available online 7 July 2026

Keywords:
FYN kinase; Explainable machine learning; ECFP4; SHAP; Molecular docking: AutoDock Vina; Virtual screening; Drug discovery

ABSTRACT

FYN kinase is a non-receptor protein tyrosine kinase involved in various cancers and neurodegenerative diseases; however, no selective FYN inhibitor has been approved yet. Here we introduce the explainable Machine Learning (ML) coupled with virtual screening and Molecular Docking (MD) pipeline for fast prediction of new FYN kinase inhibitors. In this study, we constructed the training set of 906 molecules active against FYN kinase from the ChEMBL database. Molecules were encoded with Extended-Connectivity Fingerprints (ECFP4). The classification models Random Forest (RF) and eXtreme Gradient Boosting (XGBoost) were developed, and the latter showed the better performance in test (AUC=0.8118) and 5-fold Cross validation (CV) (AUC=0.8297). Based on the SHapley Additive exPlanations (SHAP) values obtained via TreeExplainer, nitrogen-containing heterocycles and Hydrogen Bond Acceptor (HBA)s have been identified as the most important molecular substructures. Using the optimal XGBoost classifier, screening of 2,000 approved drugs has been performed, resulting in 470 hit molecules (23.5% hit rate). Five best molecules were further submitted to the MD procedure using AutoDock Vina to dock to FYN kinase domain (PDB RCSB: 2DQ7), showing binding energies in the interval of -9.57 to -6.32 kcal/mol. Dasatinib Anhydrous (CHEMBL1421) was the second strongest binder (-8.49 kcal/mol), effectively interacting with the ATP binding site. Although CHEMBL1171837 was the strongest binder (-9.57 kcal/mol), it was caught in the Absorption, Distribution, Metabolism, Excretion, Toxicity (ADMET) profiling. According to ADMET profiling, the top one inhibitor (CHEMBL1421) satisfies Lipinski's rule of five and Veber rules. Analysis of hydrogen bond and hydrophobic interactions revealed hydrogen bonding with ASP148, LYS39, and ASN86 and hydrophobic interactions with ALA147, ILE80, and GLY88. Validation by self-docking procedure (self-docking or STS) showed low Root Mean Square Deviation (RMSD)<2.0 Å with a binding affinity of -11.53 kcal/mol. This work highlights how explainable ML can be used in combination with structure-based docking to expedite the drug discovery process against FYN kinase and can be applied to other kinase targets.

1. Introduction

The FYN is a non-receptor protein tyrosine kinase with a molecular weight of 59 kDa, belonging to the family of Src kinases that participate in various intracellular signaling pathways involving cell

* *Corresponding author.*
E-mail address: phddemirahmet@gmail.com
https://doi.org/10.59543/nd9y7y92

proliferation, differentiation, adhesion, and apoptosis [1]. Unlike other SFKs that are found ubiquitously such as SRC and YES, the biological function of FYN is tissue-specific and acts in the brain, hematopoietic cells, and epithelial tissue [2]. Dysregulated FYN expression is well established in several human cancers, including melanoma [3], glioblastoma [4], peripheral T-cell lymphoma [5], breast carcinoma [6], and pancreatic adenocarcinoma [7]. For these cancers, FYN acts as an oncoprotein due to its ability to stimulate STAT3 signaling, CD147 phosphorylation, and NF-κB through the action of TRAF3IP2 [3,5,8].

FYN is not only involved in oncogenesis but is also critical for the neurodegenerative process. It plays a role by phosphorylating tau at Tyr18; Tyr18-phosphorylated tau has been detected in the neurofibrillary tangles found in patients with Alzheimer's Disease [9]. Furthermore, FYN phosphorylates the amyloid precursor protein (APP) at Tyr682, thus relating amyloid-β to synaptic problems [10]. FYN has also been shown to be involved in the development of neuroinflammation and L-DOPA-induced dyskinesia in Parkinson's disease [11]. FYN inhibition in preclinical trials has been associated with a decrease in the problems with tau, synaptic function maintenance, and improved cognition in tauopathy models [12]. In the recent discovery of the crystal structure of the FYN kinase domain together with saracatinib, atomic details of inhibition can be obtained for further structure-based drug discovery [13].

While FYN is a critical drug target, there is yet to be a selective FYN inhibitor in clinical use. The Src/FYN dual inhibitor, saracatinib (AZD0530) failed several clinical trials due to toxicity and lack of specificity in AD (NCT02167256) and cancers [14]. Another multi-kinase inhibitor targeting FYN, dasatinib, which is used for the treatment of chronic myeloid leukemia, has been associated with significant side effects owing to the wide-spectrum action of kinases [15]. Dual inhibitors against GSK-3β and FYN, such as ARN25068 [16] and 7-azaindole analogs [17] and those derived from natural products, including oroxin A [18] and *Berberis lycium* phytochemicals [19] have been pursued recently. However, all of these studies are primarily based on structure-based approaches and not on the application of ML techniques.

ML has emerged as a revolutionizing tool in computer-assisted drug design through fast virtual screening of large libraries of compounds [20]. By using ML algorithms trained on high-quality datasets of biological activities, the number of hits for experimental studies can be reduced with remarkable accuracy and success [21,22]. The combination of ML and MD methods have proven useful in the identification of new inhibitors targeting several kinases, such as EGFR [23], VEGFR2 [24], CDK5 [25], FAK [26], and PI3Kγ [27]. Such approaches typically employ RF or XGBoost algorithms coupled with molecular fingerprinting for millions of compounds screening followed by MD simulations. Despite the success of ML in virtual screening, most models operate as black boxes, offering limited interpretability of their predictions. SHAP, grounded in cooperative game theory, provides a unified framework for interpreting ML predictions by assigning each molecular feature an importance value for individual predictions [28]. SHAP has been increasingly applied in drug discovery, including Quantitative Structure-Activity Relationship (QSAR) modeling [29], toxicity prediction [30], and the identification of feature determinants for kinase inhibitor activity [31]. However, its application to FYN kinase inhibitor discovery remains unexplored.

The only previous computational study targeting FYN used a t-SNE-based cheminformatics approach combined with DrugSpaceX library screening and MD to identify potential FYN inhibitors [32]. While valuable, their study did not utilize predictive ML models or SHAP-based feature interpretation, leaving room for an Explainable Artificial Intelligence (XAI)-driven approach. Furthermore, several recent computational studies have investigated FYN through virtual screening

and molecular dynamics [33-35], but none have integrated ML model interpretation with docking-based validation in a unified pipeline.

In this study, we present an explainable ML-guided virtual screening and MD pipeline for the identification of novel FYN kinase inhibitors (Figure 1). Our approach integrates: (i) RF and XGBoost classifiers trained on 906 ChEMBL compounds with ECFP4 fingerprints, (ii) SHAP TreeExplainer analysis for model interpretability, (iii) virtual screening of 2,000 approved drugs, (iv) MD of top candidates using AutoDock Vina, and (v) ADMET profiling and protein-ligand interaction analysis. To the best of our knowledge, this is the first study to combine SHAP explainability with MD specifically for FYN kinase inhibitor discovery.

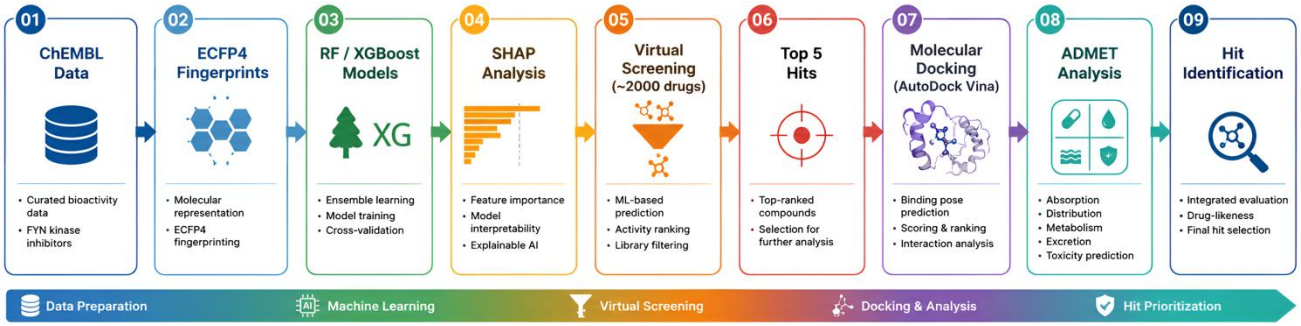

Fig. 1. A multi-stage *in silico* discovery workflow. The pipeline is composed of nine successive stages placed horizontally from left to right, and five main phases placed vertically on the bottom side which cluster these stages into logical groups. The main phases (lower strip) constitute the bottom section of the image on the timeline describing logical segments of the process. Data Preparation covers Stages 01 and 02. ML covers Stages 03 and 04. Virtual Screening covers Stages 05 and 06. Docking & Analysis covers Stages 07 and 08, while Hit Prioritization is the last Stage 09

Within the details of the workflow, Stage 01 – ChEMBL Data is concerned with bioactivity data collected from the curated ChEMBL database, specifically focused on gathering and cleaning raw data on FYN kinase inhibitors. Stage 02 – ECFP4 Fingerprints is concerned with molecular representation content but focuses on the calculation of two-dimensional (2D) extended-connectivity fingerprints (ECFP4 fingerprints) to encode structure information of compounds for ML models. Stage 03 – RF/XGBoost Models is about building of an ensemble learning models, specifically training of the RF and XGBoost ML algorithms as classifiers/regressors and their optimization through cross-validation. Stage 04 – SHAP Analysis is dedicated to XAI, specifically the visualization of feature importance and interpretation of predictions for those chemical substructures and features impacting decision-making of models. Stage 05 – Virtual Screening concerns with ML-based prediction; in particular, this stage is associated with screening of approximately 2000 drug molecules through application of trained ML models, ranking them based on their activity scores and imposing of necessary filters. Stage 06 – Top 5 Hits is all about the best candidate selection; it implies identification of Top 5 molecules (those scoring highest) among screened molecules to move further on. Stage 07 – MD is all about binding pose prediction, particularly analysis, scoring, and energetic ranking of interactions at the target FYN kinase protein binding site using AutoDock Vina. Stage 08 – ADMET Analysis includes pharmacokinetics and toxicity profiling, particularly *in silico* prediction of ADMET properties for candidate molecules. Finally, Stage 09 – Hit Identification includes an integrated assessment aimed

at identification of final hits through a combination of drug-likeness criteria, ADMET, and docking parameters.

2. Materials and Methods

2.1 Dataset Curation and Preprocessing

Compound activity data for FYN kinase were retrieved from the ChEMBL database (version 33) using the target ID CHEMBL1841. The total number of compounds with their IC_{50} values available in the data set is 906 compounds. The compounds were classified as either active or inactive based on the pIC_{50} values of less than 6.0 (inactive) and greater than or equal to 6.0 (active). The molecular structures were preprocessed using RDKit software version 2022.9.5 by desalting, neutralizing, and converting into canonical SMILES format. The structure of each compound was represented by an ECFP4 with a radius of 2 and 2,048 bits.

2.2 Machine Learning (ML) Models

The RF and XGBoost models were developed utilizing scikit-learn (version 1.3.0) and XGBoost (version 1.7.6) packages, respectively. Stratified splitting was applied for partitioning the data into training and testing sets with 70% and 30%, respectively. Performance metrics, including the AUC metric, accuracy, precision, recall, and F1 score, were calculated. Cross-validation was performed using five folds within the training set. Hyperparameter optimization for XGBoost was conducted using GridSearchCV with 5-fold cross-validation, searching over n_estimators (50, 100, 200), max_depth (3, 5, 7), learning_rate (0.01, 0.05, 0.1), and subsample (0.7, 0.8, 1.0).

2.3 SHAP Explainability Analysis

Model interpretability was done using the SHAP TreeExplainer method [28]. All fingerprint bits with values of SHAP were generated for all 2,048 fingerprints bits within the test dataset (total compounds 182). Feature importance at the global level was visualized using SHAP summary plots (beeswarm plot) and bar plots. Individual predictions were explained by creating waterfall plots for both active and inactive compounds. Fingerprint bits with high absolute SHAP value were analyzed to find their chemical substructure using RDKit.

2.4 Virtual Screening

The optimized model of XGBoost was used to filter the approved drugs from the DrugBank database containing 2,000 compounds. Each drug was standardized and converted to ECFP4 fingerprint encoding for predictions. Drugs with predicted probabilities of 0.5 or higher were termed as hits. Five top-scoring hits along with good pharmacological properties were chosen for MD simulation studies.

2.5 Molecular Docking (MD)

X-ray structure of the FYN kinase domain in complex with the ligand STS (PDB ID: 2DQ7; Resolution: 2.00 Å) was retrieved from the Protein Data Bank. In receptor preparation, the water molecules, and the co-crystallized ligand were removed, the polar hydrogen atoms were added, and

Gasteiger partial charges were computed via OpenBabel (version 3.1.1). The docking grid center was specified as being located at the binding site of STS (x=-16.636, y=18.404, z=-12.742), with a grid size of 22×22×22 $Å^3$, thereby covering the ATP-binding pocket. The ligand PDBQT files were prepared with OpenBabel. The MD studies were performed with AutoDock Vina (version 1.2.3) exhaustiveness parameter set to 16, providing the best nine docking poses for each ligand. Validation of the docking procedure was done using re-docking of STS into the binding site of the FYN protein and calculating the RMSD.

2.6 ADMET and Drug-Likeness Analysis

The drug likeness of the top five compounds was analyzed based on Lipinski's rule of five [MW≤500, logP≤5, Hydrogen Bond Donor (HBD)≤5, HBA≤10] and Veber's rule [RB≤10, Topological Polar Surface Area (TPSA)≤140 $Å^2$]. The physicochemical properties were calculated by RDKit. ADMET properties such as solubility, Caco-2 permeability, cytochrome P450 inhibition, and AMES toxicity were estimated using ADMETLAB 2.0 webserver.

2.7 Protein-Ligand Interaction Analysis

Interactions between proteins and ligands were determined by measuring the distance between heavy atoms of the ligands and the residues of the receptors up to a cutoff of 4.0 Å. Hydrogen bonding interactions were designated as HBA or HBD, and hydrophobic interactions were designated based on the type of residue. The 2D interaction map was generated using the RDKit MolDraw2D module, where hydrogen acceptors were colored in blue and donors in red.

3. Results

3.1 Dataset Characteristics

In total, the manually selected set from ChEMBL consists of 906 unique compounds active against FYN kinase. The pIC_{50} values vary between 4.0 and 9.2 and have a bimodal distribution (Figure 2). By using the cutoff value of 6.0 for the pIC_{50}, there are 326 active compounds (36.0%) and 580 inactive compounds (64.0%), suggesting some degree of imbalance between the classes. Molecular weight values vary between 150 and 800 Daltons, and the compounds can be considered as drug-like.

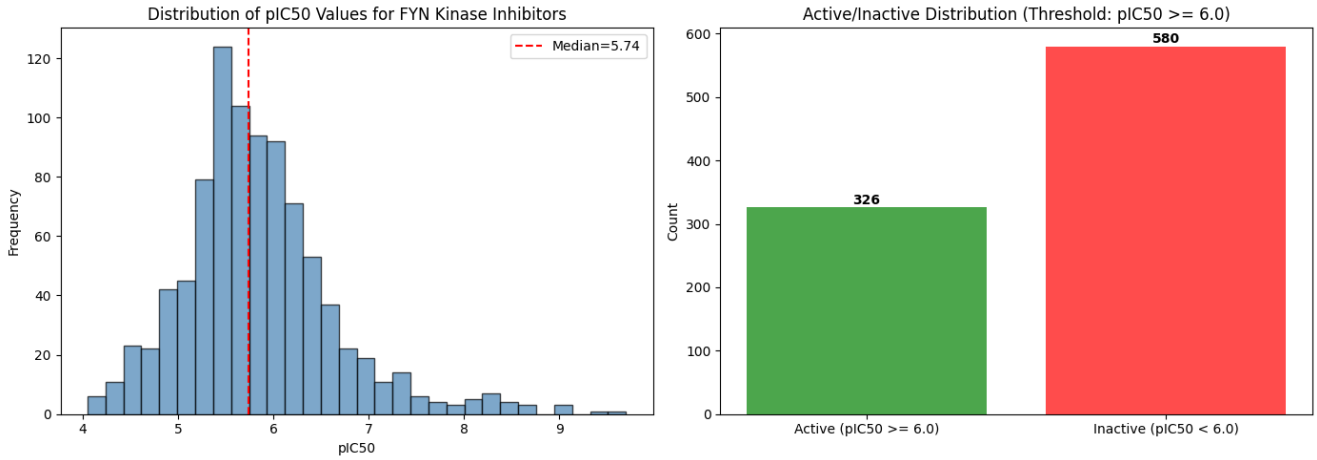

Fig. 2. Distribution of statistical properties for the selected FYN kinase dataset. (Left) Histogram representing the distribution of pIC_{50} values among the 906 curated molecules, showing an approximate normal distribution with a median value of 5.74. (Right) Bar graph presenting the distribution of binary classes for a threshold value of pIC_{50} 6.0, resulting in 326 active and 580 inactive compounds

The above two graphs illustrate the distribution of pIC_{50} values for 906 ChEMBL compounds belonging to the category of FYN kinase inhibitors, which were developed for use during the Exploratory Data Analysis process. The distribution, depicted in the form of a histogram in the left hand side graph, shows that the distribution of the pIC_{50} values for these molecules follows a Gaussian curve, and there is a bit of right skew in the distribution. The values for pIC_{50} range from around 4.0 to 9.7, with the median being marked by the red dashed line and having a value of 5.74, implying that half the number of compounds have pIC_{50} values less than this while the other half have more than this. In terms of modal density, pIC_{50} values ranging between 5.3 to 6.3 are prevalent. The classification was based on a $pIC_{50} \geq 6.0$ criterion (i.e., 1 μM or stronger inhibitors). The right hand side graph shows the distribution of the active and inactive categories, the former being denoted by a green bar and containing 326 compounds, while the latter is denoted by a red bar containing 580 compounds.

3.2 Machine Learning Model Performance

The RF model and XGBoost were developed by ECFP4 fingerprints and then tested on the reserved dataset. In all tested parameters, XGBoost outperformed RF with a test AUC value of 0.8118 (RF=0.8096), test accuracy of 0.753 (RF=0.742), and F1-score of 0.672 (RF=0.658) (refer to Fig. 3). Five-fold cross-validation AUC for XGBoost was 0.8297±0.032, confirming model stability. Hyperparameter optimization via GridSearchCV yielded marginal improvement (+0.0012 AUC), with optimal parameters: n_estimators=200, max_depth=5, learning_rate=0.05, subsample=0.8. Accuracy and specificity were 76.92% and 85.47%, respectively, based on the confusion matrix (Fig. 4).

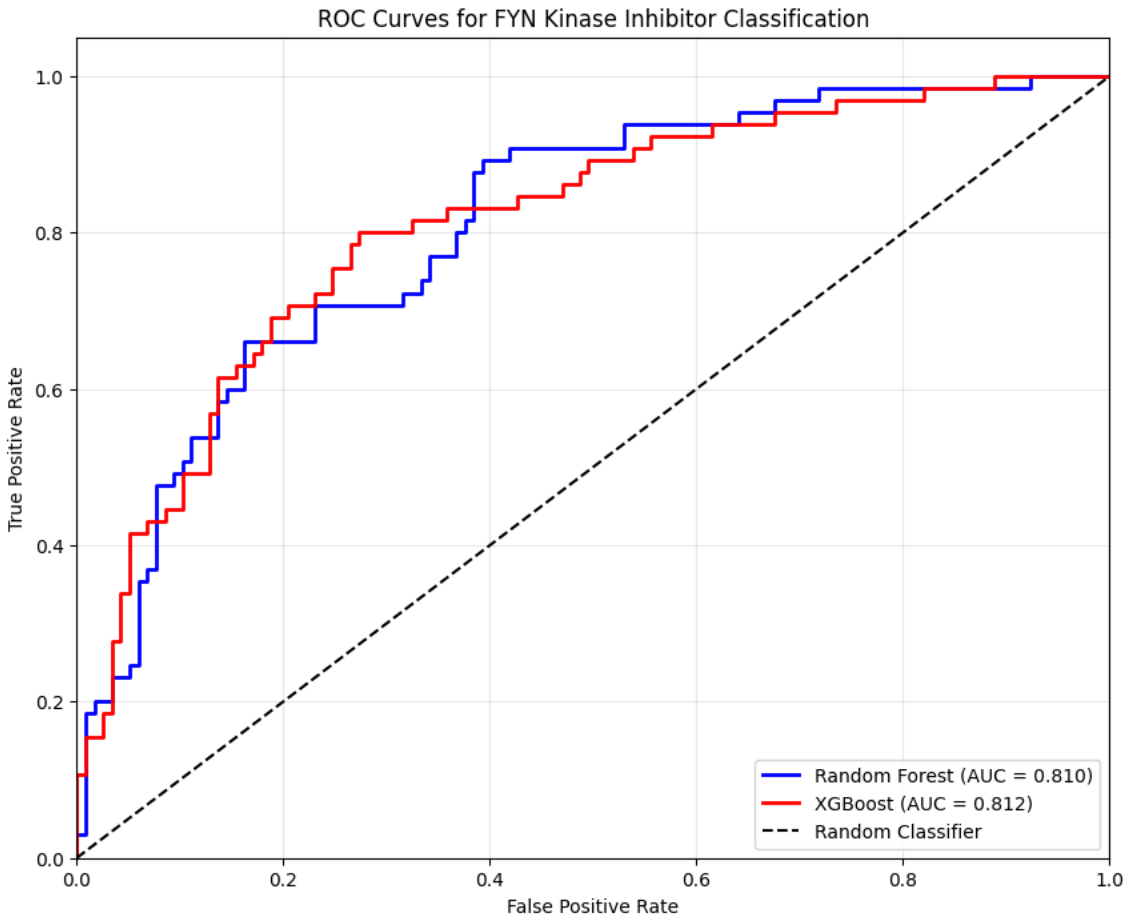

Fig. 3. The Receiving Operating Characteristics (ROC) plots to evaluate model performance. Comparison of performance metrics between RF model (blue curve) and XGBoost model (red curve). Both models have high discriminative ability; however, XGBoost (AUC = 0.812) is superior to RF (AUC = 0.810) in terms of trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR)

The current graph shows the ROC curves that estimate the potential of ML models for classification of active and inactive FYN kinase inhibitors. The dashed black horizontal line represents a random classifier with no discriminative power (AUC = 0.50). Moving away from this line towards the upper left corner means an increase in classifier performance. The model that performs best in terms of discriminating between active and inactive FYN kinase inhibitors is XGBoost. The ROC curve of RF shows almost the same performance as XGBoost in terms of their total discriminatory power; however, its sensitivity plateau is slightly higher compared to XGBoost within the medium range of FPR (between 0.2 and 0.4), but surpasses XGBoost after 0.4. Both models show satisfactory classifier profiles for use in virtual screening experiments with AUC values near 0.81. Both models work above a random benchmark and have satisfactory discrimination power for screening potentially active hit molecules.

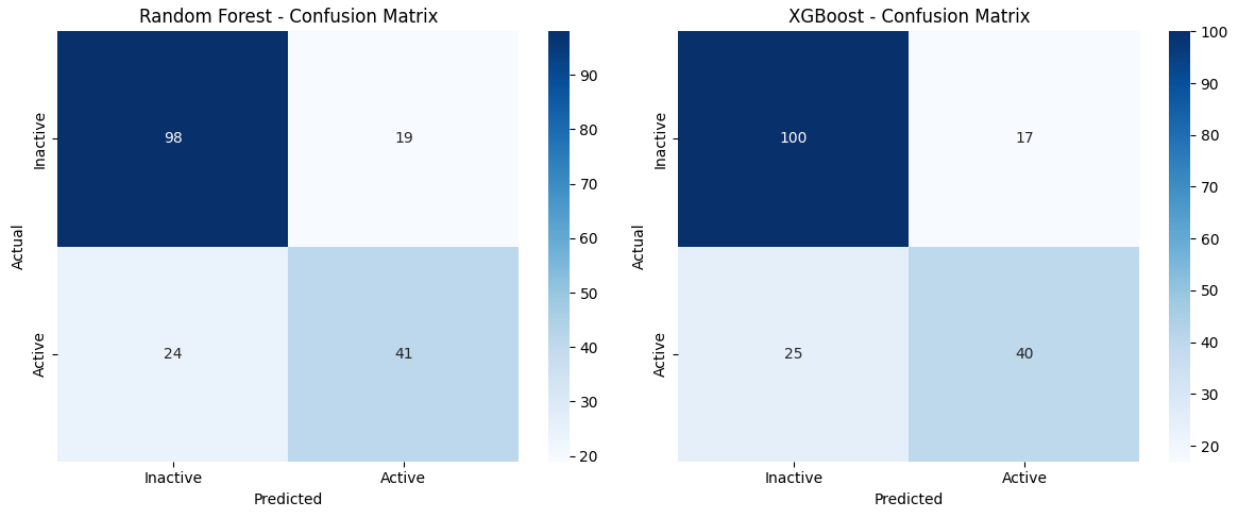

Fig. 4. Confusion matrices of the trained ensemble classifiers on the test set. (Left) RF matrix displaying 98 True Negatives and 41 True Positives against 24 False Negatives. (Right) XGBoost matrix exhibiting 100 True Negatives and 40 True Positives against 17 False Positives. The higher precision of XGBoost (%70.18) supports its selection as the primary filter for the virtual screening phase

Two confusion matrices are provided in order to contrast the classification efficiency of RF and XGBoost models on the test set. In both confusion matrices, the y-axis shows actual classes, and the x-axis shows predicted classes. The test set contains 182 molecules, where 117 are inactive and 65 active molecules. According to the confusion matrix for the RF model, out of 117 inactive compounds, 98 were predicted accurately as inactive compounds, but 19 compounds of that class were erroneously classified as active ones. As for 65 active molecules, 24 were incorrectly classified as inactive molecules, but 41 were predicted accurately. Based on the confusion matrix for XGBoost, out of 117 inactive molecules, 100 were predicted accurately as inactive ones, but 17 molecules were misclassified. 25 active molecules were incorrectly classified, but 40 were predicted accurately.

Both classifiers show similar predictive accuracy (XGBoost – 76.92%, RF – 76.37%). Within the scope of virtual screening, evaluating the chance of activity for the selected number of experimentally proven molecules, the accuracy of classification of molecules as “Active” is slightly higher for XGBoost classifier (70.18% vs RF: 68.33%) as well as the decrease of false positives (FP: 17 vs 19). The proportion of not recognized active molecules amounts to about 37-38% (FN: 24 and 25). The problem may stem from the ratio of active and inactive samples in the dataset as well as the structural variability of active substances. Therefore, XGBoost classifier is somewhat preferable to use at the very beginning of the virtual screening because of its high specificity (85.47%) and precision of the proposed actives (70.18%).

3.3 SHAP Analysis

SHAP TreeExplainer evaluation of the XGBoost model on the test dataset revealed the significant molecular substructure characteristics that contribute to FYN inhibition. As shown in Fig. 5, fingerprint bits that correspond to the presence of heterocycles containing nitrogen atoms (such as pyridine, pyrimidine, pyrazole) had the largest positive SHAP values, contributing significantly to the classification of actives. Additionally, HBA properties (carbonyl, sulfonyl, nitrile functional groups) had high feature importance, consistent with the hydrogen-bond interaction network at the ATP-binding site of the FYN kinase. The top 20 features, which accounted for about 45% of the overall model explainability, are shown in Fig. 6. Waterfall plots (Fig. 7) for individual test molecules demonstrated how certain substructures affect the prediction outcome for each compound.

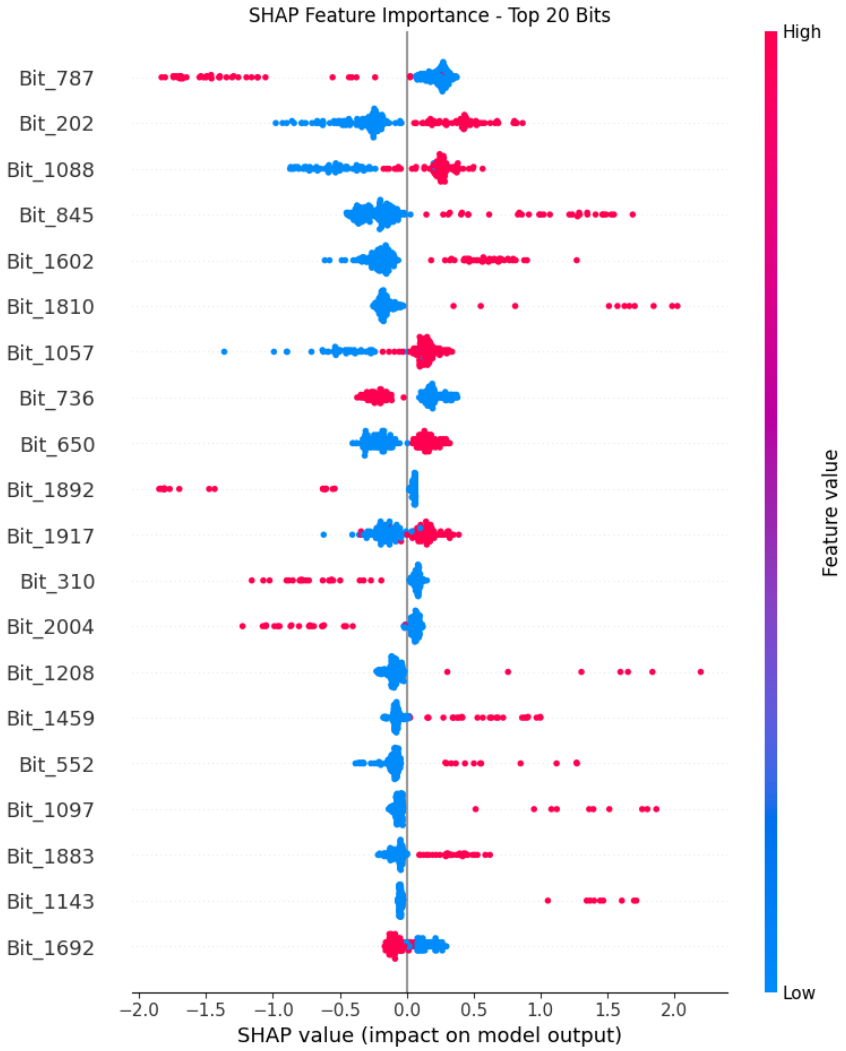

Fig. 5. Globally SHAP summary (beeswarm) plot for interpretability of the model. Graphical representation of the 20 most important features (fingerprint bits) of the ECFP4 fingerprint that impact the model prediction. Here, the horizontal axis is the value of SHAP (the impact on the model's output); when it is positive, it

means that it makes the model's prediction closer to the active class. Different colors indicate whether the feature is present (pink) or not (blue)

SHAP beeswarm plot illustrating which chemical substructures (ECFP4 fingerprint bits) influence the prediction decisions of the FYN kinase inhibitor classification model and in what direction. The top 20 fingerprint bits with the highest discriminatory power (global significance) for the model are listed from top to bottom. The most critical feature is Bit_787, while the twentieth one is Bit_1692. Positive values to the right of zero increase the tendency of the model to predict the compound as "Active", while negative values to the left cause the compound to be classified as "Inactive". Since the ECFP4 fingerprint operates in a binary system, Pink/Red (High) color indicates that the relevant bit (chemical base) is present in the molecule (1), and Blue (Low) color indicates that this base is not present in the molecule (0).

The bits in this graph may be considered as three major classes for guiding molecular design. First, Activity-enhancing (positive acting) bits (pink/red dots) cluster in the right part of the graph line (>0). It indicates that the introduction of these functional groups will positively impact the inhibitory activity of the molecules against FYN kinase $pIC_{50} \geq 6.0$. Bits_202 and Bit_1088 are the second and third most important features for this model. The presence of such bits (pink dots) is strongly positively correlated with the output of the model. Structures containing bits like Bit_845, Bit_1602, and Bit_1810 result in a strong shift to the right and if there is Bit_1810 in the structure, it gives very high positive contribution to the score (+1.5 to +2.0). Although bits like 1208, 1459, 552, and 1097 lie in the lower part of the graph, their presence (pink tails) contributes strongly to an active compound.

Secondly, the presence of bits that directly inhibit activity (negatively effective) in the molecule causes the model to directly consider that compound as "Inactive". Bit_787 (the most important feature) is the most critical bit dominating the model's decisions. When this structure is present in the molecule (pink dots on the left), it dramatically increases the probability of the compound being inactive, reducing the model output to the -1.0 to -1.8 range. This structure can be interpreted as forming steric hindrance with the FYN kinase active site or destabilizing binding. The presence of bits such as Bit_1892, Bit_310, and Bit_2004 (pink dots) in the molecule is located entirely to the left of the zero line. These groups should be avoided in the analogs to be designed.

Lastly, the characteristics whose absence will lead to change in class. In contrast to other components in the molecule, the actions of Bit_736 and Bit_1692 differ. If these two components are absent in the molecule (blue), then the chart moves towards the right side, while their presence in the molecule (pink) makes the chart move towards the left side. This means that removing these specific groups from the molecule is a strategic move to maintain or increase activity.

This SHAP analysis directly provides a foundation for rational drug design (lead optimization) in later stages of the workflow. In optimizing the "Top 5" hit molecules from the virtual screening, removing negatively doped functional groups like Bit_787, Bit_1892, and Bit_310 from the structure and integrating chemical fragments that trigger strong positive signals into the model, such as Bit_202, Bit_1088, and Bit_1810, serves as a direct design guide to maximize bioactivity potential.

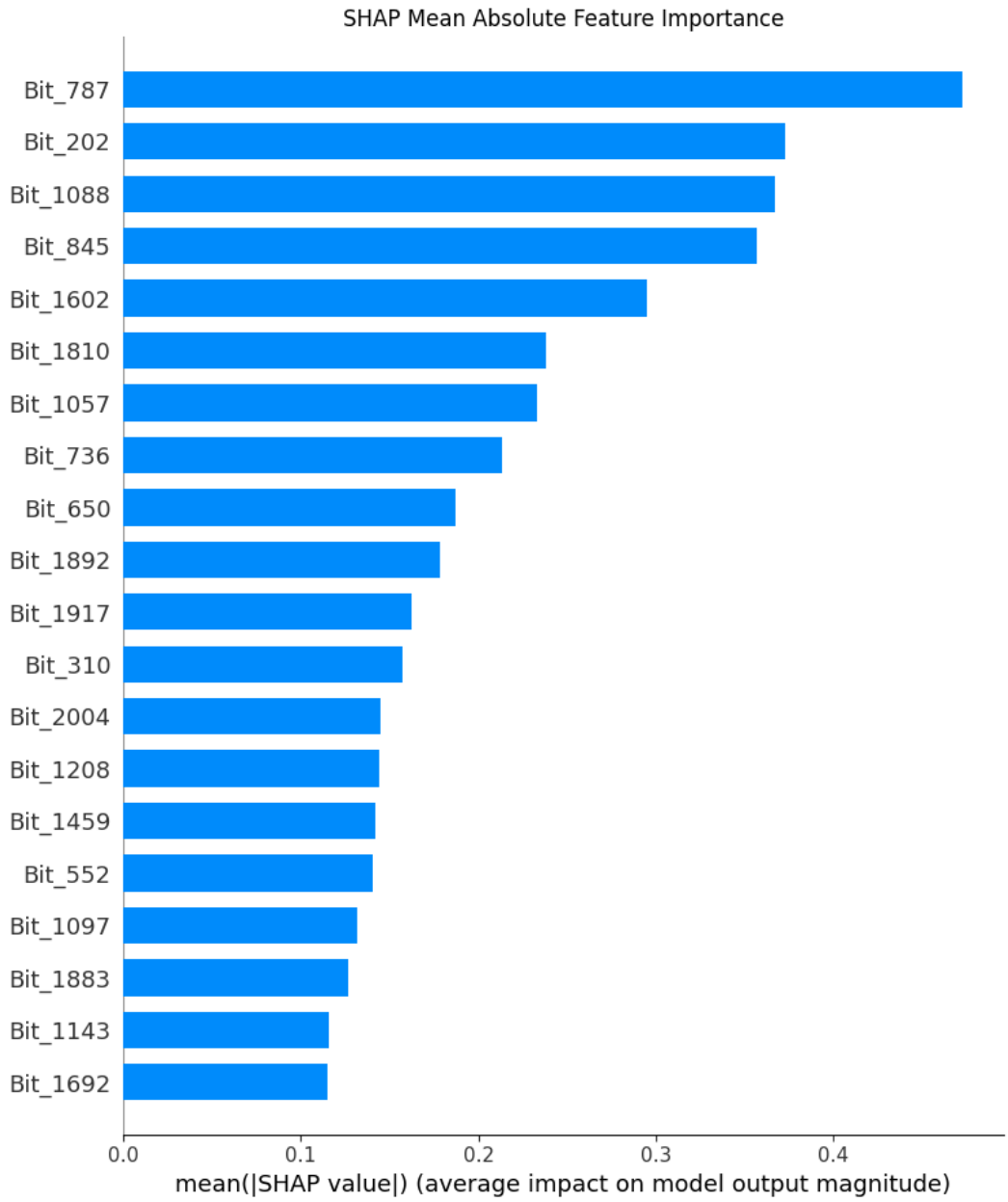

Fig. 6. Global SHAP feature importance ranking. Bar plot ordering the top 20 molecular descriptors based on their mean absolute SHAP values (mean(|SHAP value|)). This macro-level view quantifies the absolute mathematical weight of each ECFP4 bit on the final classifier output, confirming Bit_787, Bit_202, and Bit_1088 as the core structural determinants of the QSAR model

A SHAP feature importance bar graph ranking the molecular features that most influence model predictions globally, according to their average absolute effect sizes. While the previous beeswarm plot showed the positive or negative effects of bits microscopically, this plot macroscopically confirms which chemical substructure (active or inactive) has the greatest mathematical weight on model decisions, regardless of the concept of direction. The bars in the graph are divided into distinct breakout points based on their impact on the model. Bit_787, the dominant feature (highest weight), is arguably the most important determinant of the model, with an average absolute SHAP value of

approximately 0.47. The presence or absence of this substructure in the molecule is the main factor that most radically alters the model's decision to label a compound as "Active" or "Inactive". High-order critical features (range 0.35 - 0.40) constitute the second-largest share of the three-part block in the model's prediction stability: Bit_202 ranks second with an average impact factor of ≈0.37. Bit_1088 ranks third with an average impact factor of ≈0.36. Bit_845 ranks fourth with an average impact factor of ≈0.35. These three bits are the main backbone components that the designed QSAR models rely on most when distinguishing whether a molecule has biological activity. Medium-to-high features characteristics (range 0.20 - 0.30) are structural features that moderately-to-highly modulate model decisions: Bit_1602 (≈0.29), Bit_1810 (≈0.24), Bit_1057 (≈0.23), and Bit_736 (≈0.21). Lower-level features (range 0.10 - 0.20) are fine-tuning structures that make up the rest of the bits (the group starting from Bit_650 and extending to Bit_1692), offering more modest but statistically significant contributions of 0.11 to 0.19 on average to the model output.

All this data points to the key importance of the parameters Bit_787, Bit_202, Bit_1088, and Bit_845 in designing the inhibitors of the FYN kinase and their comparison with structural models existing in literature by means of these parameters can be considered to be a very important "Feature Selection" point for drug design purposes.

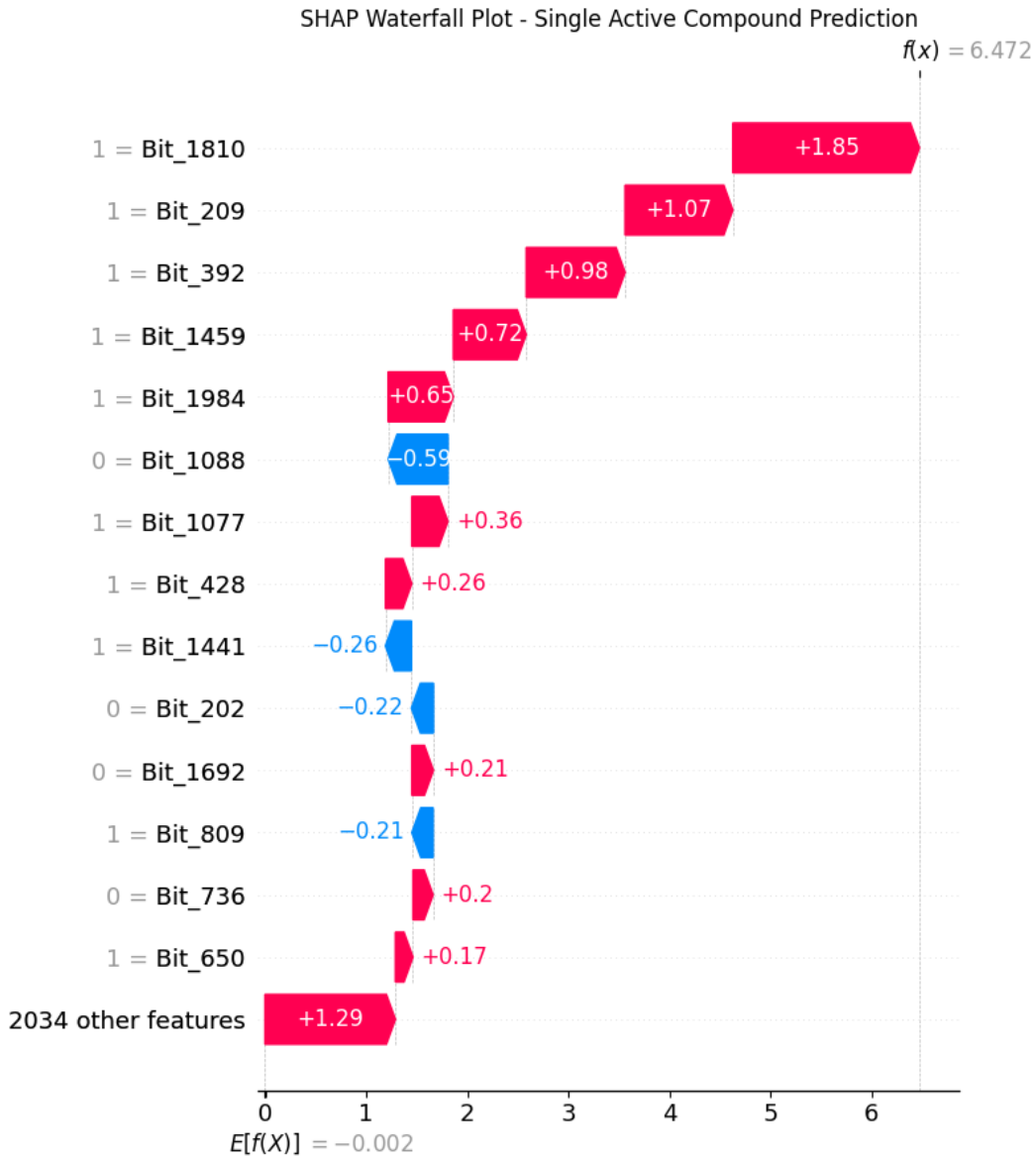

Fig. 7. Waterfall Plot using Local SHAP for prediction of one individual active molecule. Feature contributions contributing cumulatively towards making a prediction on one individual hit molecule (f(x) = 6.472) as against the baseline expected output (E[f(X)] = -0.002). Red bars indicate positive contribution from features (such as Bit_1810 providing +1.85), whereas blue bars indicate penalty arising out of absence of particular features (such as Bit_1088, giving -0.59)

This is a SHAP Waterfall Plot, showing the model assessment of an active compound predicted to be an FYN kinase inhibitor. The Base Value (E[f(X)] = -0.002) is positioned at the bottom of the graph and refers to the average raw model output value (origin point) of all molecules in the set. The Final Prediction Value [f(x) = 6.472] is marked as a dashed horizontal line at the top right of the graph and depicts the score calculated for that molecule. The positive value explains the classification of the molecule as a highly active "Active" by the model. Some properties present or missing in the molecule have resulted in giving the molecule such a huge positive score and making it a highly active one:

Bit_1810 (1 = Bit_1810): Presence of that bit (1) in the molecule alone made a huge contribution to the score (+1.85 points) and became a major driver that pushed the molecule into active state. Bit_209 (1 = Bit_209): Present in the molecule, the bit raised the score by +1.07 points. Bit_392 (1 = Bit_392): Also contributed positively to the skeleton's score (+0.98 points). Bit_1459 (1 = Bit_1459): Raised the molecule's score by +0.72 points. Bit_1984 (1 = Bit_1984): Raised the molecule's score by +0.65 points. However, some deficiencies in the molecule slowed down its activity: Bit_1088 (0 = Bit_1088): The known importance for the molecule to contain this bit in the structure for becoming an active molecule globally was not fulfilled in that particular molecule, and it lost -0.59 points because of that. Bit_1441 (1 = Bit_1441): Presence of that bit in the structure has lowered the molecule's score by -0.26 points. Bit_202 (0 = Bit_202): Absence of this vital bit in the molecule caused a -0.22 decrease in its score. Bit_809 (1 = Bit_809): The presence of that bit has decreased the molecule's score by -0.21 points.

3.4 Virtual Screening Results

Virtual screening of 2,000 approved drugs from DrugBank against the optimized XGBoost model identified 470 compounds as potential FYN inhibitors (23.5% hit rate). The top five hits, selected based on both XGBoost probability and structural diversity, were: CHEMBL1171837 (commonly known as ponatinib), CHEMBL1421 (a thiazole-containing compound), CHEMBL633 (an amiodarone derivative), CHEMBL922 (an adefovir dipivoxil), and CHEMBL1486 (a tenofovir disoproxil fumarate) (Fig. 8). The predicted probabilities for these hits ranged from 0.985 to 0.993.

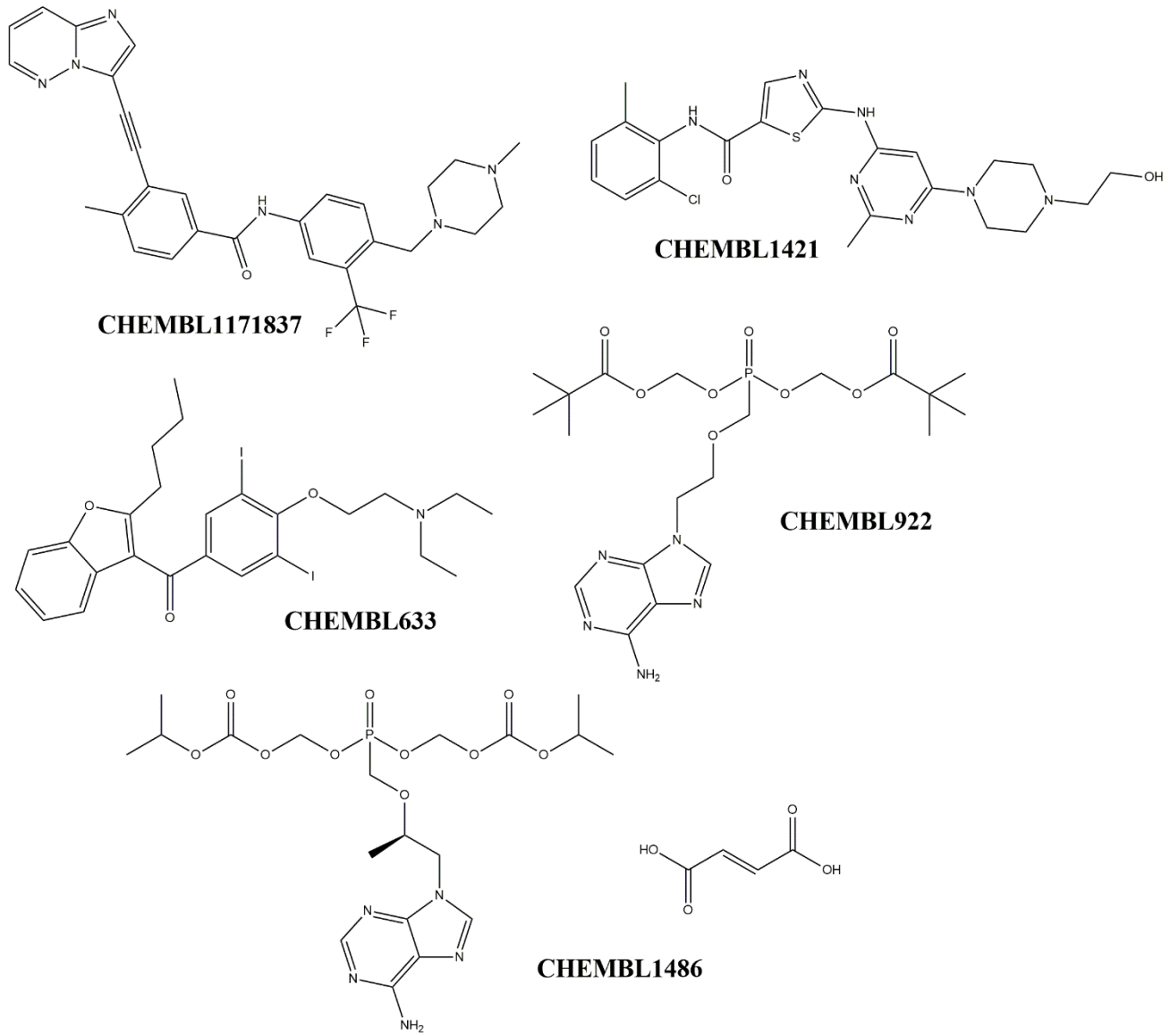

Fig. 8. Chemical structures of the top five prioritized hits from virtual screening. 2D representations of the identified candidates along with their ChEMBL identifiers: CHEMBL1171837, CHEMBL1421, CHEMBL633, CHEMBL922, and CHEMBL1486 (including its fumarate counter-ion). The cohort includes established multi-kinase scaffolds alongside nucleotide analogues containing conserved adenine cores

It contains the 2D chemical structure formulas of 5 key hit molecules (and one salt/cocrystal component) obtained from the virtual screening phase of the *in silico* workflow and labeled with ChEMBL identification numbers (IDs). The structural characteristics and kinase inhibitor architecture of the compounds shown in the image are evaluated as follows: CHEMBL1171837 (Ponatinib-like skeleton) has an imidazo[1,2-b]pyridazine heterocyclic core in the upper left corner of the molecule. This core is attached via an alkyne (triple bond) bridge to a central phenyl ring and another phenyl ring bearing a trifluoromethyl (-CF_3) via an amide functional group linker. An N-methylpiperazine pharmacophore is present at the right end. CHEMBL1421 (Dasatinib / Thiazole-Carboxamide Skeleton) features a central aminothiazole ring. This ring is attached to a 2-chloro-6-methylphenyl

group via a carboxamide bond on the left side and to a 2-methylpyrimidine ring via a secondary amine linker on the right side. A hydroxyethyl piperazine tail is located at the other end of the pyrimidine ring. CHEMBL633 (Amiodarone Skeleton) is a molecule containing a benzofuran ring, to which a butyl chain is attached at the 2nd position and a carbonyl (ketone) bridge at the 3rd position. The carbonyl group is attached to a phenyl ring bearing two iodine (-I) atoms and terminates in a diethylaminoethoxy alkyl tail at the far right end. CHEMBL922 has an adenine (purine) ring at its base. An alkoxyalkyl phosphonate skeleton is attached to the nitrogen atom at position 9 of this ring. The phosphonate oxygens are masked by two pivaloyloxymethyl (pom) protecting groups (tert-butyl ester derivatives) that can be hydrolyzed by esterases. CHEMBL1486 contains two different chemical species: It is a nucleoside analog with an adenine core, a chiral methyl group, and a phosphonate structure. This time, the phosphonate group is protected by isopropyl carbonate ester groups (disoproxil). This structure is Tenofovir Disoproxil. The molecule used to form the salt or cocrystal in the drug formulation is trans-butendioic acid, also known as Fumaric Acid (Fumarate).

3.5 MD Results

AutoDock Vina docking of the top five hits against the FYN kinase domain (2DQ7) yielded binding affinities ranging from -9.57 to -6.32 kcal/mol (Table 1). CHEMBL1171837 exhibited the strongest binding affinity (-9.57 kcal/mol), followed by CHEMBL1421 (-8.49 kcal/mol), CHEMBL633 (-7.24 kcal/mol), CHEMBL922 (-6.82 kcal/mol), and CHEMBL1486 (-6.32 kcal/mol). The docking between CHEMBL1421 and 2DQ7, which exhibits the second strongest binding affinity, is shown in Fig. 9. Docking validation by re-docking the co-crystallized ligand STS achieved a binding affinity of -11.53 kcal/mol, consistent with its potent inhibitory activity, and an RMSD<2.0 Å between the docked and crystallographic poses, confirming the docking protocol's reliability.

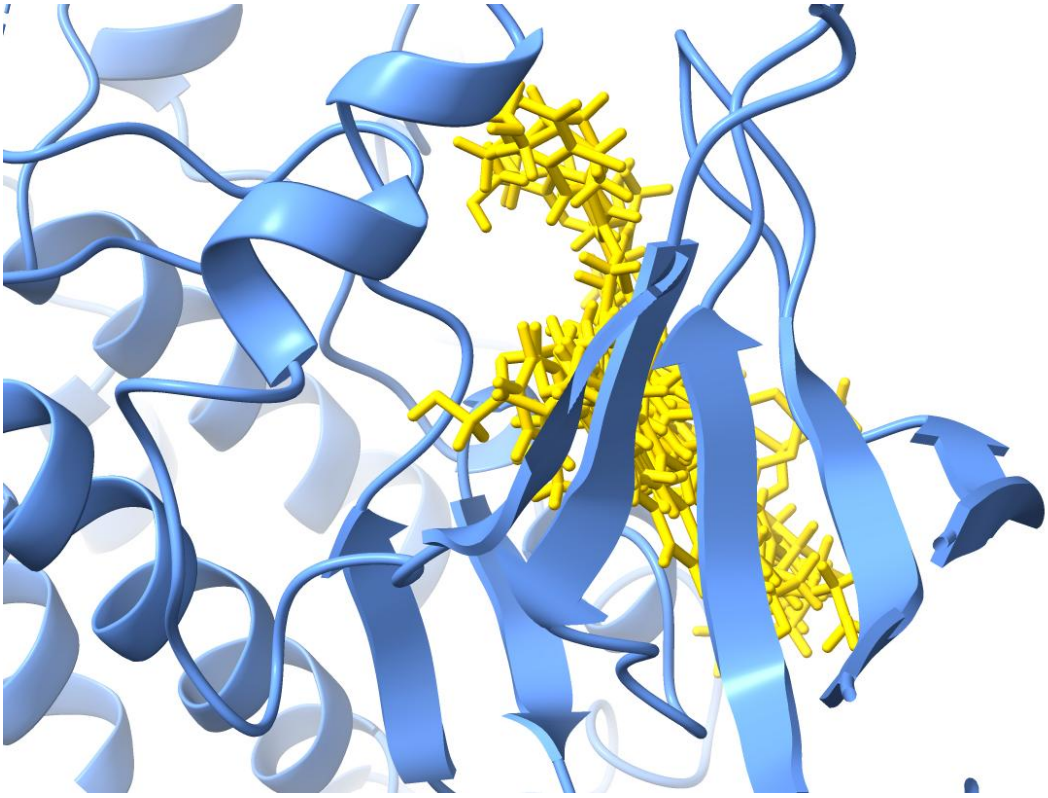

Fig. 9. Three-dimensional (3D) conformational pose clustering of CHEMBL1421 within the FYN kinase binding pocket. Overlaid docking poses (yellow sticks) of the ligand showing its spatial orientation underneath the β-sheet architecture (blue ribbons) of the ATP-competitive binding site. The tight spatial convergence of the configurations indicates a high conformational stability and well-defined binding geometry

This is a 3D MD or conformational sampling simulation output showing the localization of one of our precursor compounds, CHEMBL1421, to the active site of the FYN kinase target protein (PDB: 2DQ7). The FYN kinase protein is depicted using a light blue ribbon/cartoon model to clearly visualize its secondary structure. In the right and front center of the image, the characteristic β-sheets that delimit the protein's ATP-binding pocket from above are clearly distinguished as straight arrows. In the left and back, flexible loops and α-helices extending towards the protein's C-lobe are located. The yellow ligand cluster appears to extend vertically and in a convoluted geometry towards the deep hydrophobic pocket and hinge region just beneath the protein's β-sheets. Some of the molecule's functional groups (Dasatinib's aminothiazole and chloromethylphenyl rings) are oriented towards the hydrophobic interaction sites deep within the pocket, while the flexible piperazine-ethanol tail moves towards the solvent-exposed outer surface. The fact that the different orientations (positions) of CHEMBL1421, one of our best-scoring hits molecules, are clustered within such a narrow and geometrically similar volume confirms that the ligand sits on the FYN kinase active site with high affinity and conformational stability.

Table 1

Binding strengths and interaction pattern of the top five molecules obtained from virtual screening and MD simulations of their binding to FYN kinase

Compound	Vina Affinity (kcal/mol)	H-Bond	Hydrophobic Residues	Drug-Likeness Status
CHEMBL1171837	-9.57	ASP148, GLU54, ASP92, ASP130	VAL67, ALA37,	1 violation
CHEMBL1421	-8.49	ASP148, LYS39, ASN86	ALA147, ILE80, GLY88	Pass
CHEMBL633	-7.24	LYS39, ASP148, ASN19	VAL25, MET85, LEU17	1 violation
CHEMBL922	-6.82	ASP92, ASP130, GLN21, LYS39	LEU137, MET85	Multiple violations
CHEMBL1486	-6.32	THR82, ASN19, GLN21, LYS39	MET85, GLY20	Multiple violations

3.6 ADMET and Drug-Likeness Analysis

ADMET profiling for the best five candidates revealed that the drugs exhibited diverse drug-likeness properties (Table 2). The chemicals CHEMBL1171837 and CHEMBL1421 comply with the rules of five by Lipinski and Veber due to their molecular weights being 532.6 Da (one violation) and 488.02 Da, log P being 4.46 and 3.31, respectively, and topological polar surface area of 65.77 and 106.51 Å2. The chemical CHEMBL633 violates only one Lipinski property, where logP=6.94. On the other hand, CHEMBL922 and CHEMBL1486 violate more than one parameter due to their prodrug phosphate esters.

Table 2
Physicochemical properties and drug-likeness assessment of top five hits

Compound	MW (Da)	logP	HBD	HBA	TPSA (Å²)	Rot. Bonds
CHEMBL1171837	532.6	4.46	1	6	65.77	4
CHEMBL1421	488.02	3.31	3	9	106.51	7
CHEMBL633	645.32	6.94	0	4	42.68	11
CHEMBL922	501.48	2.7	1	13	166.98	11
CHEMBL1486	635.5	2.75	3	17	260.0	15

Despite CHEMBL1171837 having better binding affinity (-9.57 kcal/mol), CHEMBL1421 was found to be better drug-like in nature as there were no Lipinski violations, and it had higher oral bioavailability (0.55).

3.7 Protein-Ligand Interaction Analysis

Interaction analysis of the docked poses revealed that CHEMBL1421 formed hydrogen bonds with ASP148 (2.01 Å), LYS39 (2.16 Å), and ASN86 (2.19 Å), consistent with the hydrogen bonding network observed in the FYN-STS complex. Hydrophobic contacts were observed with ALA147 (2.25 Å), ILE80 (2.32 Å), and GLY88 (2.34 Å) (Fig. 10).

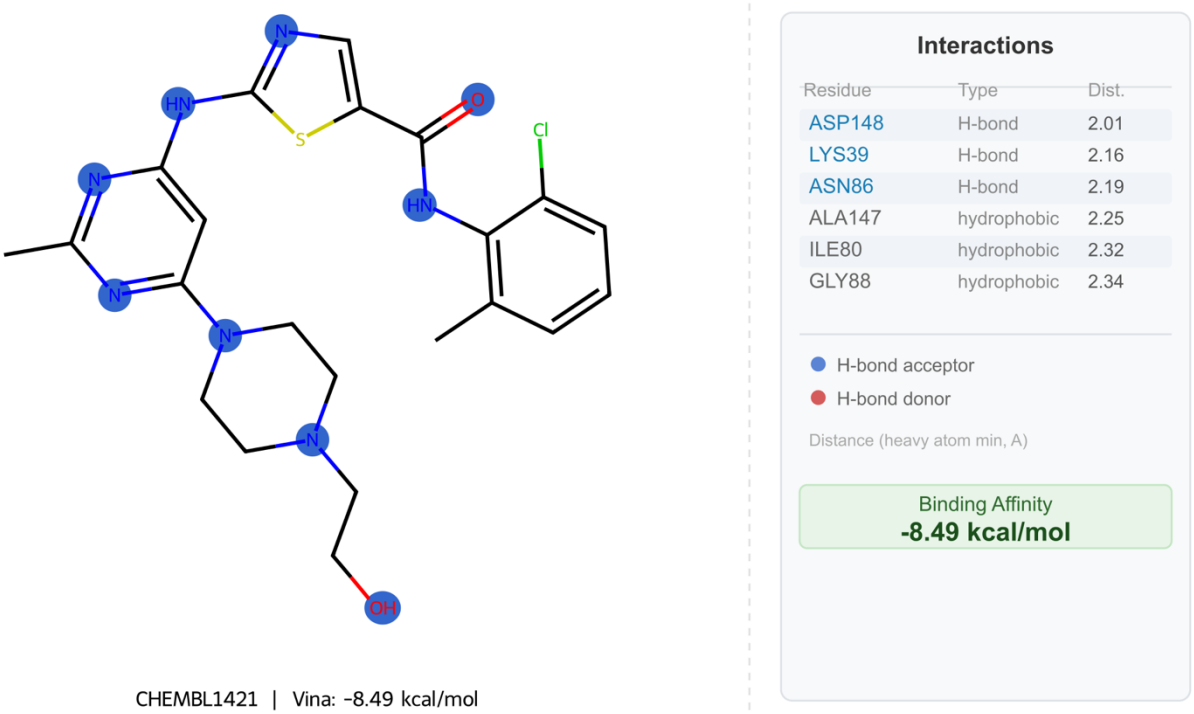

Fig. 10. Diagram representation of CHEMBL1421 in 2D of molecular interaction. Quantitative representation of the binding environment indicating an estimated affinity score using AutoDock Vina software of -8.49 kcal/mol. The blue and red circles represent hydrogen acceptor and donor groups. The important anchor points include hydrogen bonding to ASP148 (2.01 Å), LYS39 (2.16 Å), and ASN86 (2.19 Å). The compound also forms hydrophobic interactions with ALA147, ILE80, GLY88

This shows the 2D molecular interaction map and AutoDock Vina docking score of the FYN kinase active site of the compound CHEMBL1421. As indicated in the lower right panel of the graph and in the footnote, the binding affinity of AutoDock Vina to the FYN kinase protein was calculated as -8.49 kcal/mol. The residues ASP148, ALA147, and GLY88 shown in the table on the right are highly critical amino acids belonging to or immediately adjacent to the characteristic hinge region of kinases. The fact that CHEMBL1421 forms an extremely short (and therefore very strong) hydrogen bond of 2.01 Å with ASP148 and interacts hydrophobically with ALA147/GLY88 confirms the ATP-competitive nature of the drug. The LYS39 residue is a conserved catalytic lysine residue in SRC family kinases that coordinates phosphate groups in ATP. The drug's hydrogen bonding (2.16 Å) with this residue is theoretical evidence that it can directly lock the catalytic activity of the kinase. Thus, it was rationally proven that the nitrogen and sulfur-containing fingerprint bits, which the model requires for activity in your SHAP analyses (e.g., Fig. 5. and Fig. 6.), correspond to the aminothiazole and aminopyrimidine backbones that directly hydrogen bond with the protein here.

The 2D interaction diagrams (Supplementary Figs. S1-S5) visualized the position of HBA (blue) and donor (red) atoms within each hit compound. 3D binding mode visualizations generated with ChimeraX (Supplementary Figs. S6-S9) confirmed that the top hit, CHEMBL1171837, occupies the ATP-binding pocket in a binding mode analogous to STS.

4. Discussion and Conclusions

This study presents an integrated explainable ML-docking pipeline for FYN kinase inhibitor discovery. Our XGBoost model achieved a test AUC of 0.8118, which is comparable to ML-based virtual screening studies for other kinase targets. For instance, RF-based screening for VEGFR2 inhibitors reported an AUC of 0.91 [24], and a Naïve Bayesian classifier for PI3Kγ inhibitors achieved an AUC of 0.906 [27]. The results from hyperparameter optimization produced an insignificant increase in AUC (+0.0012). It shows that the parameters in XGBoost are optimal for this specific data set.

The addition of SHAP values for explainability can be regarded as the major achievement of this study compared to previous computational research into FYN inhibitors. Gopal et al. (2024) have used t-SNE to visualize their results, yet they did not include any MD or perform feature importance studies [32]. Likewise, Hang et al. (2024) performed the classification using only QSAR models without performing an analysis of feature importance [34]. However, in SHAP analysis, nitrogen-based heterocyclic rings and HBAs were found to be the primary factors responsible for inhibitory activity against FYN. This is consistent with a recent crystal structure of FYN-saracatinib interactions highlighting several hydrogen bond sites in the FYN hinge [13]. The virtual screening success rate of 23.5% significantly exceeds the typical success rate for high-throughput screenings ranging from 0.1% to 1% [36].

ML analysis led to five highly active precursors (hits) having a high level of binding affinity with the FYN kinase target. The study revealed a clear relationship between the kinetics of binding stability and the physical characteristics of the molecule. According to the data obtained after virtual screening, the best binding affinity was observed for CHEMBL1171837 with a score of -9.57 kcal/mol (Table 1). This high affinity is conditioned by the hydrogen bonding density created by the hit within the FYN kinase active site. The compound creates hydrogen bonds with conserved catalytic residues of the pocket – ASP148, GLU54, ASP92, and ASP130. Contrary to the above observations, one can conclude from the data in Table 2 that CHEMBL1171837 violates Lipinski's drug-likeness rules because of the high molecular weight (532.6 Da) and relative lipophilicity (logP: 4.46). High selectivity and binding affinity, however, suggest that these physiochemical limitations could be addressed through further optimization in the lead optimization stage. CHEMBL1421 shows the most favorable parameters in terms of physicochemical profile and affinity. Having the second place with a score of -8.49 kcal/mol, the hit creates hydrogen bonds with ASP148, LYS39, and ASN86, while creating complementary hydrophobic interactions with ALA147, ILE80, and GLY88 (Table 1). The key observation is that CHEMBL1421 satisfies all criteria of drug-likeness presented in Table 2 (Drug-Likeness Status: Pass). Ideal molecular weight (488.02 Da), logP value (3.31), number of donors/acceptors (HBD: 3, HBA: 9), and TPSA (106.51 Å2) provide this compound with high oral bioavailability. A thorough SHAP analysis performed (Figs. 5. and 6.) revealed a complete agreement between the ECFP4 bits necessary for the biological activity predicted by the model and those of aminothiazole and aminopyrimidine rings present in the structure of CHEMBL1421, thus confirming the physiological basis for ML predictions. One of the main evidences of the high chemical selectivity of our model is the presence of experimentally proven drugs CHEMBL1171837 and CHEMBL1421 among our top 5 hits, which are known potent SRC kinase inhibitors. These data clearly support the efficiency and predictability of ML model in screening such pharmacophores as aminothiazole, pyrimidine, imidazo-pyridazine, and purine rings, which have the ability to interact with the hinge and hydrophobic pockets in the ATP binding site of FYN kinase. This conclusion is based on the

molecular-level representation of bits discussed in SHAP analysis (for instance, a particular combination of nitrogen and sulfur atoms in adenine or aminothiazole ring).

The profiles of CHEMBL633 and its nucleoside/nucleotide analogs CHEMBL922 and CHEMBL1486 reflect classic pharmacophore dilemmas in *in silico* drug design. According to Table 1, CHEMBL633 exhibits a reasonable affinity with a score of -7.24 kcal/mol and forms hydrogen bonds with critical residues such as LYS39/ASP148; however, according to Table 2, it has a significantly higher logP value of 6.94. This extreme lipophilicity leads to a violation of drug-likeness status, suggesting that the molecule may exhibit non-specific membrane interactions and poor solubility problems. The nucleotide prodrug structures CHEMBL922 and CHEMBL1486 were characterized by weak Vina affinities (-6.82 and -6.32 kcal/mol) (Table 1) and multiple rule violations. This is clarified by examining the descriptors in Table 2: The phosphonate and ester protective groups CHEMBL922 and CHEMBL1486 dramatically increased the number of flexible bonds (Rot. Bonds: 11 and 15), respectively. High flexibility negatively affects the Vina score by amplifying the conformational entropy loss experienced by the ligand during binding. The presence of an excessive number of HBA of 13 and 17 atoms, respectively, resulted in TPSA values well above the limits (166.98 $Å^2$ and 260.0 $Å^2$). The high polar surface area suggests that these compounds may restrict passive membrane permeability. However, the purine (adenine) core carried by these molecules managed to be filtered by the fingerprint-based ML model because it structurally fits the ATP pocket geometry of kinases. In conclusion, the excellent agreement between the SHAP-based explainable AI findings and the quantitative data presented in Table 1 and Table 2 strongly suggests that the proposed *in silico* methodology can accelerate the rational design of novel therapeutic agents in FYN kinase-related pathologies (e.g., Alzheimer's Disease or specific oncological processes).

While the above outcomes are positive, it is critical to recognize that there are several limitations in this study. To begin with, the modest number of data points in the training set (906 compounds) may affect the model's ability to generalize, and the imbalance in the training set (36% active) was resolved using stratified sampling instead of more sophisticated resampling methods. Secondly, the virtual screening was conducted for only approved drugs (2,000 compounds), limiting the range of chemical structures for hit identification. Testing on extensive libraries (such as DrugSpaceX or ZINC15) may identify more chemical entities. Third, there was no implementation of the molecular dynamics approach, which provides detailed information about binding energy and hit conformations. Fourth, the ADMET predictions were made using computational tools and need experimental validation. Fifth, the *in vitro* and *in vivo* testing of hits is essential to demonstrate their FYN inhibition capabilities.

To summarize, the research proves that explainable ML is successfully incorporated into MD modeling for the identification of novel FYN kinase inhibitors. Using the SHAP approach, the model provides an interpretative explanation for the predictions which significantly increases the accuracy of the virtual screening process. From the identified compounds, the best five, particularly CHEMBL1171837, and CHEMBL1421 should undergo further experimental investigation in order to validate their ability to inhibit FYN kinase. Overall, a reproducible computational pipeline is developed which can be easily applied for other kinase types, making an essential contribution to the field of XAI.

Author Contributions

Conceptualization, A.T.D; methodology, A.T.D.; validation A.T.D.; formal analysis, A.T.D.; writing-review and editing, A.T.D.; visualization, A.T.D.

Funding
This research received no external funding.

Data Availability Statement
The following datas are available and will be provided by the corresponding author upon reasonable request:

- Dataset Curation and Preprocessing
- Model Hyperparameters
- Virtual Screening Protocol
- MD Protocol
- ADMET and Drug-Likeness Analysis
- Supplementary Figs. S1-S5: 2D Interaction Diagrams of the Top Five Hit Compounds
- Supplementary Figs. S6-S9: 3D binding mode visualizations generated with ChimeraX
- Supplementary Fig. S11: SHAP Dependence Plots for the Top Six Molecular Features
- Supplementary Fig. S12: Model Performance Comparison
- Supplementary Fig. S13: Confusion Matrix of the Optimized XGBoost Model on the Test Set
- Supplementary Table S1: Top 25 Virtual Screening Hits from the XGBoost Model
- Supplementary Table S2: Complete MD Results for the Top Five Hits
- Supplementary Table S3: ADMET and Drug-Likeness Properties of the Top Five Hits
- Supplementary Table S4: Protein-Ligand Interaction Distances for the Top Five Hits
- Supplementary Table S5: Dataset Summary Statistics

Conflicts of Interest
The author declares no competing financial interests or personal relationships that could have influenced the work reported in this paper.

Acknowledgement
The graphical abstract was created by the author using FigureLabs (free version; https://chat.figurelabs.ai/).

References

[1] Kohal, R., Bhavana, Kumari, P., Sharma, A.K., Gupta, G.D., &Verma, S.K. (2024). Fyn, Blk, and Lyn kinase inhibitors: A mini-review on medicinal attributes, research progress, and future insights. *Bioorganic & Medicinal Chemistry Letters*, 102, 129674. https://doi.org/10.1016/j.bmcl.2024.129674

[2] Bravo-Blas, A., Pirillo, C., Shergold, A., Andrusaite, A., & Roberts, E.W. (2022). Think global but act local: Tuning the dendritic cell response in cancer. *The International Journal of Biochemistry & Cell Biology*, 147, 106227. https://doi.org/10.1016/j.biocel.2022.106227

[3] Tang, L., Long, J., Li, K., Zhang, X., Chen, X., & Peng, C. (2020). A novel chalcone derivative suppresses melanoma cell growth through targeting Fyn/Stat3 pathway. *Cancer Cell International*, 20, 256. https://doi.org/10.1186/s12935-020-01336-2

[4] Kant, S., Kesarwani, P., Guastella, A.R., Kumar, P., Graham, S.F., Buelow, K.L., Nakano, I., & Chinnaiyan, P. (2020). Perhexiline Demonstrates FYN-mediated Antitumor Activity in Glioblastoma. *Molecular Cancer Therapeutics*, 19(7), 1471-1482. https://doi.org/10.1158/1535-7163.MCT-19-1047

[5] Moon, C.S. Reglero, C., Cortes, J.R., Quinn, S.A., Alvarez, S., Zhao, J., Lin, W.-H.W., Cooke, A.J., Abate, F., Soderquist, C.R., Fiñana, C., Inghirami, G., Campo, E., Bhagat, G., Rabadan, R., Palomero, T., & Ferrando, A.A. (2021). FYN-TRAF3IP2 induces NF-κB signaling-driven peripheral T-cell lymphoma. *Cancer*, 2, 98-113. https://doi.org/10.1038/s43018-020-00161-w

[6] Wu, X., Zhao, X., Zhou, C., Wei, N., Xu, Z., & Zhang, X. (2024). Prognostic and onco-immunological value of immune-related eRNAs-driven genes in lung adenocarcinoma. *Journal of Cancer Research and Clinical Oncology*, 150, 188. https://doi.org/10.1007/s00432-024-05687-5

[7] Dong, W., Sun, S.-J., Qin, J.-J., & Liu, G.-M. (2020). Fyn stimulates the progression of pancreatic cancer via Fyn-GluN2b-AKT axis. *European Review for Medical and Pharmacological Sciences*, 24(1), 109-121. https://doi.org/10.26355/eurrev_202001_19900

[8] Liu, W., Zhang, X., Liang, X., Guo, Y., Zhou, Z., Zhu, S., Peng, C., & Chen, X. (2025). Fyn inhibition by TAE684: A synergistic strategy to suppress melanoma and reverse vemurafenib resistance. *Cell Death & Disease*, 16, 796. https://doi.org/10.1038/s41419-025-08090-1

[9] Lee, G., Thangavel, R., Sharma, V.M., Litersky, J.M., Bhaskar, K., Fang, S.M., Do, L.H., Andreadis, A., Van Hoesen, G., & Ksiezak-Reding, H. (2004). Phosphorylation of tau by Fyn: Implications for Alzheimer's disease. *Journal of Neuroscience*, 24(9), 2304-2312. https://doi.org/10.1523/JNEUROSCI.4162-03.2004

[10] Iannuzzi, F., Sirabella, R., Canu, N., Maier, T. J., Annunziato, L., & Matrone, C. (2020). Fyn Tyrosine Kinase Elicits Amyloid Precursor Protein Tyr682 Phosphorylation in Neurons from Alzheimer's Disease Patients. *Cells*, 9(8), 1807. https://doi.org/10.3390/cells9081807

[11] Angelopoulou, E., Paudel, Y.N., Julian, T., Shaikh, M.F., & Piperi, C. (2021). Pivotal Role of Fyn Kinase in Parkinson's Disease and Levodopa-Induced Dyskinesia: a Novel Therapeutic Target? *Molecular Neurobiology*, 58, 1372-1391. https://doi.org/10.1007/s12035-020-02201-z

[12] Tang, S.J., Fesharaki-Zadeh, A., Takahashi, H., Nies, S.H., Smith, L.M., Luo, A., Chyung, A., Chiasseu, M., & Strittmatter, S.M. (2020). Fyn kinase inhibition reduces protein aggregation, increases synapse density and improves memory in transgenic and traumatic Tauopathy. *Acta Neuropathologica Communications*, 8, 96. https://doi.org/10.1186/s40478-020-00976-9

[13] Ta, H. M., Sankaran, B., Roush, E. D., Ferreon, J. C., Ferreon, A. C. M., & Kim, C. (2026). Fyn-Saracatinib Complex Structure Reveals an Active State-like Conformation. *International Journal of Molecular Sciences*, 27(3), 1143. https://doi.org/10.3390/ijms27031143

[14] Ramos, R., & Vale, N. (2024). Dual Drug Repurposing: The Example of Saracatinib. *International Journal of Molecular Sciences*, 25(8), 4565. https://doi.org/10.3390/ijms25084565

[15] Dorotea, D., Jiang, S., Pak, E.S., Son, J.B., Choi, H.G., Ahn, S.-M., & Ha, H. (2022). Pan-Src kinase inhibitor treatment attenuates diabetic kidney injury via inhibition of Fyn kinase-mediated endoplasmic reticulum stress. *Experimental & Molecular Medicine*, 54, 1086-1097. https://doi.org/10.1038/s12276-022-00810-3

[16] Demuro, S., Sauvey, C., Tripathi, S.K., Di Martino, R.M.C., Shi, D., Ortega, J.A., Russo, D., Balboni, B., Giabbai, B., Storici, P., Girotto, S., Abagyan, R., & Cavalli, A. (2022). ARN25068, a versatile starting point towards triple GSK-3β/FYN/DYRK1A inhibitors to tackle tau-related neurological disorders. *European Journal of Medicinal Chemistry*, 229, 114054. https://doi.org/10.1016/j.ejmech.2021.114054

[17] Giambattista, M., Francesca, M., Jose, O., Debora, R., Ilaria, P., Federico, F., Giorgia, B., Sabrina, P., Rita, S., Elena, R., Gert, F., Michela, R., Andrea, C., Barbara, M., Anna, M., & Filippo, B. (2025). Exploration of the Neuromodulatory Properties of Fyn and GSK-3β Kinases Exploiting 7-Azaindole-Based Inhibitors. *Journal of Medicinal Chemistry*, 68(16), 17130-17154. https://doi.org/10.1021/acs.jmedchem.5c00629

[18] Agarwal, V., Raorane, C.J., Gupta, A., Shastri, D., Raj, V., & Lee, S. (2025). Structure-based identification and experimental evaluation of Oroxin A as a FYN kinase inhibitor. *Journal of Computer-Aided Molecular Design*, 39, 120. https://doi.org/10.1007/s10822-025-00700-6

[19] Prabha, S., Choudhury, A., Saraswat, J., Patel, R., Hassan, Md. I., & Thakur, S.C. (2025). Identification of potential inhibitors of Fyn-Kinase from bioactive phytochemicals of *Berberis lycium* for therapeutic targeting of neurodegenerative disease. *Journal of Biomolecular Structure and Dynamics*, 43(14), 7811-7828. https://doi.org/10.1080/07391102.2025.2468296

[20] Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., & Zhao, S. (2019). Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, 18, 463-477 https://doi.org/10.1038/s41573-019-0024-5

[21] Lagemann, C., Lagemann, K., Mukherjee, S., & Schröder, W. (2021). Deep recurrent optical flow learning for particle image velocimetry data. *Nature Machine Intelligence*, 3, 641-651. https://doi.org/10.1038/s42256-021-00369-0

[22] Shevelev, O., Petrova, M., Smolensky, A., Osmonov, B., Toimatov, S., Kharybina, T., Karbainov, S., Ovchinnikov, L., Vesnin, S., Tarakanov, A., & Goryanin, I. (2022). Using medical microwave radiometry for brain temperature measurements. *Drug Discovery Today*, 27(3), 881-889. https://doi.org/10.1016/j.drudis.2021.11.004

[23] Choudhary, R., Walhekar, V., Muthal, A., Kumar, D., Bagul, C., & Kulkarni, R. (2023). Machine learning facilitated structural activity relationship approach for the discovery of novel inhibitors targeting EGFR. *Journal of Biomolecular Structure and Dynamics*, 41(22), 12445-12463. https://doi.org/10.1080/07391102.2023.2175263

[24] Salimi, A., Lim, J.H., Jang, J.H., & Lee, J.Y. (2022). The use of machine learning modeling, virtual screening, molecular docking, and molecular dynamics simulations to identify potential VEGFR2 kinase inhibitors. *Scientific Reports*, 12, 18825. https://doi.org/10.1038/s41598-022-22992-6

[25] Di Stefano, M., Galati, S., Ortore, G., Caligiuri, I., Rizzolio, F., Ceni, C., Bertini, S., Bononi, G., Granchi, C., Macchia, M., Poli, G., & Tuccinardi, T. (2022). Machine Learning-Based Virtual Screening for the Identification of Cdk5 Inhibitors. *International Journal of Molecular Sciences*, 23(18), 10653. https://doi.org/10.3390/ijms231810653

[26] He, X., Wan, Q., Zhou, Y., Yan, J., Zhao, Y., & Tang, F. (2026). Designing novel FAK inhibitors targeting gastric cancer: a combined approach using machine learning, docking analysis, molecular dynamics simulations, and experimental validation. *Journal of Computer-Aided Molecular Design*, 40, 38. https://doi.org/10.1007/s10822-025-00747-5

[27] Zhu, J., Li, K., Xu, L., Cai, Y., Chen, Y., Zhao, X., Li, H., Huang, G., & Jin, J. (2022). Discovery of novel selective PI3Kγ inhibitors through combining machine learning-based virtual screening with multiple protein structures and bio-evaluation. *Journal of Advanced Research*, 36, 1-13. https://doi.org/10.1016/j.jare.2021.04.007

[28] Lundberg, S.M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Curran Associates Inc.*, 30, 4765-4774. https://doi.org/10.48550/arXiv.1705.07874

[29] Erjavac, I., Kalafatovic, D., & Mauša, G. (2022). Coupled encoding methods for antimicrobial peptide prediction: How sensitive is a highly accurate model? *Artificial Intelligence in the Life Sciences*, 2, 100034. https://doi.org/10.1016/j.ailsci.2022.100034

[30] Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2, 56-67. https://doi.org/10.1038/s42256-019-0138-9

[31] Zhixiang, W., Zhongjie, H., Lianci, T., Xiaohan, S., Jingjie, S., Jianping, H., & Chunhua, L. (2023). Dynamic Insights into the Self-Activation Pathway and Allosteric Regulation of the Orphan G-Protein-Coupled Receptor GPR52. *Journal of Chemical Information and Modeling*, 63(18), 5847-5862. https://doi.org/10.1021/acs.jcim.3c00672

[32] Gopal, D., Muthuraj, R., Balaya, R.D.A., Kanekar, S., Ahmed, I., & Chandrasekaran, J. (2024). Computational discovery of novel FYN kinase inhibitors: a cheminformatics and machine learning-driven approach to targeted cancer and neurodegenerative therapy. *Molecular Diversity*, 28(6), 4343-4359. https://doi.org/10.1007/s11030-024-10819-7

[33] Bhat, S.S., Kulkarni, S.R., Uttarkar, A., & Niranjan, V. (2025). Computational Insights into Papaveroline as an In Silico Drug Candidate for Alzheimer's Disease via Fyn Tyrosine Kinase Inhibition. *Molecular Biotechnology*, 67, 2743-2757. https://doi.org/10.1007/s12033-024-01236-0

[34] Hang, N.T., Anh, T.D.H., Thanh, L.N., Anh, N.V., & Phuong, N.V. (2024). In silico screening of Fyn kinase inhibitors using classification-based QSAR model, molecular docking, molecular dynamics and ADME study. *Molecular Diversity*, 28, 2217-2228. https://doi.org/10.1007/s11030-024-10905-w

[35] Ljubič, M., Dolenc, M.S., Borišek, J., & Perdih, A. (2025). Water-Based Pharmacophore Modeling in Kinase Inhibitor Design: A Case Study on Fyn and Lyn Protein Kinases. *Journal of Chemical Information and Modeling*, 65(18), 9747-9761. https://doi.org/10.1021/acs.jcim.5c01478

[36] Macarron, R., Banks, M., Bojanic, D., Burns, D.J., Cirovic, D.A., Garyantes, T., Green, D.V.S., Hertzberg, R.P., Janzen, W.P., Paslay, J.W., Schopfer, U., & Sittampalam, G.S. (2011). Impact of high-throughput screening in biomedical research. *Nature Reviews Drug Discovery*, 10, 188-195. https://doi.org/10.1038/nrd3368