Repurposing antitumoral drugs for multi-resistant infections

Filipe Barbosa de Vasconcelos, Anália Lourenço, and Maria Olívia Pereira

Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal

Abstract. The growing global issue of antimicrobial resistance needs the development of new and more rapid treatment techniques. The traditional medication development approach is becoming less sustainable due to high costs and long durations. In this context, drug repurposing is a promising alternative since it offers a faster, cost-effective alternative by identifying new uses for existing drugs. This study applies artificial intelligence-based methods, particularly machine learning and deep learning to explore drug-pathogen interactions using open-access datasets, aiming to enhance specificity and reduce toxicity.

Keywords: Drug Repurposing · Antimicrobial resistance · Artificial intelligence.

1 Introduction

1.1 Context and motivation

Currently, the treatment of choice to combat bacterial infections is the use of antibiotics and other antimicrobial products. The excessive use of these drugs has been contributing to bacteria rise their antimicrobial resistance (AMR)[38]. This recurring problem has become increasingly serious and is now listed by the World Health Organization (WHO) as one of the main global threats to public health, as there is an increase in AMR not only from bacteria, but also from viruses, fungi and parasites[44]; some of these pathogenic organisms are no longer even affected by the available treatments, leaving populations weaker and unprotected in terms of the risk of infection, illness and even death.

To address AMR, two strategies are commonly explored: traditional drug discovery, which is costly and time-consuming, and drug repurposing, which seeks new therapeutic applications for approved drugs, reducing time and cost [35].As a result, drug repurposing has become a reliable tool to identify new drugs and quickly address public health emergencies[4].

The biggest challenge right now is finding new associations between drugs and pathogens/diseases in the middle of an enormous amount of data[8]. Among the various drug repurposing methods, the most efficient is computational drug repurposing, which takes advantage of the recent big generation of free biological and medical data and the use of analytical methods based on Artificial Intelligence (AI), particularly Machine Learning (ML) and Deep Learning (DL)[39].

1.2 Objective

The main objective is to do drug repurposing as a more efficient and economical alternative to conventional drug discovery. Even when identified through in silico methods, repurposed drugs can lack specificity for new targets, leading to off-target effects and reduced efficacy. In addition, the range of possible drug candidates may be limited and additional optimization might be necessary for their new applications. Therefore, the use of computational drug repurposing aims to reduce or even solve all these problems[5,48].

2 Background

2.1 Antimicrobial Resistance

Antimicrobial resistance is a natural process amplified by the misuse of antibiotics, antivirals, antifungals, and antiparasitics. The WHO ranks AMR as one of the top ten global health hazards, citing its rising impact on death and morbidity caused by resistant diseases[44]. According to reports, if adequate measures are not adopted, AMR might kill up to 10 million people each year by 2050, overtaking cancer as the major cause of death[38]. The molecular processes that cause AMR include spontaneous genetic mutations and horizontal gene transfer via plasmids, transposons, and integrons. These conditions promote the transmission of resistance genes among bacterial species, resulting in the development of multidrug-resistant pathogens like those in the ESKAPE group[8,5]. Indeed, the most threatening pathogens, often showing resistance to multiple or all classes of antibiotics, are grouped under the acronym ESKAPE responsible for the majority of nosocomial infections worldwide due to their ability to "escape" conventional antibiotics[40].

The COVID-19 pandemic has exacerbated the AMR situation by encouraging extensive antibiotic usage in hospitalized patients, even without verified bacterial illnesses [40]. This, along with strain on healthcare institutions and disruptions to surveillance efforts, has expedited the development of resistance. Research is investigating solutions such as novel antibiotic classes, phage treatment, and nanotechnology-based antimicrobials [23].

2.2 Drug Repurposing

Cancer remains one of the leading causes of death worldwide, driving continuous efforts to develop more effective therapies [36]. While classical chemotherapeutic agents have long played a central role in treatment, their use is often limited by toxicity, lack of specificity, and drug resistance [46]. Advances such as targeted therapies, immunotherapies, and combination regimens have improved outcomes, yet treating refractory and multidrug-resistant cancers remains a major challenge [46,36].

Drug repurposing is a promising and cost-effective method for finding new therapeutic uses for existing and licensed medications[35]. Unlike traditional

drug discovery, repurposing uses clinically existing medications to drastically reduce development time and failure rates[22]. This strategy has proven effective in severe public health situations, such as the COVID-19 pandemic, where quick treatment deployment was critical[39]. Beyond infectious illnesses, drug repurposing has gained popularity in oncology, where non-cancer therapies such as antimicrobials and cardiovascular medications have been revealed to have surprising anticancer effects[9,49].

The promise of repurposing is not just its efficiency, but also its ability to find unanticipated pharmacological effects, particularly when led by large-scale biological data and computer modeling. The application of AI-driven methods has significantly improved medication repurposing efforts[5]. ML and DL algorithms have been extremely useful in predicting drug-disease interactions, finding off-target effects, providing tools to mine existing datasets, and ranking candidate compounds with high translational potential[47]. The convergence of systems biology, cheminformatics, and omics technology has enabled the creation of computational frameworks for assessing drug-disease interactions beyond their intended applications. This is especially important in the context of AMR and multidrug-resistant diseases, where repurposed antitumoral and non-antibiotic medications may serve as alternative therapeutics[13].

2.3 Bioinformatic approach

AI in drug discovery To analyze the vast number of datasets, including drug-target interactions, genetic sequences and pharmacokinetics, needed to repurpose drugs is still to limited and time-consuming[31]. Thus, there is a need to execute new techniques sell-sufficient to overcome this problem, and that is where AI emerges as a powerful tool to resolve this problem[28]. The application of AI in the drug discovery area allows for the prediction and minimization of potential off-target interactions, analyzing both protein sequences and chemical structures, maximizing efficacy and minimizing toxicity. This helps researchers save time by focusing on drugs with the best chances of success[11].

Machine Learning has emerged as a transformational force in medication repurposing, allowing for predictive modeling, complicated pattern identification, and high-throughput hypothesis development from multidimensional biological data[3,8]. Training on datasets containing molecular structures, gene-disease associations, and pharmacological profiles allows ML algorithms to identify previously unknown therapeutic potentials in existing drugs with greater speed and scalability than traditional experimental pipelines[39].

One of the most significant advantages of machine learning is its ability to manage high-dimensional and diverse data, such as genomes, proteomics, chemical descriptors, and clinical outcomes. Random forests, support vector machines (SVM), and graph-based models have all been used extensively to repurpose frameworks, predict drug-target interactions, and infer mechanisms of action[47,10]. Furthermore, ML has shown to be useful in prioritizing repurposing options for AMR. As traditional antibiotics lose potency, combining ML

models with biological networks and phenotypic screening data aids in identifying non-antibiotic medicines with potential antibacterial properties[4].

Deep Learning allows computational models to learn representations of data that are more abstract. It tries to discover structures in, usually, larger data sets by using backpropagation to show the machine how it should change its own parameters, which are used to compute each layer, from the representations it has learned from previous layers[18].

Deep learning has become popular due to its ability to uncover latent patterns from high-dimensional biological data. Convolutional neural networks (CNNs) and autoencoders can be trained using structured datasets such as ChEMBL and DrugBank[25,42]. These models can anticipate drug-target interactions, toxicity profiles, and synergistic effects in combination therapy. The growing availability of open-access databases such as Drug Repurposing Hub and BindingDB has boosted ML by providing well-annotated training datasets[42]. When combined with unsupervised or semi-supervised learning approaches, these datasets allow for the development of innovative hypotheses in pharmacology with minimum experimental input.

3 Methodology

3.1 Search and Study selection

We conducted our review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) standards[12]. A detailed search was conducted on Google Scholar for full-length

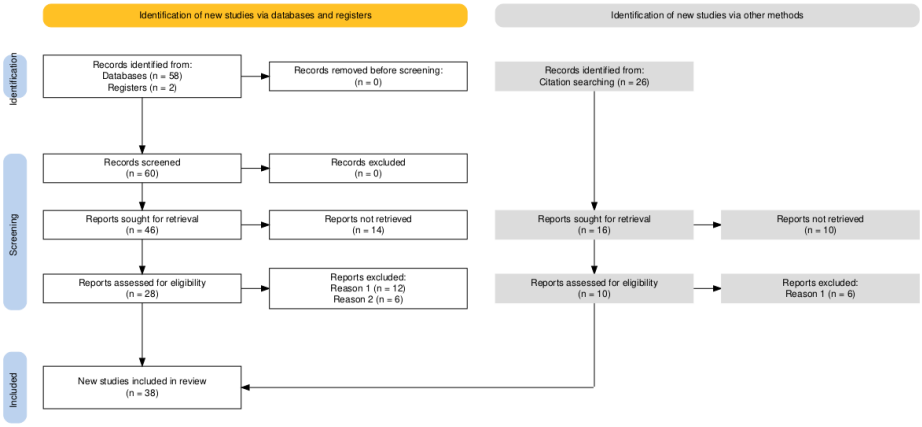

Fig. 1. Flow chart for article selection and filtering

research publications published between January 2020 and March 2025 using the

keywords: "machine learning", "deep learning", "drug repurposing" and "antimicrobial resistance".

Our search yielded 87 items, which were then narrowed down to 38 relevant research articles after screening for computational drug repurposing and drug data utilization. There were 16 papers found in PubMed, 12 papers found in Nature, and 10 other papers from citation searching, which were chosen for their focus on ML-based AMR prediction. Figure 1 presents an overview of the selection procedure.

3.2 Review of selected literature

Table 1. Summary of predictive models for drug discovery and their associated code. References are listed in the bibliography.

Ref.	Date	Predictive model	Code
[16]	2024	TxGNN	https://github.com/mims-harvard/TxGNN
[49]	2024	Shennong framework	https://github.com/PeijingZhang/Shennong
[27]	2020	BacEffluxPred	http://proteininformatics.org/mkumar/baceffluxpred/
[42]	2022	DrugRepo	https://github.com/futurerepo/drugrepo
[32]	2021	DRIAD	https://github.com/labsyspharm/DRIADrc
[20]	2020	VAMPr	https://github.com/jiwoongbio/VAMPr
[15]	2020	MM-RBM	https://github.com/LBBSoft/Multimodal-Drug-Repurposing.git
[37]	2020	Chemprop	https://github.com/swansonk14/chemprop
[41]	2024	APEX	https://gitlab.com/machine-biology-group-public/apex
[29]	2020	MOSES	https://github.com/molecularsets/moses
[17]	2022	Chemformer	https://github.com/MolecularAI/Chemformer
[45]	2021	MARS	https://github.com/yutxie/mars
[43]	2021	CCMDR	https://github.com/HoytWen/CCMDR

In addition, we included publications that apply artificial intelligence models in this context in Table 1. This table summarizes recent machine learning and deep learning models applied to drug repurposing, including model names, publication dates, and links to source code to emphasize reproducibility.

Recent advances in artificial intelligence, particularly deep learning, have significantly impacted drug repurposing. Among these, TxGNN, proposed by Wang et al., uses Graph Neural Networks to model biomedical knowledge graphs, achieving strong performance in predicting drug–disease associations [42]. Similarly, Shennong, introduced by Zhang et al., combines molecular embeddings with biological data to effectively predict drug–target interactions across multiple tasks [49].

To address interpretability, APEX offers an explainable machine learning pipeline for drug repositioning, integrating predictive modeling with result visu-

alization [41]. VAMPr applies explainable AI to predict antimicrobial resistance from genomic data, delivering accurate and interpretable results [20].

Other neural models include Chemprop, which uses message-passing neural networks to predict molecular properties from SMILE, and Chemformer, a transformer-based model pretrained on chemical data, supporting tasks like molecular translation and bioactivity prediction [37,17].

MARS adopts multi-task learning to identify antimicrobial resistance signatures in genomic data, while CCMDR leverages contrastive learning to distinguish valid drug–indication pairs, improving precision in repositioning [45,43].

MOSES provides a standardized platform to benchmark generative molecular models, and DrugRepo is a curated ML-driven database for predicting new drug indications from pharmacological data [29,4].

MM-RBM applies multimodal restricted Boltzmann machines to integrate diverse biomedical data for repositioning, while DRIAD uses deep learning to infer drug response across diseases by combining transcriptomic and pharmacological profiles [15,32]. Lastly, BacEffluxPred is an SVM-based tool for identifying bacterial efflux pump inhibitors, aiding virtual screening of potential repositioning candidates [27].

3.3 Data loading and preparation

Our primary data source for drug targets was DrugBank, a comprehensive repository of approved and experimental drugs along with their known protein targets[21]. We started by extracting the list of drug-target interactions and dividing it into two datasets: antimicrobial and antitumoral. The antimicrobial category was defined broadly to include antibiotics and drugs that were relevant to AMR; the antitumoral category included all anti-cancer drugs like chemotherapeutic agents and targeted therapies. To construct the positive and negative sets, the protein sequences of the antimicrobial targets were extracted from the DrugBank FASTA file. Each protein sequence corresponding to the target of at least one known antimicrobial drug was labelled as a positive instance. Conversely, sequences not associated with any antimicrobial drug in DrugBank were labelled as negative instances.

The resulting labelled dataset comprised approximately 1,945 protein sequences in total, with roughly 360 positives and 1,585 negatives. We randomly partitioned these data into a training set by 90% and a hold-out test set by 10%, maintaining the class proportions in each subset to ensure a representative distribution of positive and negative examples in both sets. This split presented an imbalanced class distribution in the training data, with positive sequences greatly outnumbered by negatives.

To address this class imbalance, we applied the Synthetic Minority Oversampling Technique (SMOTE) on the training set. SMOTE generates synthetic minority-class examples by interpolating between existing positive instances, thereby augmenting the positive class[2,33]. Using this approach, we increased the representation of positive sequences in the training data to achieve a more

balanced class distribution for model learning, while leaving the test set unchanged for unbiased evaluation.

3.4 Feature extraction

From the curated datasets, we extracted a set of physicochemical features for each drug based on amino acid sequence analysis. For peptide- or protein-based drugs, the amino acid sequence of the drug itself was obtained from DrugBank records; for small-molecule drugs, the amino acid sequence of the drug's primary protein target was used as a proxy, since small molecules have no inherent sequence. We utilized Biopython's ProtParam *ProteinAnalysis* module to compute six key features from each sequence: the sequence length, molecular weight, aromaticity, instability index, isoelectric point, and GRAVY (Grand Average of Hydropathy), indicating overall hydrophobicity[1,26]. These features were chosen to capture size, compositional biases, stability, charge properties, and hydrophobic character of each protein sequence, which are relevant attributes in distinguishing antimicrobial peptides. The resulting feature vectors constituted the input for model training. All feature extraction was performed uniformly for both antimicrobial and antitumoral drug entries to ensure consistency in the machine learning pipeline.

3.5 Machine learning pipeline

We designed a machine learning pipeline to train predictive models and identify antitumoral drugs with potential antimicrobial activity. First, the prepared dataset was used to train three different classifiers: a Random Forest (RF) ensemble, a Support Vector Machine (SVM), and a Gaussian Naïve Bayes (GNB) classifier. We selected these algorithms to provide a diversity of modelling approaches: RF as a non-linear ensemble of decision trees that can capture complex feature interactions, SVM as a robust linear/non-linear classifier known for solid performance in high-dimensional spaces, and GNB as a simple probabilistic model that can serve as a baseline[27,19,14]. All models were implemented using the scikit-learn library.

For training and validation, we employed a cross-validation strategy to reliably estimate performance. In particular, we used a 10-fold cross-validation on the training data: the data were shuffled and split into 10 folds, models were trained on 9 folds and validated on the remaining fold, iteratively, and performance metrics were averaged. During cross-validation, class stratification was maintained to ensure each fold had a representative class balance.

Finally, we deployed the trained models on the antitumoral drug set to predict which of those drugs might have antimicrobial activity. A drug was considered a repurposing candidate if the model classified at least one of its targets as "antimicrobial-like". We recorded the set of candidate drugs identified by each model, along with their prediction confidence scores. These results were then analyzed collectively to determine consensus predictions and to perform further analysis on the candidates.

4 Results

4.1 Predictive performance

The three machine learning models were evaluated on their ability to distinguish known antimicrobial drugs from antitumoral drugs, using the features derived from drug target proteins. Table 2 presents a summary of the performance metrics for each model, including accuracy, precision, recall, and F1-score for the positive class, as well as the ROC-AUC. Overall, the Random Forest classifier achieved the best performance, with an accuracy of about 81% and the highest F1-score. The RF model showed a balanced profile, managing to capture a large portion of the antimicrobial class while also maintaining good precision in predicting positives. This indicates that the ensemble approach was effective at modelling the complex feature patterns separating antimicrobial and cancer drug targets. The Gaussian Naïve Bayes classifier had a moderate performance: it achieved an accuracy near 70%, with precision around 85% but a lower recall, around 60%. This suggests that GNB was conservative in labelling the positives, likely due to its simplistic assumptions not fully capturing the feature correlations. The SVM had the lowest performance in our evaluations, with an accuracy of only 59% and an F1-score around 52%. Its recall was particularly low, meaning it missed about half of the known antimicrobials in the validation splits. Although the SVM's ROC-AUC was 77%, indicating that while the other classifications were suboptimal, it did rank instances moderately well by score. In contrast, the RF's ROC-AUC was around 90%, reflecting an excellent overall discrimination, and GNB's ROC-AUC was 82%, in line with its intermediate performance.

Table 2. Performance of machine learning models in model training

Model	Precision	Recall	F1-score	Accuracy
Random Forest	0.90	0.87	0.88	0.91
SVM	0.84	0.86	0.85	0.89
Gaussian Naïve Bayes	0.78	0.66	0.72	0.81

4.2 Antitumoral predictions

After validating the models, we applied them to the set of antitumoral drugs to predict which could be repurposed as antimicrobials. Each model produced a list of predicted positive drugs. We found that the Random Forest model, being more sensitive, identified a larger number of antitumoral drugs as potential hits, whereas the Naïve Bayes gave a smaller, more conservative list. The SVM identified the fewest candidates.

We compiled the predictions into Table 3, which lists a selection of the top repurposing candidate drugs, along with their original indication and which

model(s) predicted them as positives. In this table, we highlight consensus candidates as well as those predicted by at least two models. Among the consensus candidates were several drugs that have notable connections to antimicrobial action. For instance, Bleomycin (BLM), an antitumoral chemotherapeutic drug, was identified by RF and GNB. BLM is a product of *Streptomyces* bacteria and is known to have antibiotic properties, although its clinical use has been in oncology[7]. Dactinomycin (Actinomycin D), an antineoplastic that intercalates DNA, was predicted by all three models[7]. Similarly, Mitomycin-C, an anticancer drug that is also an antibiotic compound, appeared as a positive prediction in all models' outputs[34]. The models also predicted anthracyclines like Doxorubicin (Adriamycin) and Daunorubicin, these drugs are derived from bacterial sources and kill cells by targeting Topoisomerase II[24,30].

From the overlapping predictions, we infer that the more models agree on a drug, the higher the confidence in that drug's repurposing potential. For instance, Mitomycin-C might be a high-priority candidate for laboratory testing, since three independent algorithms all converged on it. On the other hand, unique predictions like Imatinib, for example, could either be false positives or findings missed by other methods, needing a case-by-case examination.

Table 3. Selected antitumoral drugs predicted as potential repurposing candidates. "✓" indicates a positive prediction by that model. Drugs predicted by all three methods are highlighted.

Drug Name	**Original Indication**	**RF**	**SVM**	**GNB**
Dactinomycin	Wilms' tumor, sarcomas, chemotherapy	✓	✓	✓
Mitomycin-C	Gastric/pancreatic cancer , chemotherapy	✓	✓	✓
Bleomycin	Hodgkin's lymphoma, testicular cancer , chemotherapy	✓	–	✓
Doxorubicin	Breast cancer, leukemia , chemotherapy	✓	–	✓
Methotrexate	Leukemia, rheumatoid arthritis, antimetabolite	✓	–	✓
Imatinib	Chronic myeloid leukemia, chemotherapy	✓	–	–
Paclitaxel	Breast, ovarian cancer, mitotic inhibitor	✓	–	–
Chlorambucil	Chronic lymphocytic leukemia, alkylating agent	–	–	✓
Daunorubicin	Leukemia, anthracycline antibiotic	✓	✓	✓
Cladribine	Hairy cell leukemia, antimetabolite	✓	✓	✓

5 Discussion

In this study, we explored the repurposing of antitumoral drugs as potential treatments for multi-resistant bacterial infections using machine learning methods. The results demonstrate both the promise and the challenges of applying machine learning to drug repurposing in the context of antimicrobial resistance.

One positive outcome is the identification of several antitumoral drugs with predicted antimicrobial activity, especially those that were consistently flagged by all models. This convergence between our model predictions and known pharmacological properties provides a measure of validation for the approach. It suggests that the models, particularly the Random Forest, learned meaningful distinctions. The overlap of predictions also implies that combining different ML algorithms can increase confidence in certain results; when an antitumor drug is independently identified by algorithms with different biases and assumptions, it strengthens the case for that drug's potential to act as an antibiotic.

Despite these successes, the approach also has clear limitations. One challenge was the imbalance and heterogeneity of data. The number of known antimicrobials was smaller than antitumoral drugs, and their feature distribution was quite distinct. We softened this by oversampling with SMOTE, but the SVM's lower performance indicates that some models may struggle in this setting. For example, the SVM's poor sensitivity suggests it might have overlooked candidates that Random Forest and Gaussian Naïve Bayes caught, whereas Random Forest's broader net means it could have captured some false positives that GNB filtered out.

Additionally, the text-mining approach to build the dataset, using keywords to define antimicrobials and antitumorals, has its own weaknesses. It is possible some drugs were misclassified or omitted. For instance, some antitumoral drugs that also have known antimicrobial effects might have been labelled only as antitumoral in our set if their DrugBank entries didn't clearly list a bacterial target. On the other side, a few drugs might have been included as antimicrobial because of a bacterial-related keyword in their entry, even if they are not primarily used as antibiotics, for example, certain antivirals or antiparasitics could have been selected if they mention bacteria in their description. We attempted to minimize such errors, but it is an area for improvement.

The Random Forest's superior performance implies that nonlinear combinations of the protein features were important to distinguish antimicrobial targets from human cancer targets. The RF's variable importance analysis highlighted some intuitive patterns: for example, protein sequence length and isoelectric point emerged as influential features. Many known antibiotic targets tend to be smaller and have different pI ranges compared to human signalling proteins targeted by antitumoral drugs[6]. Additionally, the presence of certain amino acid composition signatures helped the RF model identify bacterial proteins. The SVM, being less flexible with default hyperparameters, may not have captured these nonlinear patterns as effectively, whereas the ensemble of trees in the RF could partition the feature space more distinctively.

Overall, the cross-validation results gave us confidence to proceed with using these models for prediction on antitumoral drugs. We decided to primarily rely on the Random Forest model's predictions for identifying candidates, given its better sensitivity-specificity balance, while using the other models to cross-verify certain predictions. The performance evaluation also underscored the need for further improvement; for instance, SVM might benefit from advanced kernels or

feature selection, and more sophisticated methods, like deep learning on protein sequences or molecular graphs, could potentially push accuracy higher. Nonetheless, the chosen models were sufficient to provide a list of promising candidate drugs for repurposing.

6 Conclusion

We presented a computational study aimed at repurposing antitumoral drugs to combat multidrug-resistant bacterial infections. By constructing a dataset from DrugBank and applying machine learning models, we successfully identified several approved antitumoral drugs with potential antimicrobial properties. The Random Forest model emerged as the most effective predictor, and consensus among different models highlighted a set of high-confidence candidates, including drugs with known antibiotic origins and some new suggestions. These findings contribute to the growing evidence that *in silico* approaches can uncover unconventional uses for existing drugs, which is particularly crucial in the fight against antimicrobial resistance.

In conclusion, this project underscores the utility and value of drug repurposing in addressing urgent health threats. As antimicrobial resistance continues to rise, repurposed drugs provide a hopeful strategy to fill the antibiotic pipeline with minimal delay. Our results, while requiring experimental confirmation, lay the groundwork for subsequent research and development. Future work will focus on updating the model with new data, including feedback from lab results, thereby creating an evolving AI-driven platform for drug repurposing. Ultimately, the convergence of bioinformatics and pharmacology showed that we can accelerate the discovery of new treatments.

References

1. Alberto, N., Venanzi, E., Basciu, A., Vargiu, A.V., Kiparissides, A., Dalby, P.A., Dikicioglu, D.: Machine learning integrating protein structure, sequence, and dynamics to predict the enzyme activity of bovine enterokinase variants. J. Chem. Inf. Model **64**, 2025 (2024). https://doi.org/10.1021/acs.jcim.3c00999, https://doi.org/10.1021/acs.jcim.3c00999
2. Altalhan, M., Algarni, A., Alouane, M.T.H.: Imbalanced data problem in machine learning: A review. IEEE Access **13**, 13686–13699 (2025). https://doi.org/10.1109/ACCESS.2025.3531662
3. Anahtar, M.N., Yang, J.H., Kanjilal, S.: Applications of machine learning to the problem of antimicrobial resistance: an emerging model for translational research. Journal of Clinical Microbiology **59**, e01260–20 (6 2021). https://doi.org/10.1128/JCM.01260-20, https://pmc.ncbi.nlm.nih.gov/articles/PMC8218744/
4. Anokian, E., Bernett, J., Freeman, A., List, M., Santamaría, L.P., Tanoli, Z., Bonnin, S.: Machine learning and artificial intelligence in drug repurposing—challenges and perspectives. Drug Repurposing **1**, 20240004 (7 2024). https://doi.org/10.58647/DRUGREPO.24.1.0004, https://drugrepocentral.scienceopen.com/hosted-document?doi=10.58647/DRUGREPO.24.1.0004

5. Arnold, A., McLellan, S., Stokes, J.M.: How ai can help us beat amr. npj Antimicrobials and Resistance 2025 3:1 **3**, 1–15 (3 2025). https://doi.org/10.1038/s44259-025-00085-4, https://www.nature.com/articles/s44259-025-00085-4
6. Bakheet, T.M., Doig, A.J.: Properties and identification of antibiotic drug targets. BMC Bioinformatics **11**, 195 (4 2010). https://doi.org/10.1186/1471-2105-11-195, https://pmc.ncbi.nlm.nih.gov/articles/PMC2873537/
7. Bano, N., Parveen, S., Saeed, M., Siddiqui, S., Abohassan, M., Mir, S.S.: Drug repurposing of selected antibiotics: An emerging approach in cancer drug discovery (2024). https://doi.org/10.1021/acsomega.4c00617, https://doi.org/10.1021/acsomega.4c00617
8. Cesaro, A., Hoffman, S.C., Das, P., de la Fuente-Nunez, C.: Challenges and applications of artificial intelligence in infectious diseases and antimicrobial resistance. npj Antimicrobials and Resistance 2025 3:1 **3**, 1–10 (1 2025). https://doi.org/10.1038/s44259-024-00068-x, https://www.nature.com/articles/s44259-024-00068-x
9. Dallavalle, S., Dobričić, V., Lazzarato, L., Gazzano, E., Machuqueiro, M., Pajeva, I., Tsakovska, I., Zidar, N., Fruttero, R.: Improvement of conventional anti-cancer drugs as new tools against multidrug resistant tumors. Drug Resistance Updates **50**, 100682 (5 2020). https://doi.org/10.1016/J.DRUP.2020.100682
10. Greener, J.G., Kandathil, S.M., Moffat, L., Jones, D.T.: A guide to machine learning for biologists. Nature Reviews Molecular Cell Biology 2021 23:1 **23**, 40–55 (9 2021). https://doi.org/10.1038/s41580-021-00407-0, https://www.nature.com/articles/s41580-021-00407-0
11. Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R.K., Kumar, P.: Artificial intelligence to deep learning: machine intelligence approach for drug discovery. Molecular Diversity 2021 25:3 **25**, 1315–1360 (4 2021). https://doi.org/10.1007/S11030-021-10217-3, https://link.springer.com/article/10.1007/s11030-021-10217-3
12. Haddaway, N.R., Page, M.J., Pritchard, C.C., McGuinness, L.A.: Prisma2020: An r package and shiny app for producing prisma 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis. Campbell Systematic Reviews **18**, e1230 (6 2022). https://doi.org/10.1002/CL2.1230, https://onlinelibrary.wiley.com/doi/full/10.1002/cl2.1230https://onlinelibrary.wiley.com/doi/abs/10.1002/cl2.1230https://onlinelibrary.wiley.com/doi/10.1002/cl2.1230
13. Han, L., Zhang, X., Fu, W.Q., Sun, C.Y., Zhao, X.M., Zhou, L.R., Liu, G.X.: A systematic review of the budget impact analyses for antitumor drugs of lung cancer. Cost Effectiveness and Resource Allocation **18**, 1–10 (12 2020). https://doi.org/10.1186/S12962-020-00253-5/TABLES/3, https://resource-allocation.biomedcentral.com/articles/10.1186/s12962-020-00253-5
14. Hanson, G., Adams, J., Kepgang, D.I., Zondagh, L.S., Bueh, L.T., Asante, A., Shirolkar, S.A., Kisaakye, M., Bondarwad, H., Awe, O.I.: Machine learning and molecular docking prediction of potential inhibitors against dengue virus. Frontiers in Chemistry **12**, 1510029 (12 2024). https://doi.org/10.3389/FCHEM.2024.1510029/XML/NLM
15. Hooshmand, S.A., Ghobadi, M.Z., Hooshmand, S.E., Jamalkandi, S.A., Alavi, S.M., Masoudi-Nejad, A.: A multimodal deep learning-based drug repurposing approach for treatment of covid-19. Molecular Diversity **25**, 1717–1730 (8 2021). https://doi.org/10.1007/S11030-020-10144-9/FIGURES/8, https://link.springer.com/article/10.1007/s11030-020-10144-9

16. Huang, K., Chandak, P., Wang, Q., Havaldar, S., Vaid, A., Leskovec, J., Nadkarni, G.N., Glicksberg, B.S., Gehlenborg, N., Zitnik, M.: A foundation model for clinician-centered drug repurposing. Nature Medicine 2024 30:12 **30**, 3601–3613 (9 2024). https://doi.org/10.1038/s41591-024-03233-x, https://www.nature.com/articles/s41591-024-03233-x
17. Irwin, R., Dimitriadis, S., He, J., Bjerrum, E.J.: Chemformer: A pre-trained transformer for computational chemistry
18. Issa, N.T., Stathias, V., Schürer, S., Dakshanamurthy, S.: Machine and deep learning approaches for cancer drug repurposing. Seminars in Cancer Biology **68**, 132–142 (1 2021). https://doi.org/10.1016/J.SEMCANCER.2019.12.011
19. Jukič, M., Bren, U.: Machine learning in antibacterial drug design. Frontiers in Pharmacology **13**, 864412 (5 2022). https://doi.org/10.3389/FPHAR.2022.864412, https://pmc.ncbi.nlm.nih.gov/articles/PMC9110924/
20. Kim, J., Greenberg, D.E., Pifer, R., Jiang, S., Xiao, G., Shelburne, S.A., Koh, A., Xie, Y., Zhan, X.: Vampr: Variant mapping and prediction of antibiotic resistance via explainable features and machine learning. PLoS computational biology **16** (2020). https://doi.org/10.1371/JOURNAL.PCBI.1007511, https://pubmed.ncbi.nlm.nih.gov/31929521/
21. Knox, C., Wilson, M., Klinger, C.M., Franklin, M., Oler, E., Wilson, A., Pon, A., Cox, J., Chin, N.E.L., Strawbridge, S.A., Garcia-Patino, M., Kruger, R., Sivakumaran, A., Sanford, S., Doshi, R., Khetarpal, N., Fatokun, O., Doucet, D., Zubkowski, A., Rayat, D.Y., Jackson, H., Harford, K., Anjum, A., Zakir, M., Wang, F., Tian, S., Lee, B., Liigand, J., Peters, H., Wang, R.Q.R., Nguyen, T., So, D., Sharp, M., da Silva, R., Gabriel, C., Scantlebury, J., Jasinski, M., Ackerman, D., Jewison, T., Sajed, T., Gautam, V., Wishart, D.S.: Drugbank 6.0: the drugbank knowledgebase for 2024. Nucleic Acids Research **52**, D1265–D1275 (1 2024). https://doi.org/10.1093/NAR/GKAD976,, https://pubmed.ncbi.nlm.nih.gov/37953279/
22. Li, B., Shao, H., Gao, L., Li, H., Sheng, H., Zhu, L.: Nano-drug co-delivery system of natural active ingredients and chemotherapy drugs for cancer treatment: a review. Drug Delivery **29**, 2130–2161 (12 2022). https://doi.org/10.1080/10717544.2022.2094498, https://www.tandfonline.com/doi/abs/10.1080/10717544.2022.2094498
23. Liu, G.Y., Yu, D., Fan, M.M., Zhang, X., Jin, Z.Y., Tang, C., Liu, X.F.: Antimicrobial resistance crisis: could artificial intelligence be the solution? Military Medical Research 2024 11:1 **11**, 1–23 (1 2024). https://doi.org/10.1186/S40779-024-00510-1, https://mmrjournal.biomedcentral.com/articles/10.1186/s40779-024-00510-1http://creativecom-mons.org/publicdomain/zero/1.0/
24. Mattioli, R., Ilari, A., Colotti, B., Mosca, L., Fazi, F., Colotti, G.: Doxorubicin and other anthracyclines in cancers: Activity, chemoresistance and its overcoming. Molecular Aspects of Medicine **93** (10 2023). https://doi.org/10.1016/j.mam.2023.101205, https://pubmed.ncbi.nlm.nih.gov/37515939/
25. Melo, M.C., Maasch, J.R., de la Fuente-Nunez, C.: Accelerating antibiotic discovery through artificial intelligence. Communications Biology 2021 4:1 **4**, 1–13 (9 2021). https://doi.org/10.1038/s42003-021-02586-0, https://www.nature.com/articles/s42003-021-02586-0
26. Mishra, S., Rastogi, Y.P., Jabin, S., Kaur, P., Amir, M., Khatun, S.: A deep learning ensemble for function prediction of hypothetical proteins from pathogenic bacterial species. Computational Biology and Chemistry **83**, 107147

(12 2019). https://doi.org/10.1016/J.COMPBIOLCHEM.2019.107147, https://www.sciencedirect.com/science/article/pii/S1476927119305936?casa_token=fp3yJdop9YkAAAAA:kQWxx71knULPDh4uUNnlaZEIKy1Gdhru1erlp7qsa3pU0_742IarD5FgBF7reP9av-2EaSrFBOE

27. Pandey, D., Kumari, B., Singhal, N., Kumar, M.: Baceffluxpred: A two-tier system to predict and categorize bacterial efflux mediated antibiotic resistance proteins. Scientific reports **10** (12 2020). https://doi.org/10.1038/S41598-020-65981-3, https://pubmed.ncbi.nlm.nih.gov/32518231/
28. Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., Tekade, R.K.: Artificial intelligence in drug discovery and development. Drug Discovery Today **26**, 80–93 (1 2021). https://doi.org/10.1016/J.DRUDIS.2020.10.010
29. Polykovskiy, D., Zhebrak, A., Sanchez-Lengeling, B., Golovanov, S., Tatanov, O., Belyaev, S., Kurbanov, R., Artamonov, A., Aladinskiy, V., Veselov, M., Kadurin, A., Johansson, S., Chen, H., Nikolenko, S., Aspuru-Guzik, A., Zhavoronkov, A.: Molecular sets (moses): A benchmarking platform for molecular generation models. Frontiers in pharmacology **11** (12 2020). https://doi.org/10.3389/FPHAR.2020.565644, https://pubmed.ncbi.nlm.nih.gov/33390943/
30. Pourmadadi, M., Ghaemi, A., Shamsabadipour, A., Rajabzadeh-Khosroshahi, M., Shaghaghi, M., Rahdar, A., Pandey, S.: Nanoparticles loaded with daunorubicin as an advanced tool for cancer therapy. European Journal of Medicinal Chemistry **258** (10 2023). https://doi.org/10.1016/j.ejmech.2023.115547, https://pubmed.ncbi.nlm.nih.gov/37327678/
31. Rehman, A.U., Li, M., Wu, B., Ali, Y., Rasheed, S., Shaheen, S., Liu, X., Luo, R., Zhang, J.: Role of artificial intelligence in revolutionizing drug discovery. Fundamental Research (5 2024). https://doi.org/10.1016/J.FMRE.2024.04.021
32. Rodriguez, S., Hug, C., Todorov, P., Moret, N., Boswell, S.A., Evans, K., Zhou, G., Johnson, N.T., Hyman, B.T., Sorger, P.K., Albers, M.W., Sokolov, A.: Machine learning identifies candidates for drug repurposing in alzheimer's disease. Nature Communications 2021 12:1 **12**, 1–13 (2 2021). https://doi.org/10.1038/s41467-021-21330-0, https://www.nature.com/articles/s41467-021-21330-0
33. Sakho, A., Malherbe, E., Scornet, E.: Do we need rebalancing strategies? a theoretical and empirical study around smote and its variants (2 2024), https://arxiv.org/pdf/2402.03819
34. Scilipoti, P., Ślusarczyk, A., de Angelis, M., Soria, F., Pradere, B., Krajewski, W., D'Andrea, D., Mari, A., Giudice, F.D., Pichler, R., Subiela, J.D., Afferi, L., Albisinni, S., Mertens, L., Laukhtina, E., Mori, K., Radziszewski, P., Shariat, S.F., Necchi, A., Xylinas, E., Gontero, P., Rouprêt, M., Montorsi, F., Briganti, A., Moschini, M.: The role of mitomycin c in intermediate-risk non–muscle-invasive bladder cancer: A systematic review and meta-analysis. European Urology Oncology **7**, 1293–1302 (12 2024). https://doi.org/10.1016/J.EUO.2024.06.005, https://www.sciencedirect.com/science/article/pii/S2588931124001524?casa_token=Hw7cj8bTfNIAAAAA:BMDHpnhiTpp1R8g_fgNdB1d0O5hoIfF4I8cwG74NbfzE5rPAybvZuwFQ0MyyN2IW7Am7_OQHktI
35. Singh, A.: Artificial intelligence for drug repurposing against infectious diseases. Artificial Intelligence Chemistry **2**, 100071 (12 2024). https://doi.org/10.1016/J.AICHEM.2024.100071
36. Son, D.S., Lee, E.S., Adunyah, S.E.: The antitumor potentials of benzimidazole anthelmintics as repurposing drugs. Immune Network **20**, 1–20 (7 2020). https://doi.org/10.4110/IN.2020.20.E29, https://doi.org/10.4110/in.2020.20.e29

37. Stokes, J.M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N.M., MacNair, C.R., French, S., Carfrae, L.A., Bloom-Ackerman, Z., Tran, V.M., Chiappino-Pepe, A., Badran, A.H., Andrews, I.W., Chory, E.J., Church, G.M., Brown, E.D., Jaakkola, T.S., Barzilay, R., Collins, J.J.: A deep learning approach to antibiotic discovery. Cell **180**, 688–702.e13 (2 2020). https://doi.org/10.1016/J.CELL.2020.01.021/ATTACHMENT/012CE4DE-FC28-489F-95C0-CA386FECBFBF/MMC7.XLSX, https://www.cell.com/action/showFullText?pii=S0092867420301021https://www.cell.com/action/showAbstract?pii=S0092867420301021https://www.cell.com/cell/abstract/S0092-8674(20)30102-1
38. Tang, K.W.K., Millar, B.C., Moore, J.E.: Antimicrobial resistance (amr). British Journal of Biomedical Science **80**, 11387 (6 2023). https://doi.org/10.3389/BJBS.2023.11387/BIBTEX, https://pubmed.ncbi.nlm.nih.gov/37448857/
39. Tanoli, Z., Vähä-Koskela, M., Aittokallio, T.: Artificial intelligence, machine learning, and drug repurposing in cancer. Expert opinion on drug discovery **16**, 977–989 (2021). https://doi.org/10.1080/17460441.2021.1883585, https://pubmed.ncbi.nlm.nih.gov/33543671/
40. Walsh, T.R., Gales, A.C., Laxminarayan, R., Dodd, P.C.: Antimicrobial resistance: Addressing a global threat to humanity. PLOS Medicine **20**, e1004264 (7 2023). https://doi.org/10.1371/JOURNAL.PMED.1004264, https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1004264
41. Wan, F., Torres, M.D., Peng, J., de la Fuente-Nunez, C.: Deep-learning-enabled antibiotic discovery through molecular de-extinction. Nature Biomedical Engineering 2024 8:7 **8**, 854–871 (6 2024). https://doi.org/10.1038/s41551-024-01201-x, https://www.nature.com/articles/s41551-024-01201-x
42. Wang, Y., Aldahdooh, J., Hu, Y., Yang, H., Vähä-Koskela, M., Tang, J., Tanoli, Z.: Drugrepo: a novel approach to repurposing drugs based on chemical and genomic features. Scientific Reports 2022 12:1 **12**, 1–13 (12 2022). https://doi.org/10.1038/s41598-022-24980-2, https://www.nature.com/articles/s41598-022-24980-2
43. Wen, Q., Liu, R., Zhang, P.: Clinical connectivity map for drug repurposing: using laboratory results to bridge drugs and diseases. BMC medical informatics and decision making **21** (9 2021). https://doi.org/10.1186/S12911-021-01617-4, https://pubmed.ncbi.nlm.nih.gov/34560862/
44. World Health Organization: 10 global health issues to track in 2021. https://www.who.int/news-room/spotlight/10-global-health-issues-to-track-in-2021 (2021), accessed: 2025-06-20
45. Xie, Y., Shi, C., Zhou, H., Yang, Y., Zhang, W., Yu, Y., Li, L.: Mars: Markov molecular sampling for multi-objective drug discovery https://github.com/yutxie/mars.
46. Xu, J., Meng, L.H., Qing, C.: The clinical application and development of traditional antitumor drugs. Yao Xue Xue Bao **56**, 1551–1561 (6 2021). https://doi.org/10.16438/J.0513-4870.2020-1895, http://dx.doi.org/10.16438/j.0513-4870.2020-1895
47. Yang, F., Zhang, Q., Ji, X., Zhang, Y., Li, W., Peng, S., Xue, F.: Machine learning applications in drug repurposing. Interdisciplinary Sciences – Computational Life Sciences **14**, 15–21 (3 2022). https://doi.org/10.1007/S12539-021-00487-8/TABLES/1, https://link.springer.com/article/10.1007/s12539-021-00487-8
48. Yang, X., Yang, G., Chu, J.: Self-supervised learning for label sparsity in computational drug repositioning. IEEE/ACM transactions on computational biol-

ogy and bioinformatics **20**, 3245–3256 (9 2023). https://doi.org/10.1109/TCBB.2023.3254163, https://pubmed.ncbi.nlm.nih.gov/37028367/

49. Zhang, P., Wang, X., Cen, X., Zhang, Q., Fu, Y., Mei, Y., Wang, X., Wang, R., Wang, J., Ouyang, H., Liang, T., Xia, H., Han, X., Guo, G.: A deep learning framework for in silico screening of anticancer drugs at the single-cell level. National Science Review **12** (1 2025). https://doi.org/10.1093/NSR/NWAE451, https://dx.doi.org/10.1093/nsr/nwae451