International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2023, 7 (2), 47-52
DOI: 10.36287/ijmsit.7.2.2
Received: August 3, 2023; Accepted: November 3, 2023

Detection of new candidate compounds against four antibiotic targets using explainable artificial intelligence by molecular fingerprints

Kevser Kübra Kırboğa [1,2], Naeem Abdul Ghafoor [3] and Ömür Baysal [3,*]

[1] *Bioengineering Department/Bilecik Seyh Edebali University, Bilecik, Türkiye (kubra.kirboga@bilecik.edu.tr) (ORCID: 0000-0000-0000-0000)*
[2] *Informatics Institute, Istanbul Technical University, Maslak, Istanbul, 34469, Türkiye*
[3*] *Faculty of Science, Department of Molecular Biology and Genetics, Molecular Microbiology Unit, Muğla Sıtkı Koçman University, Kötekli, Türkiye (omurbaysal@mu.edu.tr) (ORCID: 0000-0001-5104-0983, (N.A.G. ORCID ID: 0000-0002-4200-7679)*

Abstract – Antibiotic resistance is a threat that renders bacteria ineffective against antibiotics and makes it difficult to treat infections. Therefore, finding new target compounds is essential in discovering and developing new antibiotics. In this study, we developed an artificial intelligence algorithm that can predict and explain the pIC50 values for four antibiotic targets (Penicillin Binding Proteins (PB), β-Lactamase (BL), DNA Gyrase (DG), and Dihydrofolate Reductase(DR)). The algorithm uses molecular fingerprints of the molecules to predict the pIC50 values using the random forest regression method. We created the algorithm in a transparent and interpretable way. We used permutation feature importance (PFI) and Shapley explanations methods to identify the different molecular fingerprints that have the most influence on the pIC50 values. The results obtained from these methods show that different molecular fingerprints are essential for different antibiotic targets. According to the permutation importance results, KRFPC1646 (number of hydrogen bond donors of the compound) for BL and DR targets; 579 (a substructure with 5 bonded radius around the atom) for DG target; SubFPC182 (number of aromatic rings in the molecule) for PB target, are important fingerprints. With explainable artificial intelligence (XAI) (SHAP), KRFPC1646 (the number of hydrogen bond donors of the compound) for BL; KRFPC4274 (C1CCCCC1) for DR; 401 (C1CCCCC1) for DG; SubFPC182 (number of aromatic rings in the molecule) were determined as important fingerprints for PB. These results demonstrate the effectiveness and potential of using molecular fingerprints with explainable artificial intelligence to find new antibiotic candidates.

Keywords *– antibiotics, explainable artificial intelligence, shapley explanations, machine learning*

Citation: Kırboğa, Kevser., Ghafoor, Naeem., Baysal, Ömür. (2023). Detection of new candidate compounds against four antibiotic targets using artificial intelligence explained by molecular fingerprints. International Journal of Multidisciplinary Studies and Innovative Technologies, 7(2): 47-52

I. INTRODUCTION

Antibiotics have formed the basis of modern medicine since the discovery of Penicillin, and the continued effectiveness of these drugs is uncertain due to the global spread of antibiotic resistance[1]. Furthermore, the development of new antibiotics has been declining due to a lack of economic incentives in the private sector, exacerbating this problem [2, 3]. Antibiotic resistance is a problem that reduces the effectiveness of existing antibiotics and complicates the treatment of infections. Unfortunately, the current antimicrobial crisis, which is to blame for 700,000 deaths globally each year, is the result of the enticing gradual growth of antibiotic resistance [4]. Therefore, discovering new antibiotics can contribute to health by providing alternative options for treating resistant infections. Antibiotics are one of the cornerstones of modern medicine, and bacterial resistance is known to occur to a minimal extent immediately after or immediately after the introduction of an antimicrobial agent [5-9]. It is stated that there are significant differences between organisms and antibiotics at the time of the emergence of resistance. New antibiotics stimulate innovation in medicine and pharmacology. It allows for the improvement of existing treatment methods and the development of more effective, safer antibiotics. This diversifies treatment options and can help control infections more effectively. In the last two decades, critical perspectives such as resistance gene detection, genome sequencing, and rapid pathogen identification have been developed to develop new antibiotics. But technologies such as artificial intelligence (AI), machine learning (ML), and neural networks (NN) process enormous amounts of data almost instantly, ushering in a new golden age in drug discovery and synthesis [10, 11]. Table 1 summarizes AI's studies on antibiotic drugs. Traditional drug discovery techniques require high costs, long synthesis, testing and application processes, expensive equipment and extensive human resources, which are the most difficult to obtain. Alternatively, automated computer-assisted drug discovery techniques are more economical and rapid, enabling faster progression to preclinical and clinical testing phases [12].

Table 1. Artificial Intelligence technologies used in antibiotics studies.

Highlights	Technology	Ref.
The use of spectroscopy to identify specific biochemical fingerprints and machine learning to evaluate and predict the mode of action and potency of various antibiotics.	ML + high-throughput Fourier-transform infrared spectroscopy(FTIR)	[13]
Estimating phenotypic changes and antibacterial potency of various substances using a random forest model.	ML-random forest model	[14]
Estimation of possible antibacterial agents using DL and NN by searching databases.	DL + NN	[15]
Using ML to search and find possible candidates with beta-lactamase inhibition properties.	ML-random forest model	[16]
Using neural networks to distinguish between bacteriocin-containing sequences and those containing con-bacteriocin.	RNN	[17]
Using ANN to build computational chemistry models, identify and classify antimicrobial compounds.	ANN	[18]
Evaluating antimicrobial susceptibility and phenotypic polymyxin resistance using ML.	ML	[19]
Development of an ML algorithm for identifying toxin-like substances that also function as antimicrobial agents.	ML	[20]
Applying NN to examine the characteristics and composition of amino acids and peptides in order to find fresh antibacterial peptides.	NN	[21]
Using Mask-Loss 1D convolutional neural network (ML-ConvNet) for antibiotic resistance prediction in datasets with missing labels	ML-ConvNet	[22]
Using a ligand-based virtual scanning method, a deep neural network (DNN) model called a multilayer perceptron (MLP) is created to categorize molecules into "active" and "inactive" substances.	DNN, MLP	[23]
Examining the capacity of logistic regression, conditional trees and C5.0 rule-based models to evaluate the impact of critical interpretable prediction approaches using a dataset.	logistics regression	[24]

This study aims to apply an interpretable algorithm that supports the prodrug discovery stage against Penicillin Binding Proteins (PB), β-Lactamase (BL), DNA Gyrase (DG), and Dihydrofolate Reductase (DR) enzymes on the deficiency seen in the literature. Our study offers a transparent and interpretable perspective on new antibiotic discovery based on various molecular fingerprints (Figure 1).

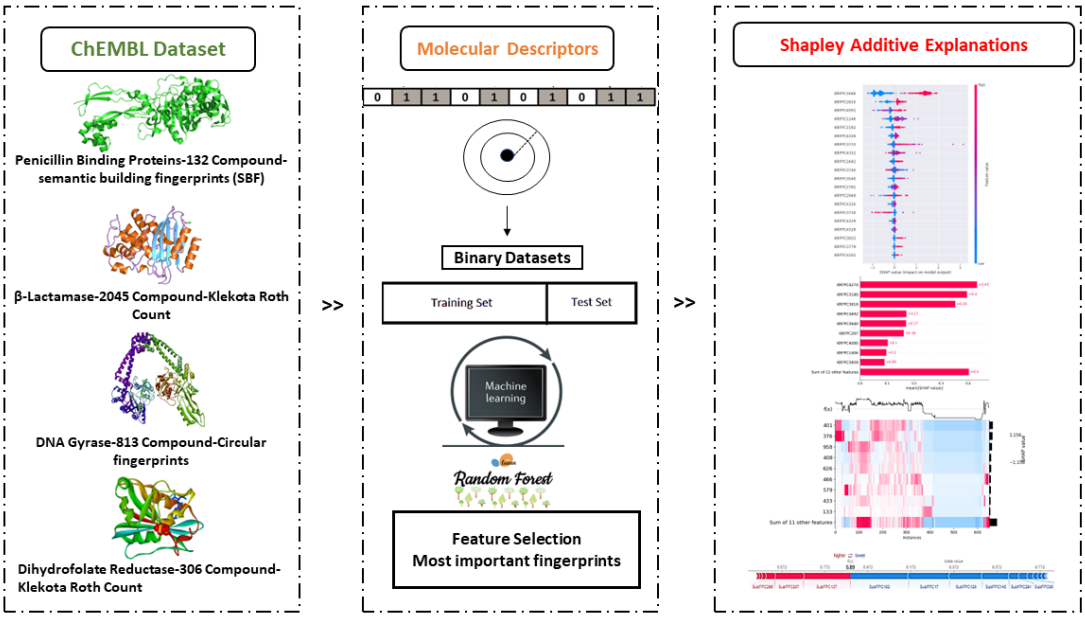


Fig. 1. Workflow of molecular fingerprint similarity study against Penicillin Binding Proteins, β-Lactamase, DNA Gyrase, and Dihydrofolate Reductase enzymes with explainable artificial intelligence methods.

II. Materials and Method

The examination of candidate antibiotics within the framework of molecular fingerprints (Klekota Roth Count, Circular Fingerprint and Infrastructure Fingerprint Number) by XAI is shown in Figure 1. In this study, new Penicillin Binding Proteins (PBP), and β-Lactamase (BL), DNA Gyrase (DG), and Dihydrofolate Reductase (DR) enzymes were analyzed using ChEMBL 32, PubChem, BindingDB databases. The molecular properties of anti-antibiotic candidates were predicted by XAI with the help of molecular fingerprints [25-27]. The research proceeds in the form of data collection, determination of molecular fingerprints, and application of ML and XAI. Python version 3.11 was used to conduct the research. Matplotlib version 3.7.1, pip version 22.0.4, Sklearn version 1.2.2, Pandas version 2.0.1, RDkit version 2023.3.1, Shap version 0.41.0, eli5 version 0.13.0, sci-kit-plot version 0.3.7, and NumPy version 1.24.3 were the

libraries used throughout the application. The application was carried out on a computer with Intel(R) Core (TM) i5-8300H CPU 2.30GHz, 64-bit operating system, x64-based processor, and 32 GB RAM.

Data Collection and Preprocessing

According to the results of the preliminary examinations on the ChEMBL and BindingDB databases, since the known inhibitors against the selected target protein were mainly determined based on IC50, the independent variable IC50 was selected in this study, and only compounds with known IC50 values were collected. The prepared library was constructed from the SMILES, ID (ChEMBL or PubChem SID/CID), compound detecting institution, and IC50 values of each compound. Since the library will consist of 4 different sources, in the preprocessing of the libraries, (i) all SMILES were converted to canonical format, (ii) they were free of repetitive compounds/SMILES, (iii) SMILES in crystalline form were desalted, and (iv) IC50 All values were linearized with the -log(M) transformation (pIC50) by converting Molar units.

Calculation of chemical descriptors and fingerprints

Different sets of "Attributes" were calculated from the SMILES of the compounds in the library to make the compounds in the four libraries suitable for use in machine learning applications. BL and DR are based on the Klekota Roth Count, DG Circular Fingerprint, and SBP Infrastructure Fingerprint Count attributes.

Permutation Feature Importance (PFI)

Due to the increasing complexity of ML models, better explanations are needed on how predictions are made and which input properties are most important in a model's decision. Also known as model explainability, providing clear details and reasons for ML predictions and performance, validating and improving models is essential for the ethical evaluation of model performance and reliability of results[28, 29]. One method that can be used to understand and explain the models' predictions is through the feature importance (FI) calculation, which estimates each feature's contribution to the model's predictions [30]. There are various FI techniques, but this article uses Permutation Feature Importance (PFI), a PFI technique that is very simple to implement and understand. RandomForestRegressor model and PFI method were applied to antibiotic candidates determined against PBP, BL, DG and DR enzymes. Among the data obtained against these enzymes, 20 important molecular fingerprints were identified, and pre-screened features for ML and XAI were used.

Explainable Artificial Intelligence (Shapley Additive exPlanation)

Shapley values are a concept in a collaborative game theory that aims to measure each player's contribution to the game. The method of obtaining Shapley values was proposed by Lioyd Shapley [31] in 1953. Shapley values emerge from the context in which "n" players collectively participate, and each of the "n" earns a "p" reward that is intended to be distributed fairly. Such a contribution is a Shapley value relative to the individual players' contribution. Shapley Additive Explanations (SHAP) is a method introduced by Lundberg and Lee in 2017 for interpreting the estimations of ML models through Shapely values [32]. The basic idea of SHAP is to calculate Shapley values for each feature of the sample to be interpreted, where each Shapley value represents the effect of the associated feature on the prediction. In this study, the TreeSHAP method was used to determine the molecular properties of chemical compounds. The TreeSHAP method is an effective method for describing the predictions of decision tree-based models. Decision trees to predict the molecular properties of chemical compounds can be annotated with the TreeSHAP method to determine the contribution of specific properties to the prediction [33-35].

III. RESULTS

Data Collection and Preprocessing

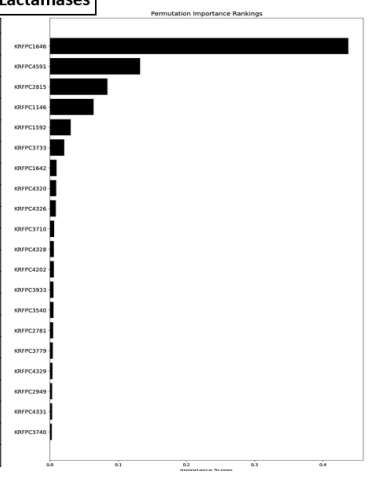
After preprocessing the data in the study, the size of the libraries prepared against each target protein group was finalized as follows: 2045 specific compounds for β-Lactamases (BL), 813 specific compounds for DNA Gyrases (DG), 306 specific compounds for Dihydrofolate Reductases (DR), 132 unique compounds for Penicillin Binding Proteins (PB) (**Fig. 2**).

Permutation Feature Importance (PFI)

According to the results of the Permutation Feature Importance (PFI), KRFPC1646 (0.438), KKRFPC4591 (0.132), and KRFPC2815 (0.085) fingerprints have an excellent importance score against the β-Lactamases (BL) target. KRFPC1646 (0.438), KRFPC3180 (0.133), and KRFPC4274(0.133) fingerprints are essential against the target of Dihydrofolate Reductases (DR). 579(0.054), 376 (0.054), and 401 (0.051) circular fingerprints against DNA Gyrases (DG) targets are essential. Against Penicillin Binding Proteins (PB) target, SubFPC182 (0.392), SubFPC17 (0.334), and SubFPC291 (0.08) fingerprints are significant (**Fig.2**).

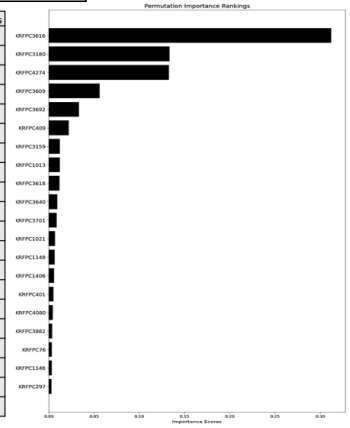
β-Lactamases

Klekota-Roth fingerprints	Importance Scores
KRFPC1646	0.438
KRFPC4591	0.132
KRFPC2815	0.085
KRFPC1146	0.064
KRFPC1592	0.031
KRFPC3733	0.021
KRFPC1642	0.01
KRFPC4320	0.009
KRFPC4326	0.009
KRFPC3710	0.006
KRFPC4328	0.006
KRFPC4202	0.006
KRFPC3933	0.006
KRFPC3540	0.005
KRFPC2781	0.005
KRFPC3779	0.004
KRFPC4329	0.004
KRFPC2949	0.003
KRFPC4331	0.003
KRFPC3740	0.003



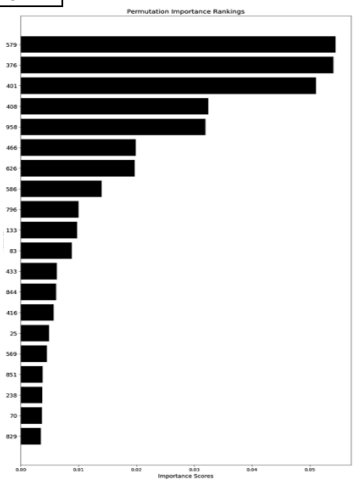
Dihydrofolate Reductase

Klekota-Roth fingerprints	Importance Scores
KRFPC1646	0.438
KRFPC3180	0.133
KRFPC4274	0.133
KRFPC3609	0.056
KRFPC3692	0.033
KRFPC409	0.022
KRFPC3159	0.012
KRFPC1013	0.012
KRFPC3618	0.012
KRFPC3640	0.009
KRFPC3701	0.009
KRFPC1021	0.007
KRFPC1148	0.007
KRFPC1406	0.006
KRFPC401	0.005
KRFPC4080	0.005
KRFPC3882	0.004
KRFPC76	0.004
KRFPC1146	0.003
KRFPC297	0.003



DNA Gyrase

Circular Fingerprints	Importance Scores
579	0.054
376	0.054
401	0.051
408	0.032
958	0.032
466	0.02
626	0.02
586	0.014
796	0.01
133	0.01
83	0.009
433	0.006
844	0.006
416	0.006
25	0.005
569	0.005
851	0.004
238	0.004
70	0.004
829	0.004



Penicillin Binding Proteins

SUB fingerprints	Importance Scores
SubFPC182	0.392
SubFPC17	0.334
SubFPC291	0.08
SubFPC128	0.064
SubFPC137	0.041
SubFPC143	0.018
SubFPC1	0.011
SubFPC209	0.008
SubFPC100	0.004
SubFPC135	0.004
SubFPC127	0.003
SubFPC287	0.002
SubFPC139	0.002
SubFPC14	0.001
SubFPC88	0.001
SubFPC296	0.001
SubFPC299	0.001
SubFPC84	0.001
SubFPC171	0.001
SubFPC298	0

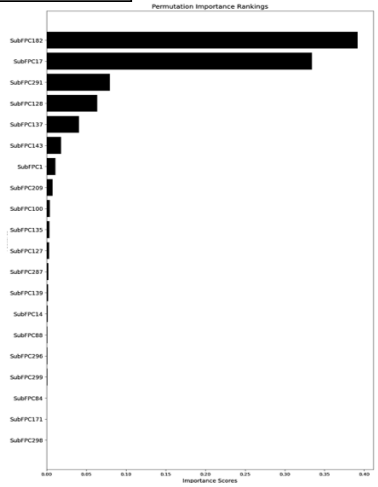


Fig. 2. Ranking of molecular fingerprints by Permutation Feature method.

Explainable Artificial Intelligence (Shapley Additive exPlanation)

When explainable artificial intelligence (SHAP) is applied to the studied data sets, KRFPC1646, KRFPC2815, KRFPC4591 for β-Lactamases (BL); KRFPC4274, KRFPC3180, KRFPC2815, KRFPC3616 for Dihydrofolate Reductases (DR); 401, 376, 958 for DNA Gyrases (DG); For Penicillin Binding Proteins (PB), SubFPC182, SubFPC17 and SubFPC128 were determined to be among the critical fingerprints (Fig.3).

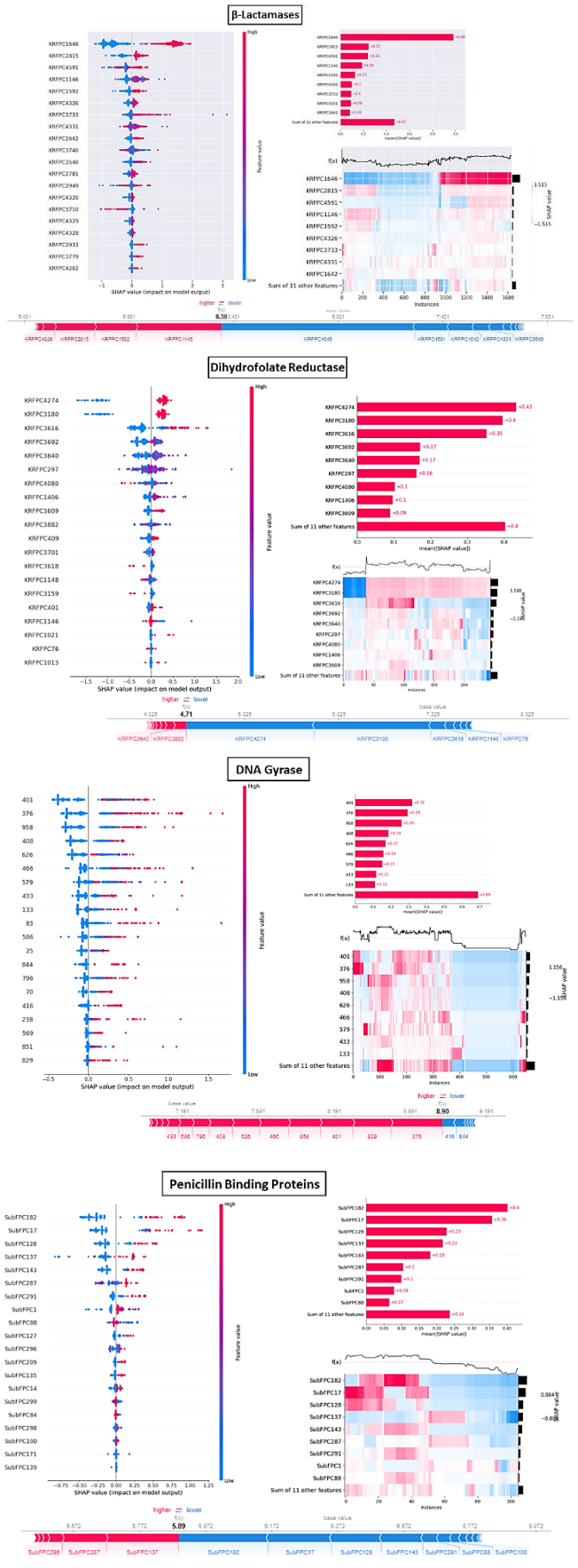


Fig. 3. Beeswarm, bar and heatmap graphics of SHAP analysis, one of the explainable artificial intelligence methods.

IV. DISCUSSION

In this study, we developed an explainable artificial intelligence (XAI) algorithm that can predict the pIC50 values for four antibiotic targets. Using the random forest regression method, the algorithm uses molecular fingerprints of the compounds to predict the pIC50 values. The algorithm also uses permutation feature importance and Shapley explanations methods to identify the molecular fingerprints that have the most influence on the pIC50 values. The results obtained from these methods show that different molecular fingerprints are essential for the four antibiotic targets. These results demonstrate the effectiveness and potential of using molecular fingerprints with XAI to find new antibiotic candidates. Molecular fingerprints are numerical vectors representing chemical compounds' structural or physicochemical properties [36].

The results of ShAP analyses show that different molecular fingerprints play an important role. Both PFI Dihydrofolate Reductases and SHAP β-Lactamases target and methods show KRFPC1646 as one of the important fingerprints. KRFPC1646 fingerprint indicates the number of hydrogen bond donors of the compound. This result may suggest that compounds with high hydrogen bond donors are more effective against β-Lactamases and Dihydrofolate Reductases. In our Shap results, we present graphs showing the contribution of each molecular fingerprint to the pIC50 value.

On the other hand, we see that the KRFPC1146 fingerprint has both the longest and the reddest bar. This indicates that the KRFPC1146 fingerprint is the most increasing and the most important fingerprint for the pIC50 value for the BL target. KRFPC1146 contains a pyridine ring, an amide group, and a nitrile group. We see that KRFPC4274 and KRFPC3180 fingerprints have both the longest and the bluest bars for Dihydrofolate Reductases. These fingerprints indicate that they are the most decreasing and the most important fingerprints for pIC50 value. KRFPC4274 indicates the number of hydroxyl groups in the molecule; KRFPC3180 indicates the number of carbonyl groups in the molecule. For DNA Gyrase, 401-376-958 fingerprints have a red bar. This indicates that these fingerprints are the most increasing and significant for pIC50 value. Circular fingerprints are topological fingerprints that encode substructures with a certain radius around each atom in the molecule [37]. The circular fingerprint 401 represents a substructure with 4 bonded radii around the atom [38]. Circular fingerprints 376 represent a substructure with 3 bonded radii around the atom. Of the circular fingerprints, 958 represent a hydroxyl group found in molecules. The hydroxyl group can affect the solubility and acidity of molecules [39]. For Penicillin Binding Proteins, we found that the SubFPC182, SubFPC17 and SubFPC128 fingerprints had both the longest and the bluest bar. SubFPC182 is the number of aromatic rings in the molecule; SubFPC17 is the number of hydrogen bond donors; SubFPC128 shows the number of carbonyl groups in the molecule [40]. It was seen that the SubFPC137 fingerprint had the reddest bar. This shows that the SubFPC317 fingerprint is the one that reduces the Pic50 value the most. SubFPC137 are fingerprints called substructure fingerprint count. This fingerprint shows the number of vinylogen ester groups in the molecule [41]. The vinylogen ester group is a functional group with a carbonyl group attached to an alkene. This group may be a property that may affect the interaction of the compound with penicillin-binding proteins. The SHAP and PFI methods results show that different molecular fingerprints are important for different antibiotic targets. These results demonstrate the efficacy and potential of using molecular fingerprints and XAI to find new antibiotic candidates. In addition to this study, some future studies are to estimate and compare pIC50 values with different molecular fingerprints and different machine learning (ML) methods, to estimate and explain pIC50 values for different antibiotic targets, to design new antibiotic candidates using predicted pIC50 values, and to synthesize, and experimentally confirm the predicted pIC50 values. The limitations of this study are that molecular fingerprints may not reflect all the structural or physicochemical properties of molecules, random forest regression may not be suitable for complex and high-dimensional data, permutation feature significance and Shapley annotations may be computationally costly, and pIC50 values may not be the only determinants of antibiotic efficacy.

V. CONCLUSION

The explainable algorithm estimates pIC50 values by random forest regression method using molecular fingerprints of molecules. The algorithm also uses permutation feature significance and Shapley annotations to identify the molecular fingerprints that have the most influence on pIC50 values. The results obtained with these methods show that different molecular fingerprints are essential for different antibiotic targets. These results demonstrate the efficacy and potential of using molecular fingerprints and XAI to find new antibiotic candidates.

ACKNOWLEDGMENT

This study emerged from the TUBITAK 1002, “Developing a Machine Learning-Based Bioinformatics Framework for the Identification of New Antibacterial Agents, 122E082”.

Authors' Contributions

The authors' contributions to the paper are equal.

Statement of Conflicts of Interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The authors declare that this study complies with Research and Publication Ethics

REFERENCES

[1] J. M. Stokes *et al.*, "A Deep Learning Approach to Antibiotic Discovery," *Cell*, vol. 180, no. 4, pp. 688-702.e13, 2020/02/20/ 2020, doi: https://doi.org/10.1016/j.cell.2020.01.021.

[2] E. D. Brown and G. D. Wright, "Antibacterial drug discovery in the resistance era," *Nature*, vol. 529, no. 7586, pp. 336-343, 2016.

[3] P. E. W. Trusts. "Five-year analysis shows continued deficiencies in antibiotic development." https://www.pewtrusts.org/en/research-and-analysis/data-visualizations/2019/five-year-analysis-shows-continued-deficiencies-in-antibiotic-development (accessed.

[4] J. O'Neill. "Antimicrobial Resistance:Tackling a crisis for the health and wealth of nations." https://www.ecdc.europa.eu/en/publications-data/ecdcemea-joint-technical-report-bacterial-challenge-time-react (accessed 29.05, 2023).

[5] A. P. Ball *et al.*, "Future trends in antimicrobial chemotherapy: expert opinion on the 43rd ICAAC," (in eng), *J Chemother*, vol. 16, no. 5, pp. 419-36, Oct 2004, doi: 10.1179/joc.2004.16.5.419.

[6] R. P. Bax *et al.*, "Antibiotic resistance - what can we do?," *Nature Medicine*, vol. 4, no. 5, pp. 545-546, 1998/05/01 1998, doi: 10.1038/nm0598-545.

[7] A. R. Coates and Y. Hu, "Novel approaches to developing new antibiotics for bacterial infections," (in eng), *Br J Pharmacol*, vol. 152, no. 8, pp. 1147-54, Dec 2007, doi: 10.1038/sj.bjp.0707432.

[8] K. Chaibi *et al.*, "What to Do with the New Antibiotics?," *ANTIBIOTICS-BASEL*, vol. 12, no. 4, APR 2023, Art no. 654, doi: 10.3390/antibiotics12040654.

[9] O. V. Kisil, N. I. Gabrielyan, and V. V. Maleev, "Antibiotic resistance - what can be done? A review," *TERAPEVTICHESKII ARKHIV*, vol. 95, no. 1, pp. 90-95, 2023, doi: 10.26442/00403660.2023.01.202040.

[10] K. K. Kırboğa, S. Abbasi, and E. U. Küçüksille, "Explainability and white box in drug discovery," *Chemical Biology & Drug Design*, vol. n/a, no. n/a, doi: https://doi.org/10.1111/cbdd.14262.

[11] L. David *et al.*, "Artificial Intelligence and Antibiotic Discovery," *Antibiotics*, vol. 10, no. 11, p. 1376, 2021. [Online]. Available: https://www.mdpi.com/2079-6382/10/11/1376.

[12] K. K. Kırboğa and E. U. Küçüksille, "Perspectives on Computer Aided Drug Discovery," (in en scheme="ISO639-1"), *11*, Review Articles 2023, doi: https://dergipark.org.tr/en/pub/dufed/issue/70232/1103457.

[13] B. Cunha, L. Fonseca, and C. Calado, "Simultaneous elucidation of antibiotics mechanism-of-action and potency with high-throughput fourier-transform Infrared spectroscopy and machine-learning," *App. Microb. Biot*, vol. 105, pp. 1269-1286, 2021.

[14] S. Zoffmann *et al.*, "Machine learning-powered antibiotics phenotypic drug discovery," *Scientific reports*, vol. 9, no. 1, p. 5013, 2019.

[15] J. M. Stokes *et al.*, "A deep learning approach to antibiotic discovery," *Cell*, vol. 180, no. 4, pp. 688-702. e13, 2020.

[16] N. Parvaiz, F. Ahmad, W. Yu, A. D. MacKerell Jr, and S. S. Azam, "Discovery of beta-lactamase CMY-10 inhibitors for combination therapy against multi-drug resistant Enterobacteriaceae," *PLoS One*, vol. 16, no. 1, p. e0244967, 2021.

[17] M.-N. Hamid and I. Friedberg, "Identifying antimicrobial peptides using word embedding with deep recurrent neural networks," *Bioinformatics*, vol. 35, no. 12, pp. 2009-2016, 2019.

[18] A. Badura, J. Krysiński, A. Nowaczyk, and A. Buciński, "Application of artificial neural networks to prediction of new substances with antimicrobial activity against Escherichia coli," *Journal of Applied Microbiology*, vol. 130, no. 1, pp. 40-49, 2021.

[19] N. Macesic, O. J. Bear Don't Walk IV, I. Pe'er, N. P. Tatonetti, A. Y. Peleg, and A.-C. Uhlemann, "Predicting phenotypic polymyxin resistance in Klebsiella pneumoniae through machine learning analysis of genomic data," *Msystems*, vol. 5, no. 3, pp. e00656-19, 2020.

[20] E. N. Grafskaia *et al.*, "Discovery of novel antimicrobial peptides: A transcriptomic study of the sea anemone Cnidopus japonicus," *Journal of Bioinformatics and Computational Biology*, vol. 16, no. 02, p. 1840006, 2018.

[21] X. Su, J. Xu, Y. Yin, X. Quan, and H. Zhang, "Antimicrobial peptide identification using multi-scale convolutional network," *BMC bioinformatics*, vol. 20, no. 1, pp. 1-10, 2019.

[22] M. Tharmakulasingam, B. Gardner, R. L. Ragione, and A. Fernando, "Explainable Deep Learning Approach for Multilabel Classification of Antimicrobial Resistance With Missing Labels," *IEEE Access*, vol. 10, pp. 113073-113085, 2022, doi: 10.1109/ACCESS.2022.3216896.

[23] M. I. Oladunjoye, O. O. Obe, and O. D. Alowolodu, *A deep neural network for the identification of lead molecules in antibiotics discovery* (Explainable Artificial Intelligence in Medical Decision Support Systems). 2022, pp. 381-400.

[24] B. Cánovas-Segura *et al.*, "Exploring Antimicrobial Resistance Prediction Using Post-hoc Interpretable Methods," in *Artificial Intelligence in Medicine: Knowledge Representation and Transparent and Explainable Systems*, Cham, M. Marcos *et al.*, Eds., 2019// 2019: Springer International Publishing, pp. 93-107.

[25] A. Gaulton *et al.*, "The ChEMBL database in 2017," *Nucleic Acids Research*, vol. 45, no. D1, pp. D945-D954, 2017-01-04 2017, doi: 10.1093/nar/gkw1074.

[26] T. Liu, Y. Lin, X. Wen, R. N. Jorissen, and M. K. Gilson, "BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities," (in eng), *Nucleic Acids Res*, vol. 35, no. Database issue, pp. D198-201, Jan 2007, doi: 10.1093/nar/gkl999.

[27] S. Kim *et al.*, "PubChem Substance and Compound databases," (in eng), *Nucleic Acids Res*, vol. 44, no. D1, pp. D1202-13, Jan 4 2016, doi: 10.1093/nar/gkv951.

[28] aporia-ai. "Permutation-importance." Github. https://github.com/aporia-ai/Permutation-importance/blob/main/Permutation%20importance/RegressionTask_carl_house.ipynb (accessed 23.05.2023, 2023).

[29] A. Altmann, L. Toloşi, O. Sander, and T. Lengauer, "Permutation importance: a corrected feature importance measure," *Bioinformatics*, vol. 26, no. 10, pp. 1340-1347, 2010.

[30] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. null, pp. 2825–2830, 2011.

[31] L. S. Shapley, "17. A Value for n-Person Games," in *Contributions to the Theory of Games (AM-28), Volume II*, K. Harold William and T. Albert William Eds. Princeton: Princeton University Press, 1953, pp. 307-318.

[32] S. Lundberg and S.-I. Lee, *A Unified Approach to Interpreting Model Predictions*. 2017.

[33] R. Mitchell, E. Frank, and G. Holmes, "GPUTreeShap: massively parallel exact calculation of SHAP scores for tree ensembles," (in eng), *PeerJ Comput Sci*, vol. 8, p. e880, 2022, doi: 10.7717/peerj-cs.880.

[34] C. Molnar, "Interpretable Machine Learning," *Self published*, 2020. [Online]. Available: https://christophm.github.io/interpretable-ml-book/.

[35] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "Consistent Individualized Feature Attribution for Tree Ensembles," 2019-03-07T00:06:09 2019.

[36] A. Capecchi, D. Probst, and J.-L. Reymond, "One molecular fingerprint to rule them all: drugs, biomolecules, and the metabolome," *Journal of Cheminformatics*, vol. 12, no. 1, p. 43, 2020/06/12 2020, doi: 10.1186/s13321-020-00445-4.

[37] R. L. Apodaca. "Computing Extended Connectivity Fingerprints." https://depth-first.com/articles/2019/01/11/extended-connectivity-fingerprints/ (accessed.

[38] C. M. Consulting. "Examples of Fingerprint and Descriptors." https://www.cambridgemedchemconsulting.com/resources/hit_identification/examples_descriptors.php (accessed.

[39] R. C. Glem, A. Bender, C. H. Arnby, L. Carlsson, S. Boyer, and J. Smith, "Circular fingerprints: flexible molecular descriptors with applications from physical chemistry to ADME," (in eng), *IDrugs*, vol. 9, no. 3, pp. 199-204, Mar 2006.

[40] N. Suvannang *et al.*, "Probing the origin of estrogen receptor alpha inhibition via large-scale QSAR study," *RSC Advances*, 10.1039/C7RA10979B vol. 8, no. 21, pp. 11344-11356, 2018, doi: 10.1039/C7RA10979B.

[41] C. Phanus-Umporn, W. Shoombuatong, V. Prachayasittikul, N. Anuwongcharoen, and C. Nantasenamat, "Privileged substructures for anti-sickling activity via cheminformatic analysis," (in eng), *RSC Adv*, vol. 8, no. 11, pp. 5920-5935, Feb 2 2018, doi: 10.1039/c7ra12079f.