UNIVERSITAT DE BARCELONA

Facultat de Farmàcia
i Ciències de l'Alimentació

Final Project

Degree in Pharmacy

ARTIFICIAL INTELLIGENCE IN THE BATTLE AGAINST SUPERBUGS

SOPHIE ANGULO BROHS

Main field: Biochemistry

Secondary fields:

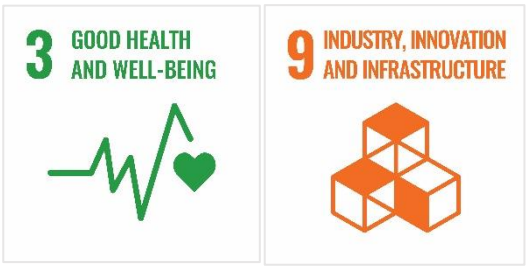
Microbiology, Pharmacology & Therapeutics and Pharmaceutical Chemistry

Department of Biochemistry and Molecular Biology

Faculty of Pharmacy and Food Sciences

University of Barcelona

Barcelona, July 4th 2024



https://www.un.org/sustainabledevelopment/news/communications-material/

Acknowledgements

I would like to express my sincere gratitude to all those who have contributed to the experience throughout these past five years of college at the Faculty of Pharmacy at the University of Barcelona. I would like to thank the Department of Biochemistry and Physiology for accepting my topic proposal.

I am particularly grateful to my tutor for the guidance and insightful suggestions during the writing process of this paper. I wish to extend my thanks to my friends for their support and ability to lighten my mood during challenging times, particularly when writing the paper seemed to stagnate.

Special thanks are due to Dr. DI Friedrich Brohs, my uncle, for his enthusiasm and engagement with the topic, which allowed me to discuss and improve my writing as well as my comprehension of computer science.

Finally, I want to acknowledge the love, support, and encouragement from my parents and grandmother, which have been invaluable.

I dedicate this work to the memory of those who are no longer with us but whose influence has profoundly shaped my life and pursuits.

Abbreviations

ABC: ATP-binding cassette

AI: Artificial intelligence

AMR: Antimicrobial resistance

CFU: Colony-forming units

CRE: *Carbapenem-resistant Enterobacteriaceae*

DL: Deep learning

DOX: Doxycycline

EMA: European Medicines Agency

ESKAPE: *Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa, and Enterobacter spp.*

FDA: Food and Drug Administration

ICU: Intensive Care Unit

IV: Intravenous

JNK: c-Jun N-terminal kinase

LPS: Lipopolysaccharides

MAPKs: Mitogen-activated protein kinases

MBC: Minimum bactericidal concentration

MDR: Multidrug resistant

MIC: Minimum Inhibitory Concentration

MIT: Massachusetts Institute of Technology

ML: Machine learning

ROS: Reactive oxygen species

R&D&I: Research and development and innovation

SEM: Scanning electron microscopy

SMILES: Simplified Molecular Input Line Entry System

TEM: Transmission electron microscopy

WHO: World Health Organization

Abstract

The rise of multidrug-resistant superbugs underscores the urgent need for novel antibiotics. The breakthrough discovery of three significant antibiotics - Halicin, Abaucin, and Zosurabalpin - was achieved through AI algorithms, showcasing its potential to revolutionize antibiotic research and development and innovation (R&D&I) by repurposing these molecules. Models like Naive Bayes Classification, Decision Trees, Random Forests, Support Vector Machines, and Artificial Neural Networks analyze chemical and biological data to predict and optimize potential antibiotics, highlighting the importance of data quality and standardization for accurate predictions.

Halicin (SU3327), which disrupts bacterial membrane integrity, showed bacteriostatic effects against a range of bacteria including *Mycobacterium tuberculosis* and *carbapenem-resistant Enterobacteriaceae*. Abaucin (RS102895), predicted by machine learning to combat *Acinetobacter baumannii* infections, interferes with lipopolysaccharide transport in bacterial cell membranes and reduces inflammation by antagonizing CCR2-selective chemokine receptors. Finally, the recent discovery of Zosurabalpin (RG6006), which acts against *carbapenem-resistant A. baumannii* by a unique mechanism of action, emphasizes the need to diversify antibiotic targets.

While AI presents promising opportunities in accelerating antibiotic discovery, challenges such as bias mitigation and standardization imperatives necessitate attention. This work sheds light on the utilization of AI in revolutionizing drug development, but it also underscores the role of data privacy and regulatory frameworks in shaping the ethical landscape of AI applications in healthcare.

Keywords

Artificial intelligence, superbugs, antimicrobial resistance, machine learning

Resumen

El aumento de las superbacterias resistentes a múltiples fármacos pone de manifiesto la urgente necesidad de nuevos antibióticos. El descubrimiento revolucionario de tres antibióticos - Halicina, Abaucina y Zosurabalpina - se logró mediante algoritmos de inteligencia artificial (IA), demostrando que esta revoluciona la investigación, el desarrollo y la innovación (I+D+I) de antibióticos.

La Halicina (SU3327), altera el gradiente electroquímico de la membrana bacteriana y, por consiguiente, su integridad. Mostró efectos bacteriostáticos contra una variedad de bacterias, incluyendo *Mycobacterium tuberculosis* y *Enterobacteriaceae resistentes a carbapenems*.

La Abaucina (RS102895), combate infecciones de *Acinetobacter baumannii* al interferir con el transporte de lipopolisacáridos de las membranas celulares bacterianas. Además, reduce la inflamación al antagonizar los receptores de quimiocinas selectivos de CCR2.

La Zosurabalpina (RG6006), actúa contra *A. baumannii resistente a carbapenems*, mediante un mecanismo de acción único, enfatizando la necesidad de diversificar las dianas de los antibióticos.

Modelos como clasificación Naive Bayes, los árboles de decisión, los bosques aleatorios, las máquinas de vectores de soporte y las redes neuronales artificiales, son empleados para analizan datos químicos y biológicos para poder predecir y optimizar posibles antibióticos, destacando la importancia de la calidad y estandarización de los datos para predicciones precisas.

A pesar de estos beneficios de la IA, es necesario tener en consideración conceptos como la mitigación de sesgos y los imperativos de estandarización. La finalidad de este trabajo es presentar esta nueva tendencia de la utilización de la IA para revolucionar el I+D+I de fármacos en un ámbito en el que la normativa legal y aplicación de IA debe ajustarse a la protección de datos y el marco ético del sector sanitario.

Palabras clave

Inteligencia artificial, superbacterias, resistencia a los antibióticos, aprendizaje automático

Integration of the different scopes

This article represents a fusion of diverse scientific disciplines aimed at gaining deeper insights into the latest trends in utilizing artificial intelligence (AI) within pharmaceutical R&D&I. While the primary focus lies within the field of biochemistry, the paper focuses to shed light on the mechanisms of action of AI-generated antibiotics, intended for clinical application in treating multidrug-resistant (MDR) infections. To explore these processes comprehensively, biochemists collaborate closely with microbiologists, forming highly skilled interdisciplinary research teams. Moreover, pharmaceutical chemists undertake the crucial task of designing optimal synthesis pathways to yield these novel biologically active molecules. Notably, computer scientists play a pivotal role in developing algorithms and filters for these programs. Each expert brings its unique perspective and expertise to the collaborative effort.

Leveraging advanced computer science, such as AI, as a tool for designing and predicting molecular properties like binding affinity or activity, facilitates expedited and enhanced R&D&I processes.

It is paramount to underscore that regulatory bodies will only authorize clinical trials for these innovative active molecules if researchers possess a thorough understanding of their mechanism of action and the biochemical reactions they catalyze or inhibit.

Identification and reflection on the Sustainable Development Goals (SDGs)

The World Health Organization (WHO) has devised the "17 Sustainable Development Goals" to provide a coordinated response to the challenges confronting humanity. As aspiring pharmacists, we are committed to ensure a prosperous future for the communities, species, and ecosystems from a pharmaceutical point of view. This article is part of my contribution to this matter.

The third goal (outlined by the WHO) emphasizes health and well-being (1). The proliferation of MDR microorganisms poses a significant threat globally, particularly in low- and middle-income countries with fragile healthcare systems and limited resources. Using AI for the discovery of novel antibiotics directly contributes to the achievement of goal 3.b. By developing remedies for communicable infections caused by MDR microorganisms, we move closer to realizing the objective 3.8 of providing "access to quality essential health-care services and access to safe, effective, quality, and affordable essential medicines for all". Progress can be monitored by examining indicators such as the "total net assistance to medical research and basic health sectors" or the availability and affordability of a "core set of relevant essential medicines" in health facilities on a sustainable basis.

Beyond the health and economic advantages of uncovering new antibiotics to combat MDR, there has been a noticeable surge in the promotion of green pharmacy practices (goal 12.4) (2,3). The WHO and various governments advocate for pharmaceutical companies to analyze the life cycle of antibiotics while carefully considering their mechanisms of action. Alongside efforts to optimize management of antibiotic use and waste, the development of more specific and narrow-spectrum antibiotics, as discussed in this paper, holds promise for significantly reducing the environmental impact of antibiotic usage. This approach represents a crucial step towards mitigating the concerning trend of antibiotic resistance.

INDEX

1. Introduction

What would our daily lives look like without antibiotics? We would most likely still be perishing form a minor cold or a (badly) infected wound. However, this alone should not give us a reason to celebrate victory yet. Recent reports from agencies like the European Medicines Agency (EMA) and the Food and Drug Administration (FDA) paint a concerning picture of the escalating emergence of resistant microorganisms. Among these, the most menacing are the so-called superbugs or MDR pathogens. Some experts go as far as labeling them "extreme drug-resistant" or "pan-drug-resistant." To earn the title of superbug, a microorganism must defy more than one agent in three or more antimicrobial categories as outlined by clinical guidelines (4). Within this classification, the ESKAPE superbugs stand out as the most prevalent and menacing nosocomial infections. ESKAPE comprises *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter spp*.

Despite the urgent demand for new treatments effective against MDR pathogens, the pharmaceutical industry shows little inclination to support and invest in R&D&I in this area. The journey from discovery to commercialization averages between $161 million to $4.54 US billion, not to mention the considerable time investment required for this process (5). This sector represents a high-risk, low-revenue endeavor, as the risk of becoming an obsolete treatment due to new resistance mechanisms before it becomes profitable is deemed too great. According to the WHO, unless the pharmaceutical industry commits to research into active antibacterial compounds and the irrational or misuse of antibiotics is regulated, the trajectory suggests that by 2050 more lives will be claimed by untreatable infections than cancer with estimates projecting up to ten million deaths annually.

The emergence of resistant microorganisms is not a novel occurrence. Even during the 20th century, hailed as the golden age of antibiotic discovery, scientists encountered mutated strains of bacteria. Some of these random mutations proved advantageous to the bacteria and a great threat to humans, as it enables bacteria to circumvent the effects of antibiotics shortly after their commercialization or discovery.

Bacteria employ various strategies to enhance mutation rates in pursuit of increased survival chances. One key method involves the integration of mobile genetic elements, which predominantly occurs through three mechanisms: transformation, conjugation, and transduction. Transformation entails bacteria incorporating DNA fragments from the environment, while conjugation involves the active exchange of genetic material between two bacteria via plasmids. Another less common path for acquiring resistance genes is through phages injecting new DNA fragments (6).

Repeated exposure to antibiotics in the environment - originating from factors like self-medication, inadequate medical supervision, incomplete treatment courses, overprescription, poor hygiene practices, or excessive use in livestock - boosts mutation rates and, consequently, resistance probabilities. This phenomenon not only facilitates the

release of DNA from deceased cells but also ensures the survival of the fittest bacteria, those resistant to antibiotics.

Moreover, considering the various mutation mechanisms, a logical classification method is based on the processes they affect, such as spatial exclusion, drug modification, target modification, or bypass (7). Mutations may enhance membrane permeability to prevent the antibiotics to enter the eukaryotic cells. Alternatively, bacteria may develop efflux pumps to actively eject antibiotics and prevent their antibacterial action.

Regarding drug modification, mutated enzymes are pivotal. These enzymes may alter antibiotic molecules by adding functional groups, hindering their binding to target sites, or simply deactivating the drug. In a contrary way, bacteria may modify the binding sites to impede drug interaction. Additionally, an increase in the expression level of enzymes, according to the Michaelis-Menten law, can lead to more free enzymes available to counteract inhibitory drug molecules, thereby bypassing inhibition. In either scenario, a beneficial mutation for the bacteria inevitably comes at the expense of human health, diminishing available treatment options.

Yet another mechanism involves bacteria acquiring substitute pathways for essential reactions interrupted by medication. Despite drug inhibition, the cell responds as if the blockade were absent, enabling the continuation of vital processes.

Given the intricate range of resistance mechanisms, researchers are actively seeking innovative tools and support systems to meet the escalating demands of healthcare for potent, cost-effective antibiotics, ideally with straightforward dosing protocols. Employing advanced computer science techniques, such as AI, to model and predict molecular characteristics such as binding affinity or activity, streamlines the R&D&I process. Computer science is just beginning to explore the countless possibilities for aiding scientists in the preclinical stages of R&D&I, potentially freeing up resources to focus on critical processes like clinical trials (8,9).

Before looking deeper into this issue, it is essential to clarify the term AI, which refers to computer systems that emulate certain human brain functions, such as interpreting information, problem-solving, and learning from input data (10). AI has subcategories as shown in **Figure 1**. One is known as machine learning (ML), which takes a step further by analyzing data without explicit human programming, thereby enhancing task performance through the analysis of new data. Another one, even more sophisticated is deep learning (DL) the most sophisticated form of machine learning, employing multiple layers of data processing to achieve results (11).

ML normally utilizes categorized data to find patterns and predict information. It can also be supervised, where human interaction corrects algorithms and adjusts the output and unsupervised, where the machine can reiterate and self-adjust its processing. To implement ML a smaller amount of structured data is sufficient, as there is mostly some human interaction.

Deep learning is suitable for vast amounts of unstructured data. Like the human brain, which it is supposed to resemble, it combines a multitude of nodes (resembling neurons) organized in layers. They connect to their neighbors and through statistically weighed functions perform their task. Neural network nodes accumulate data and if a certain threshold is surpassed, the transfer of information to the next node in the next layer is triggered. Neural networks are capable of solving complex problems autonomously, with this however comes the disadvantage, that it is not always possible to know how the solution was obtained. Normally neural networks connect data forward, from the lower to the higher layer. It is possible to implement a backward data propagation. This allows to calculate errors nodes generated.

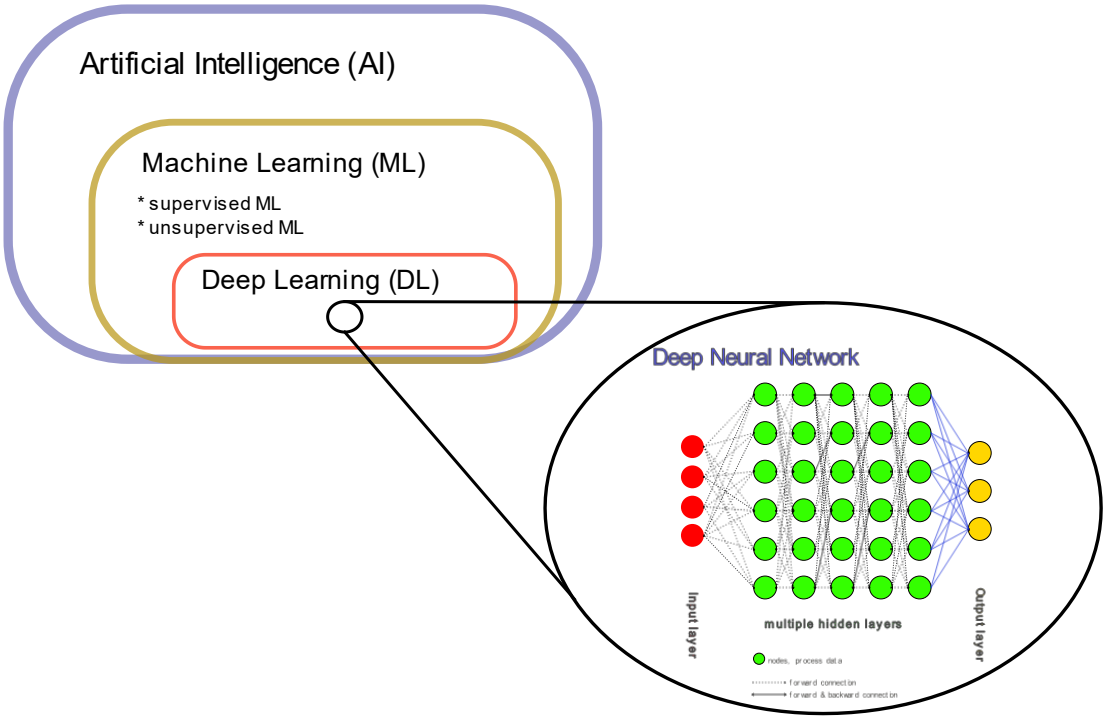


Figure 1: AI and its subcategories

2. Objectives

The objective of this article is not to cover the entirety of AI's current role in the intricate process of new drug development, but furthermore, to gather information on three active substances Halicin, Abaucin and Zosurabalpin discovered by AI. This exploration seeks to highlight the significant potential of neural network programs in R&D&I for novel antibiotics combating MDR microorganisms.

3. Materials and methods

This final degree project presents a comprehensive bibliographic analysis of scientific publications sourced from renowned international databases including PubMed, Nature, and Elsevier, accessible through the University of Barcelona. Employing Mesh Terms and keywords, the search was conducted to gather pertinent references spanning from August 2023 to April 2024. Google Scholar proved instrumental in identifying peer-reviewed documents. To optimize access to information and enhance efficiency, bibliographic references were meticulously reviewed and summarized. Furthermore, the platform Zotero facilitated the organization of cross-references and citations of the sourced materials.

Throughout the data collection process, emphasis was placed not only on exploring the technological applications of AI in biochemistry, but also on selecting articles, and monographs based on publication dates. Priority was given to documents published within the past six years to ensure currency of information, particularly concerning the status of clinical trials. To this end, the official US website ClinicalTrials.gov was consulted. Given the project's focus on the contemporary relevance of AI in pharmaceutical research and its potential future implications, reliance on the latest available information was imperative.

4. Results and Discussion

4.1. Identification of three antibiotics by AI

The identification of three noteworthy molecules - Halicin, Abaucin, and Zosurabalpin - in a chronological sequence, was conducted by distinct AI algorithms. Initially part of various databases for diverse indications or diseases, these compounds were singled out by neural networks for repurposing as antibiotics (12,13). A common thread among all three prediction processes was the pre-analysis filtration of input data. Molecules bearing resemblance to known antibiotics were deliberately excluded to avoid the possibility of pre-existing cross-resistances.

Halicin (SU3327) (**Figure 2**), belonging to the thiadiazoles family, has exhibited in-vivo antibacterial efficacy against *Mycobacterium tuberculosis*, *carbapenem-resistant Enterobacteriaceae* (CRE), *Clostridium difficile*, *pan-resistant A. baumannii* (14), *E. faecium* and *Enterococcus faecalis* (15). Regrettably, its effectiveness against *P. aeruginosa* was compromised, likely because of inadequate cell permeability. The minimum inhibitory concentration (MIC) varied across studies and targeted strains, spanning from 2 to 8µg/mL (14,15). In a particular investigation, Hussain et al. (15) examined the toxicity in combination with doxycycline (DOX) due to potential synergistic effects. Both in-vivo and in-vitro evaluations revealed no significant cytotoxic properties. A reduced dosage exhibited bacteriostatic effects on *E. faecalis*, alongside a modest anti-inflammatory response in the vicinity of infected wounds in laboratory mice.

Figure 2: 2D Chemical Structure Depiction of Halicin (16).

Moving forward, the next drug under examination is Abaucin (RS102895), identified as a benzoxazine compound (**Figure 3**). Its efficacy against *A. baumannii* infections was predicted by an ML program. With a MIC of 2 µg/mL, Abaucin demonstrated optimal effectiveness during the growth and division phases of the microorganisms. The distinctive structural and functional attributes of Abaucin pave the way for innovative, targeted treatment options against *A. baumannii*, offering a narrowed spectrum of activity. Notably, it exhibited no activity in nutrient-depleted PBS. Additionally, in-vivo experiments conducted on mice revealed that alternative antibiotic treatments showed quicker wound healing when the infection was not caused by resistant microorganisms (17).

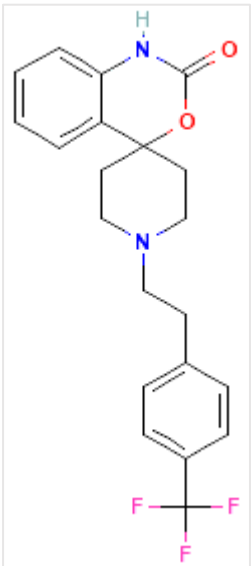


Figure 3: 2D Chemical Structure Depiction of Abaucin (18).

Finally, the most recent addition to the arsenal of fully synthetic antibiotics, discovered in 2024 through ML, is Zosurabalpin (RG6006) (19,20). This narrow-spectrum tethered macrocyclic peptide exhibits selective activity against *carbapenem-resistant A. baumannii* (**Figure 4**). Two phase I clinical trials have already been concluded, with one study concentrating on Zosurabalpin's pharmacokinetic properties following a single intravenous (IV) dose administered to Intensive Care Unit (ICU) patients with bacterial infections (21) . The second study encompasses evaluations of pharmacokinetics, safety,

and tolerability after IV administration in healthy volunteers (22). Despite its favorable profile, the rapid clearance and short half-life of Zosurabalpin may complicate its clinical utilization.

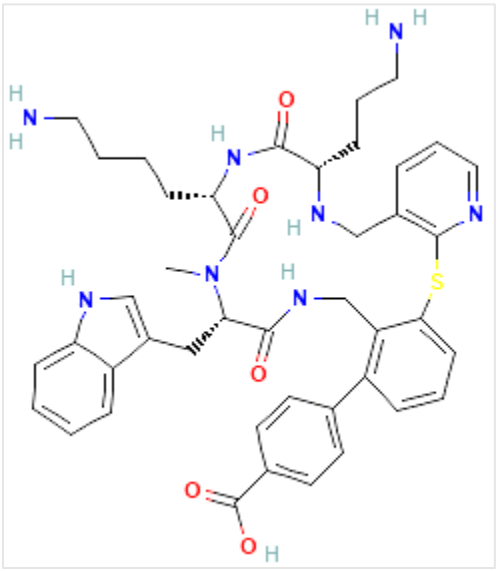


Figure 4: 2D Chemical Structure Depiction of Zosurabalpin (23).

4.2. AI models for R&D&I in the field of AMR

The term AI encompasses a wide range of models and algorithms, each markedly distinct from the others. Not every AI program is equally suitable or sensitive for all applications. AI can significantly support and enhance the work of health professionals across various fields, as seen in **Figure 5**. For instance, computer programmers can create predictive models to foresee the emergence of new resistances or develop monitoring systems that encourage the rational use of antibiotics (23,24).

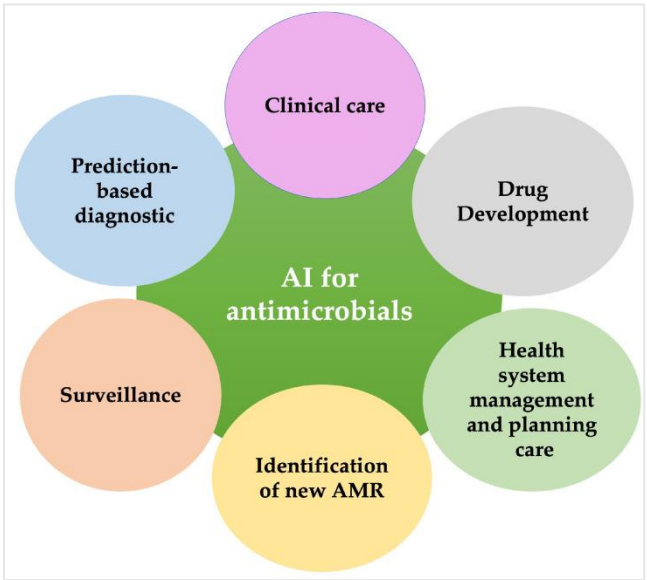


Figure 5: Practical Application of AI in the Antimicrobial Sector (24).

One handicap in research that ML has notably mitigated is the challenge of identifying antimicrobial resistance (AMR). By swiftly and accurately screening the MDR database, ML surpasses the efficiency of traditional in-vivo experiments, yielding clearer results. Additionally, ML-driven in-silico data processing significantly reduces false negatives in AMR labeling. The algorithm's learning capacity enables it to recognize previously undetected resistance sequences, which lacked apparent similarities, thus correctly flagging them as AMR (24,25).

As previously mentioned, ML and DL models belong to the advanced category of neural AI programs. These models excel at identifying underlying connections and patterns in data, allowing them to interpret, and provide insights and possible outcomes. The most exciting aspect of this field lies in three specific types of DL models, which highlight the immense power these programs currently possess and will continue to develop. Their effectiveness is measured by the accuracy of their outcomes relative to the quantity and quality of data provided to the algorithm.

The three types of DL models, each trained for regression and prediction tasks, varying in complexity and approach: unsupervised learning, reinforcement learning, and supervised learning (24). Unsupervised learning models group similar compounds based on selected features. Reinforcement learning models evaluate data by associating it with positive or negative outcomes through a “penalty and reward” system. These models are particularly effective for control and operational tasks. Supervised learning models are the most complex. They involve a workflow where the target outcomes are predefined. Then, the algorithm tries to establish non-linear connections hidden within the dataset to achieve accurate predictions.

This article will focus exclusively on the supervised learning of algorithms, since they are the most commonly used in R&D&I.

Firstly, Naïve Bayes Classification is an algorithm that bases its predictions purely on statistical analysis and assumes, that the probability of the outcome is a consequence of two antecedents, features used for classification that are independent (26). Secondly, the Decision Tree is utilized for predicting the characteristics of molecules. A more complex version of this model is the Random Forest, which selects samples of molecules chosen randomly by another decision tree within the same system. Alternatively, researchers can apply a Support Vector Machine as a classification tool. The algorithm classifies data according to different mathematical principles and then makes binary decisions. Lastly, the most complex and extensive program is the Artificial Neural Network, developed to emulate neural activity in order to analyze information and provide a weighted response or "added value" to the data (25). **Figure 6** is intended to help visualize these complex concepts.

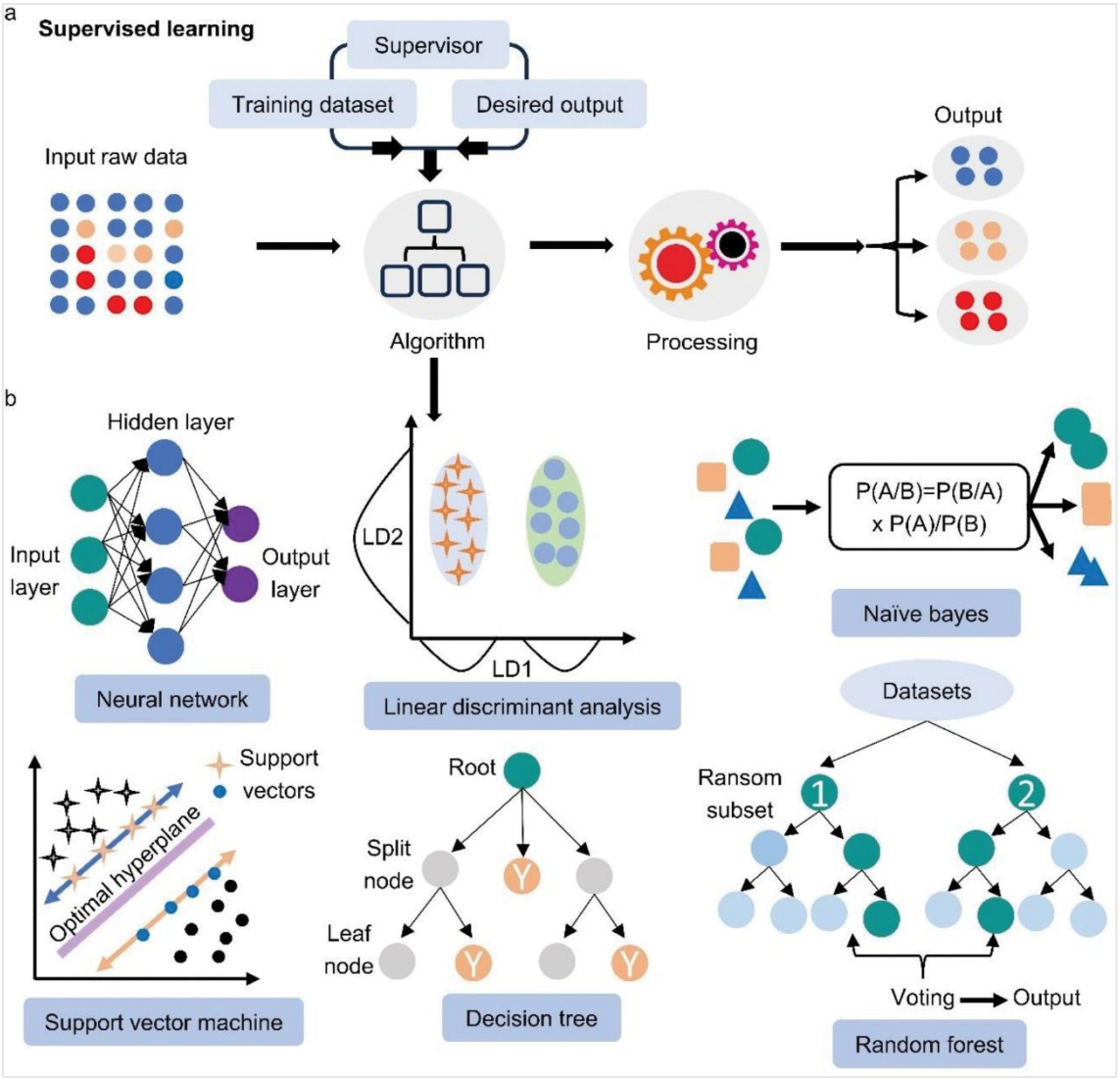


Figure 6: Supervised learning. a) Basic architecture of supervised learning. b) Examples present the commonly used supervised algorithms for NP discovery: neural networks, linear discriminant analysis, naive Bayes, support vector machine, decision tree, and random forest (27).

While some "basic" AI programs struggle with vast amounts of data, DL models thrive on extensive datasets - the more data they have, the better they perform. Without sufficient data, it is challenging, if not impossible, for the algorithm to determine the associative impacts of the known features and generate reliable and accurate outcomes. AI models autonomously adjust based on the data they process, eliminating the need for manual intervention, unlike traditional computer algorithms, which require user adjustments to obtain useful results.

But how do these algorithms perform such comprehensive analyses? The innovation lies in AI, particularly ML's capability to represent novel molecules as vectors. In programming, a vector is a dynamic one-dimensional collection of similar data elements. Essentially, it stores data (28,29). The network gathers information by sending multiple "messages" from one atom to its neighboring atoms, archiving all information and

insights into a vector. This vector comprises comprehensive details about the entire molecule, from atom features such as atomic number, bond count, formal charge, chirality, bonded hydrogen count, hybridization, aromaticity, and atomic mass, to bond features like bond type, conjugation, ring membership, and stereochemistry (**Figure 7**).

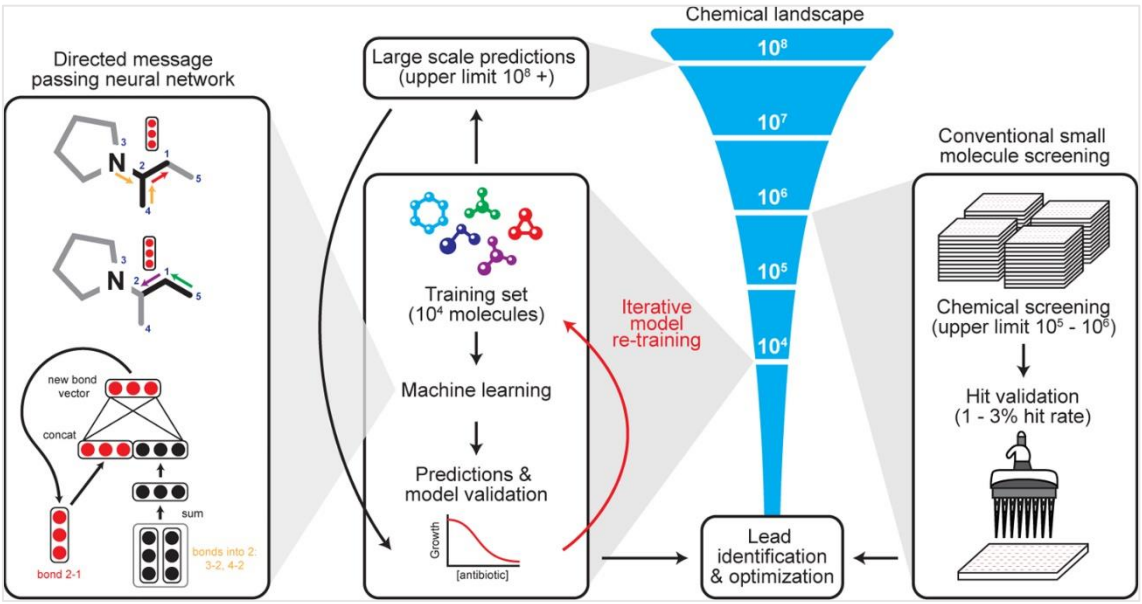


Figure 7: Machine learning in antibiotic discovery. A neural network model workings by building a molecular graph based on a specific property, in this case the inhibition of the growth of *E. coli*, using a directed message passing approach. After ranking the candidates according to the model's predicted score, a list of promising candidates was selected (14).

Prior to conducting this in-silico analysis, it is also important to define the objectives of the analysis to determine the optimal setup and select the pertinent data for input. These algorithms represent intricate systems perfected to explore and analyze chemical landscapes, with the aim of forecasting their antibacterial efficacy. The training process involves the input of diverse data collected under controlled conditions to ensure a broad prediction capacity. Emphasis is placed on establishing a foundation with minimal structural resemblance to already existing antibiotics, hence molecules with similar chemical structure were excluded beforehand.

4.3. Data

The most important base for AI is a significant amount of data. The main difference between AI and regular development is the approach. Normally using databases and filters, the programmer would change the algorithm "manually" to fit the requirements for the results. For example, when converting temperature, the programmer would use values and fit the algorithm (in this case linear regression), which in the end would lead to the well-known equation:

$$℃ = \frac{5}{9}(℉ - 32)$$

For AI the input would be two datasets, one input array for Celsius = [-40, -10, 0, 8, 15, 22, 38], and the output array would be the Fahrenheit array = [-40, 14, 32, 46, 59, 72, 100]. After training and learning, the model would then be able to make predictions e.g. using the code: print(predict(100C)). The output should be 212F, which is correct, but the user would not know how the model came to generate the result (30). This is why the model needs to be validated. This can be done by comparing the output of data obtained from an experiment and examining if the results align with the simulated training scenario. This way the researchers can establish that the program's "behavior" of all its executions generate accurate results. Additionally, it is recommended to also conduct a conceptual validation, meaning it is necessary to question if the hypothesis and the assumptions to be concluded from this model are justifiable (31).

The chosen example is a very simple model, that represents only 1 variable. In order to perform more complex predictions, data needs to be transformed in a machine-readable format. This can be done by vectorization. A vector is normally a linear array of n numbers:

$$\vec{v} = (x_1, x_2, x_3, \dots, x_n)$$

The data "stored" in the vector can be any significant parameters a molecule has to be characterized: atoms, bonds, rings, partial charges, and many more. As for the classification, one parameter is either active against a certain organism or not - a simple binary decision.

In order to make all this information understandable, a simple example such as Messi kicking a soccer ball will make these programming concepts more visual.
The input variables could be:

- force = 70N
- horizontal angle = 30 degree off center
- vertical angle = 15 degree below center
- speed of the ball before kick in relation to Messi = 0 (penalty)
- ball weight, inflation pressure, etc.

This could be transformed in a vector that would create an immense quantity of mathematical operations for each number representing a variable.

$$\underset{Ball}{\longrightarrow} = (70, 30, -15, 0, ball\ weight, ball\ inflation\ pressure, \dots\dots\dots\dots x_n)$$

With this idea in mind, the same concept can be applied for pharmaceutical applications. There have been many algorithms published, containing collection of more than hundred molecular descriptors that convert specific physical and chemical characteristics into vectors. These databases are partially available on the internet for free and can be combined. The basic for many of those modules is SMILES "Simplified Molecular Input Line Entry System", a way of creating linear letter combinations from 3D structural formulas that can be understood by software (32). Additionally, they can be augmented with supplementary information and the whole set of data can be transformed into a vector.

To illustrate the process here **Figure 8** - an example found online:

```
# import packages
from rdkit import Chem
from rdkit.ML.Descriptors.MoleculeDescriptors import MolecularDescriptorCalculator

# define SMILES string of molecule
smiles_string = "O=C(Cc1cncc2ccccc12)Nc1ccccc1"

# convert SMILES string to RDKit mol object
mol = Chem.MolFromSmiles(smiles_string)

# choose 200 molecular descriptors
chosen_descriptors = ['BalabanJ', 'BertzCT', 'Chi0', 'Chi0n', 'Chi0v', 'Chi1',
'Chi1n', 'Chi1v', 'Chi2n', 'Chi2v', 'Chi3n', 'Chi3v', 'Chi4n', 'Chi4v',
'EState_VSA1', 'EState_VSA10', 'EState_VSA11', 'EState_VSA2', 'EState_VSA3',
'EState_VSA4', 'EState_VSA5', 'EState_VSA6', 'EState_VSA7', 'EState_VSA8',
'EState_VSA9', 'ExactMolWt', 'FpDensityMorgan1', 'FpDensityMorgan2',
'FpDensityMorgan3', 'FractionCSP3', 'HallKierAlpha', 'HeavyAtomCount',
'HeavyAtomMolWt', 'Ipc', 'Kappa1', 'Kappa2', 'Kappa3', 'LabuteASA',
'MaxAbsEStateIndex', 'MaxAbsPartialCharge', 'MaxEStateIndex', 'MaxPartialCharge',
'MinAbsEStateIndex', 'MinAbsPartialCharge', 'MinEStateIndex', 'MinPartialCharge',
'MolLogP', 'MolMR', 'MolWt', 'NHOHCount', 'NOCount', 'NumAliphaticCarbocycles',
'NumAliphaticHeterocycles', 'NumAliphaticRings', 'NumAromaticCarbocycles',
'NumAromaticHeterocycles', 'NumAromaticRings', 'NumHAcceptors', 'NumHDonors',
'NumHeteroatoms', 'NumRadicalElectrons', 'NumRotatableBonds',
'NumSaturatedCarbocycles', 'NumSaturatedHeterocycles', 'NumSaturatedRings',
'NumValenceElectrons', 'PEOE_VSA1', 'PEOE_VSA10', 'PEOE_VSA11', 'PEOE_VSA12',
'PEOE_VSA13', 'PEOE_VSA14', 'PEOE_VSA2', 'PEOE_VSA3', 'PEOE_VSA4', 'PEOE_VSA5',
'PEOE_VSA6', 'PEOE_VSA7', 'PEOE_VSA8', 'PEOE_VSA9', 'RingCount', 'SMR_VSA1',
'SMR_VSA10', 'SMR_VSA2', 'SMR_VSA3', 'SMR_VSA4', 'SMR_VSA5', 'SMR_VSA6', 'SMR_VSA7',
'SMR_VSA8', 'SMR_VSA9', 'SlogP_VSA1', 'SlogP_VSA10', 'SlogP_VSA11', 'SlogP_VSA12',
'SlogP_VSA2', 'SlogP_VSA3', 'SlogP_VSA4', 'SlogP_VSA5', 'SlogP_VSA6', 'SlogP_VSA7',
'SlogP_VSA8', 'SlogP_VSA9', 'TPSA', 'VSA_EState1', 'VSA_EState10', 'VSA_EState2',
'VSA_EState3', 'VSA_EState4', 'VSA_EState5', 'VSA_EState6', 'VSA_EState7',
'VSA_EState8', 'VSA_EState9', 'fr_Al_COO', 'fr_Al_OH', 'fr_Al_OH_noTert', 'fr_ArN',
'fr_Ar_COO', 'fr_Ar_N', 'fr_Ar_NH', 'fr_Ar_OH', 'fr_COO', 'fr_COO2', 'fr_C_O',
'fr_C_O_noCOO', 'fr_C_S', 'fr_HOCCN', 'fr_Imine', 'fr_NH0', 'fr_NH1', 'fr_NH2',
'fr_N_O', 'fr_Ndealkylation1', 'fr_Ndealkylation2', 'fr_Nhpyrrole', 'fr_SH',
'fr_aldehyde', 'fr_alkyl_carbamate', 'fr_alkyl_halide', 'fr_allylic_oxid',
'fr_amide', 'fr_amidine', 'fr_aniline', 'fr_aryl_methyl', 'fr_azide', 'fr_azo',
'fr_barbitur', 'fr_benzene', 'fr_benzodiazepine', 'fr_bicyclic', 'fr_diazo',
'fr_dihydropyridine', 'fr_epoxide', 'fr_ester', 'fr_ether', 'fr_furan', 'fr_guanido',
'fr_halogen', 'fr_hdrzine', 'fr_hdrzone', 'fr_imidazole', 'fr_imide', 'fr_isocyan',
'fr_isothiocyan', 'fr_ketone', 'fr_ketone_Topliss', 'fr_lactam', 'fr_lactone',
'fr_methoxy', 'fr_morpholine', 'fr_nitrile', 'fr_nitro', 'fr_nitro_arom',
'fr_nitro_arom_nonortho', 'fr_nitroso', 'fr_oxazole', 'fr_oxime',
'fr_para_hydroxylation', 'fr_phenol', 'fr_phenol_noOrthoHbond', 'fr_phos_acid',
'fr_phos_ester', 'fr_piperdine', 'fr_piperzine', 'fr_priamide', 'fr_prisulfonamd',
'fr_pyridine', 'fr_quatN', 'fr_sulfide', 'fr_sulfonamd', 'fr_sulfone',
'fr_term_acetylene', 'fr_tetrazole', 'fr_thiazole', 'fr_thiocyan', 'fr_thiophene',
'fr_unbrch_alkane', 'fr_urea', 'qed']

# create molecular descriptor calculator
mol_descriptor_calculator = MolecularDescriptorCalculator(chosen_descriptors)

# use molecular descriptor calculator on RDKit mol object
list_of_descriptor_vals = list(mol_descriptor_calculator.CalcDescriptors(mol))
print(list_of_descriptor_vals)
```

Figure 8: RDKit is an open-source toolkit for cheminformatics. The collection of 200 molecular descriptors (red) was taken from the paper Molecular representation learning with language-models and domain-relevant auxiliary tasks by Fabian et. al. published 2020 (33,34).

These vectors can then be classified as substances that have certain characteristics.

Figure 9 is another example found online for the encoding of substances and researching their potential against HIV (35).

		smiles	activity	HIV_active
0	CCC1=[O+][Cu-3]2([O+]=C(CC)C1)[O+]=C(CC)CC(CC)...	CI	0
1	C(=Cc1ccccc1)C1=[O+][Cu-3]2([O+]=C(C=Cc3ccccc3...	CI	0
2	CC(=O)N1c2ccccc2Sc2c1ccc1ccccc21	CI	0
3	Nc1ccc(C=Cc2ccc(N)cc2S(=O)(=O)O)c(S(=O)(=O)O)c1	CI	0
4	O=S(=O)(O)CCS(=O)(=O)O	CI	0
...	...	...	...
41122	CCC1CCC2c3c([nH]c4ccc(C)cc34)C3C(=O)N(N(C)C)C(...	CI	0
41123	Cc1ccc2[nH]c3c(c2c1)C1CCC(C(C)(C)C)CC1C1C(=O)N...	CI	0
41124	Cc1ccc(N2C(=O)C3c4[nH]c5ccccc5c4C4CCC(C(C)(C)C...	CI	0
41125	Cc1cccc(N2C(=O)C3c4[nH]c5ccccc5c4C4CCC(C(C)(C)...	CI	0
41126	CCCCCC=C(c1cc(Cl)c(OC)c(-c2nc(C)no2)c1)c1cc(Cl...	CI	0

	0	1	2	3	4	5	6	7
0	0.578654	2.969831	-1.355018	-0.007029	4.534556	2.134673	-8.292525	-0.340837
1	0.582014	-0.608330	-5.723546	11.420366	1.116713	-0.251324	-19.153555	0.641661
2	0.239901	-3.120830	-3.069606	6.872059	-0.438361	-2.307343	-9.542530	0.802866
3	2.331327	-0.651370	-6.820276	4.460179	2.888736	2.596256	-10.667131	1.601155
4	1.245269	-2.941263	-0.752463	-0.470454	1.706033	1.063615	-3.766093	2.574552
...	...	...	...	...	...	...	...	...
41122	2.608390	-6.096492	-1.043300	4.551602	-1.026818	-4.465841	-10.286528	0.344521
41123	2.587630	-7.165244	-3.397100	5.840809	-1.007105	-6.378576	-13.567078	-0.032094
41124	2.172952	-6.267159	-2.897906	6.958421	-0.721393	-7.033988	-13.599429	-0.071477
41125	2.144235	-6.164550	-2.714616	7.038599	-0.714388	-6.902690	-13.486728	0.139255
41126	0.006827	-6.562754	-6.269722	11.268374	3.681344	3.841276	-15.926844	-2.676855

[41127 rows x 300 columns]

Figures 9: SMILES and correspondent result columns for the encoding of HIV drugs.

This second example created a database of 41127 substances (vectors) with 300 columns each.

4.4. Decision-making and predictions

The next step during analyses is the extraction of information within the data available to the program and finally the AI models generate decisions. This takes place due to algorithm training. During training the models use mathematical operations to identify patterns and use these to predict with which probability a vector will have a certain property, for instance, how likely it is for one vector to be antibacterial against a certain strain. One of the transformations used to bridge the gap is the "Kernel Trick" which can be explained in a simplified manner for a one-dimensional problem but can be "extended" mathematically for n-dimensions.

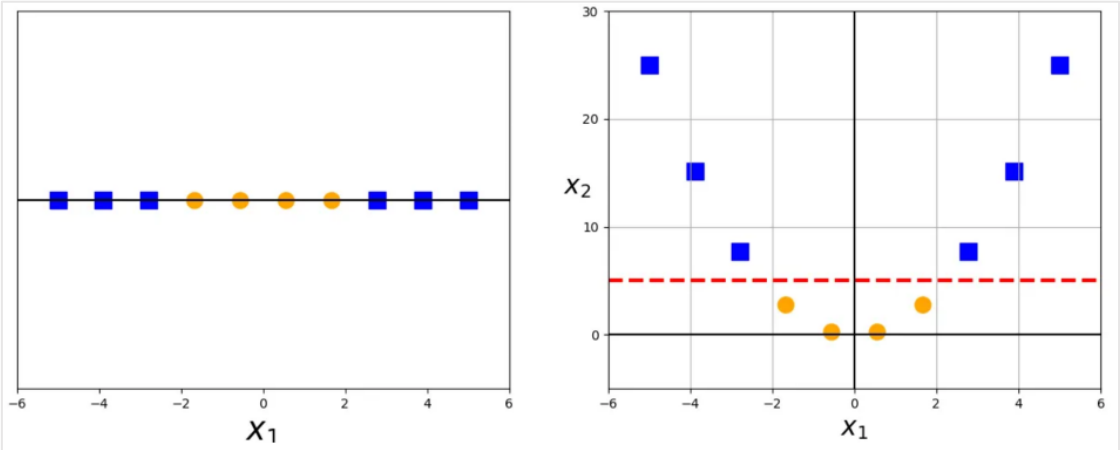


Figure 10: Graphical representation of the Kernel Trick. The data becomes linearly separable after a quadratic transformation to 2-dimensions.

The following transformation takes place: each data point is assigned a y value $f(x) = x^2$, which leads to the two-dimensional graph on the right (**Figure 10**). Here a classification can be done easily using a linear (red dotted line) separator. If, however, the data is not linearly separable, the same mathematical approach can be used for two dimensions. In this case, the result would be transformed into three dimensions and the "separator"

becomes a plane. This enables the use of vector operations to predict, with which probability a new data (point, vector) will have a certain property. In the graph above (**Figure 10**), the value of $x = 10$ would be transform to the vector (10,100), meaning that with a very high probability the result will be a blue square (36).

The equivalent can be done for more complex operations and the transformation equations can be any mathematical functions suitable, e.g. mod, sin (for recurring criteria…)

All in all, AI learns to weigh probabilities and connect "neurons". If a certain threshold is surpassed, a trigger value will then forward information to a next layer. Referring back to the previously illustrated example involving Messi, when he kicks the ball, the following decision model (**Figure 11**) could be executed by the neural network:

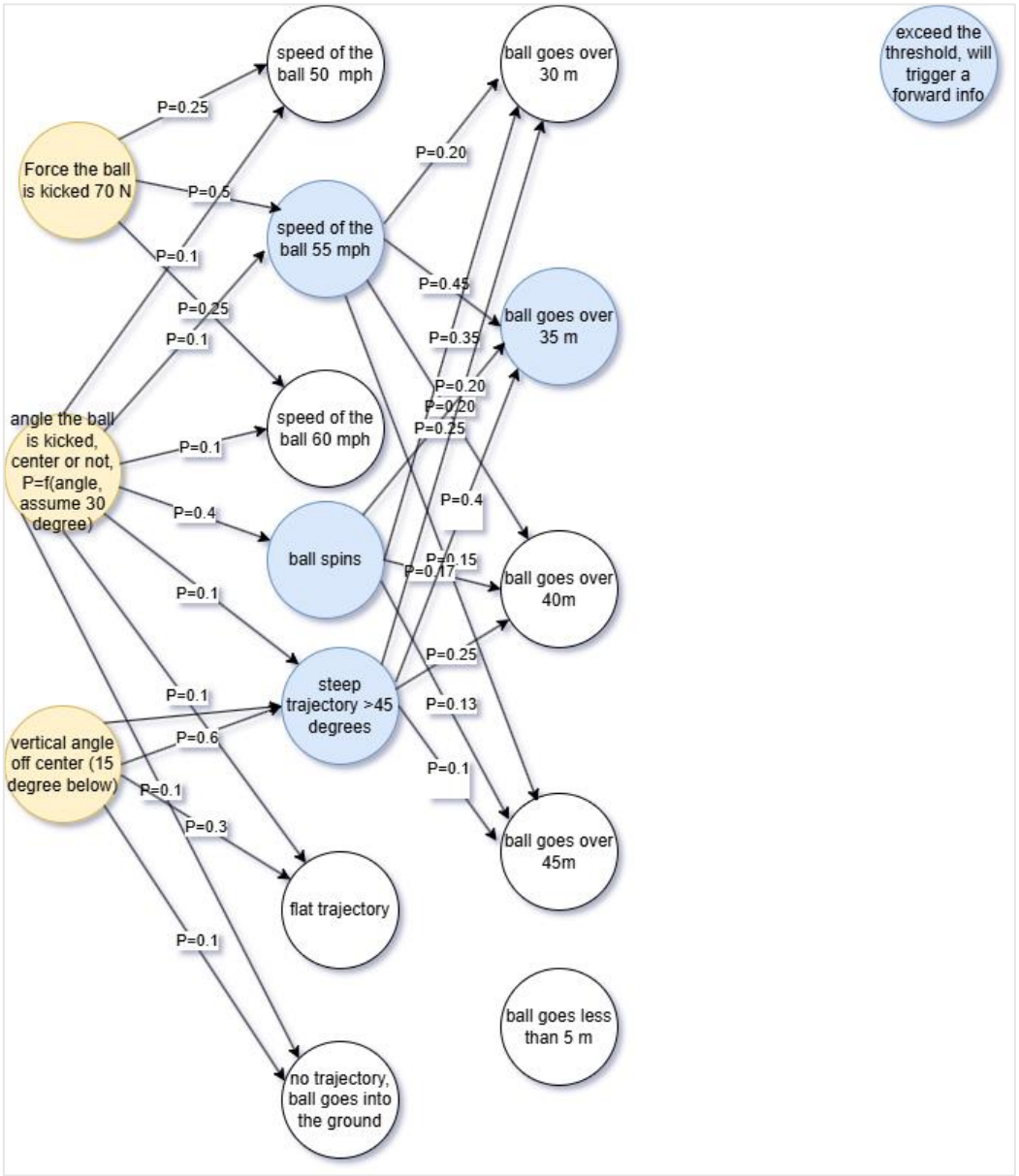


Figure 11: Simple decision model for Messi's weighted probabilities

Here are some fictional "scenarios" or data that a decision model might use to predict how far Messi will kick the ball. In particular, if Messi slips and the vertical angle of the kick is 65 degrees above the center height, he would essentially kick the ball into the ground. In this case, the probability of no significant trajectory would be very high (fictional assumption = 0.9). In the meantime, the probabilities of the ball bouncing back up or traveling any distance would be very low (= 0.05 each). Therefore, the "no trajectory" outcome would be most likely, making the probability of a long-distance kick very small (around 0.05) and the probability of a kick traveling less than 5 meters very high.

To develop such a model, extensive data must be collected, filtered, and standardized to avoid unintended cross-effects. Some of these steps can involve iterative, supervised machine learning processes.

Merely to present the other option briefly, the counterpart to the SMILE system are nodes. An example for this different molecular representation is KNIME. It is an open-source analytic software which offers an analysis interphase able to easily access, prime and model data. KNIME uses nodes, which are the colorful boxes - below in **Figure 12a** - and represent matrices that can perform different tasks that together organize, transform, train the algorithm and/or create the molecular visualizations (37,38).

The following images, **Figures 12a-13**, were taken from a seminar at the University of Southern California (USC), about Computer-aided Drug Design held by PhD. Ian S Haworth in July 2023. The aim is to illustrate the structure of a simple AI workflow design.

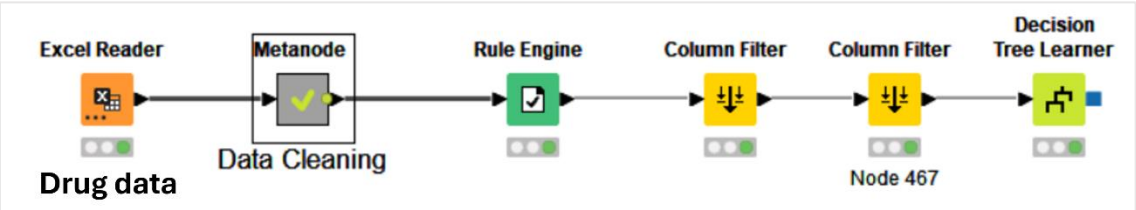


Figure 12a: Decision tree KNIME

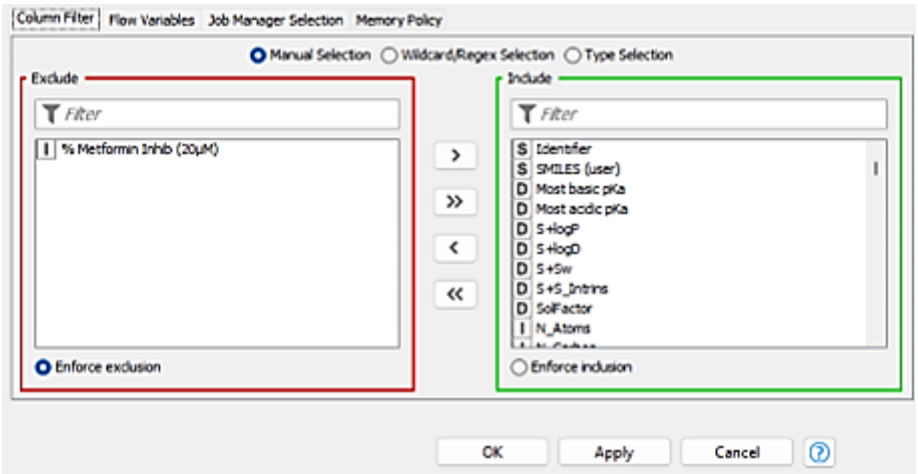


Figure 12b: Details- Column filter KNIME

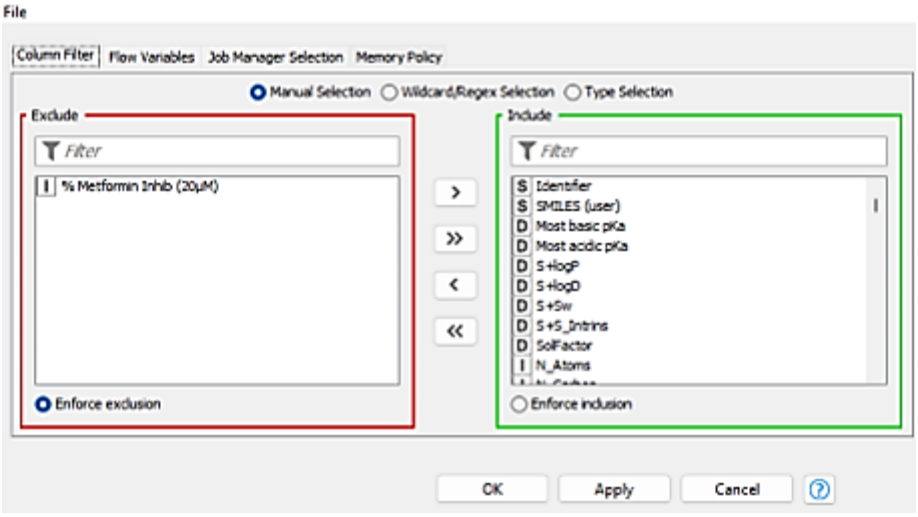


Figure 12c: Details- Decision tree learner KNIME

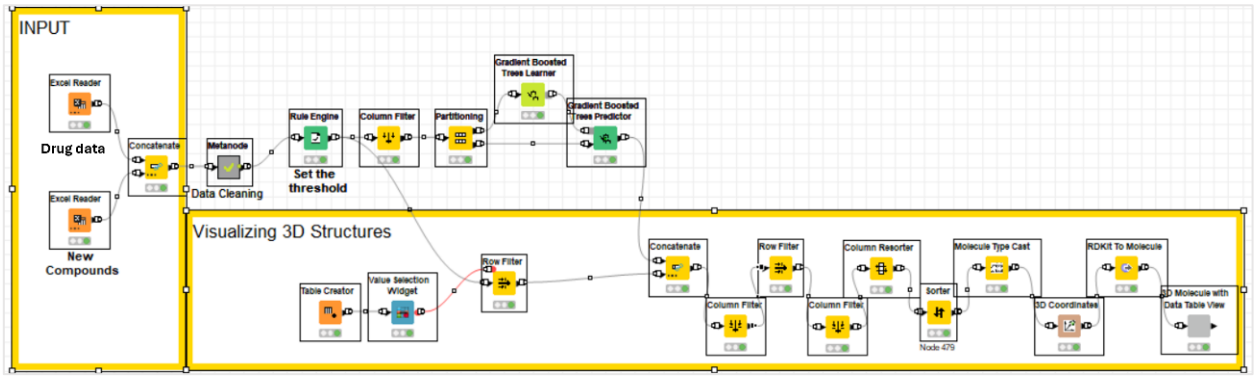


Figure 13: Full in-silico prediction workflow KNIME

4.5. Ethics of AI in health care

With AI, a great opportunity presents to all healthcare professionals to have an additional tool to support and increase the quality of their services. With the rise of AI application in the health sector however, new concerns and ethical dilemmas regarding its safety and regulation come to the surface. It is of upmost importance to prioritize patient data and privacy protection while at the same time meeting the data quantity and quality requirements of AI. This new situation requires adjusted governmental regulations, technological advances and, therefore, funding incentives. Addressing these ethical challenges is essential to be able to fully harness the benefits of AI while safeguarding patient rights and maintaining trust in the healthcare system (39,40).

On the one hand, research and clinical decision-making could be supported by considering various evidence-based sources, as well as electronic health records, clinical trials, lab results. Yet again, the need for data standardization is required to allow interoperability between medical AI systems, coordination and exchange of data between healthcare providers to create cohesive data pools and scientific discovery. Not only the

individual patient would benefit from this thorough analysis, but also in general the population's health.

On the other hand, a significant problem arises regarding responsibility in case of mistakes, particularly when AI is involved. Quintessence to this discussion is, that AI should currently only be seen as a tool, because machines inherently consider their results as absolutely accurate without questioning alternative truths, which requires human approval. Exaggerated accuracy due to intrinsic problems in algorithmic design, known as overfitting, occurs when artificial neural networks learn random fluctuations in the training data rather than generalizable concepts. Overfitting exaggerates accuracy and overestimates the model's clinical performance. AI could also develop wrong results if fed wrong data. This may be accidentally biased, or even with malicious intent to send competing researchers on a wrong path. Therefore, while AI holds great potential, it is essential to address these challenges to ensure ethical and effective use in healthcare.

Additionally, laws defining data ownership must be in place to control, process, and analyze this data effectively. Data protection should be regulated according to accessibility, including the frequency, purpose, and method of data access. For example, the Austrian digital patient information service ELGA requires pharmacists to justify their motive for accessing patient information, regulated by healthcare service providers. In the European Union, regulations such as the General Data Protection Regulation (GDPR), the Cybersecurity Directive, the Medical Devices Regulation, and the Cybersecurity Act of 2016 govern data protection. In the United States, Agencies like the Health Insurance Portability and Accountability Act (HIPAA), Genetic Information Non-discrimination Act (GINA), as well as the Food and Drug Administration (FDA) ensure companies comply with cybersecurity standards in their approved products. The European Union's GDPR sets a high standard for data protection, but there are discrepancies amongst different countries in how these regulations are implemented and enforced. These discrepancies can lead to problems, such as legal gaps, and expose vulnerabilities to hackers, potentially compromising patient information and the consent process if not addressed. It is crucial to understand that data sharing is possible without compromising patient confidentiality. Technological methods to increase data privacy include deidentification, pseudonymization, anonymous resolution, and privacy audits to ensure standards are some possible solutions to face these adversities (41).

4.6. Halicin's discovery

One of Halicin's standout features, beyond its efficacy against various superbugs, is its considerably distinct structure compared to traditional antibiotics. This was only possible due to AI's capacity to gather information, process it and based on this, generate generalized insights from vast datasets (14).

The discovery of Halicin was made after an exploration of the ZINC15 database which compilates compounds that have already undergone clinical trials. As Jo Marchant stated in a Nature news article, the novel AI methodologies are engineered to discover molecules with specific activities rather than solely relying on predetermined structural attributes (42). The efficacy and precision of the predictions significantly depend on the training of the ML algorithm, as illustrated subsequently through an example **Figure 14**.

The experiment by Stokes et al. was divided into four phases, with the initial phase dedicated to training the algorithm, an essential prerequisite for ensuring maximum accuracy (14). During this initial phase, the program underwent evaluation through non-ensembled optimization using 80% - randomly selected - of a database comprising 2335 molecules. Following thirty iterations of running the program with the selected 80%, the team determined the most pivotal parameters for identifying antibacterial drugs according to the model.

The second phase trained the ensemble of optimized models. Ensembling, a technique aimed at enhancing ML performance, involves calculating the average prediction derived from varying data or weights. Initially, 90% of the Drug Repurposing Hub database was processed through the network, followed by 90% of the WuXi anti-tuberculosis library. The top and bottom predicted molecules for each round of analysis (twenty in total) were subjected to in-vitro testing for growth inhibition against *E. coli*. These in-vitro inhibition results, coupled with 10% of new data from the initial database, the remaining unused 10% of the Drug Repurposing Hub dataset, and 10% of the WuXi anti-tuberculosis library, constituted the new final training set. Upon retraining the algorithm, it identified 232 out of the 2911 molecules (7.79%) molecules with inhibitory activity from the new collection.

Lastly, the retrained model analyzed a predetermined segment of the ZINC15 database, exclusively encompassing compounds with low structural similarity as previously justified, to derive the final prediction values.

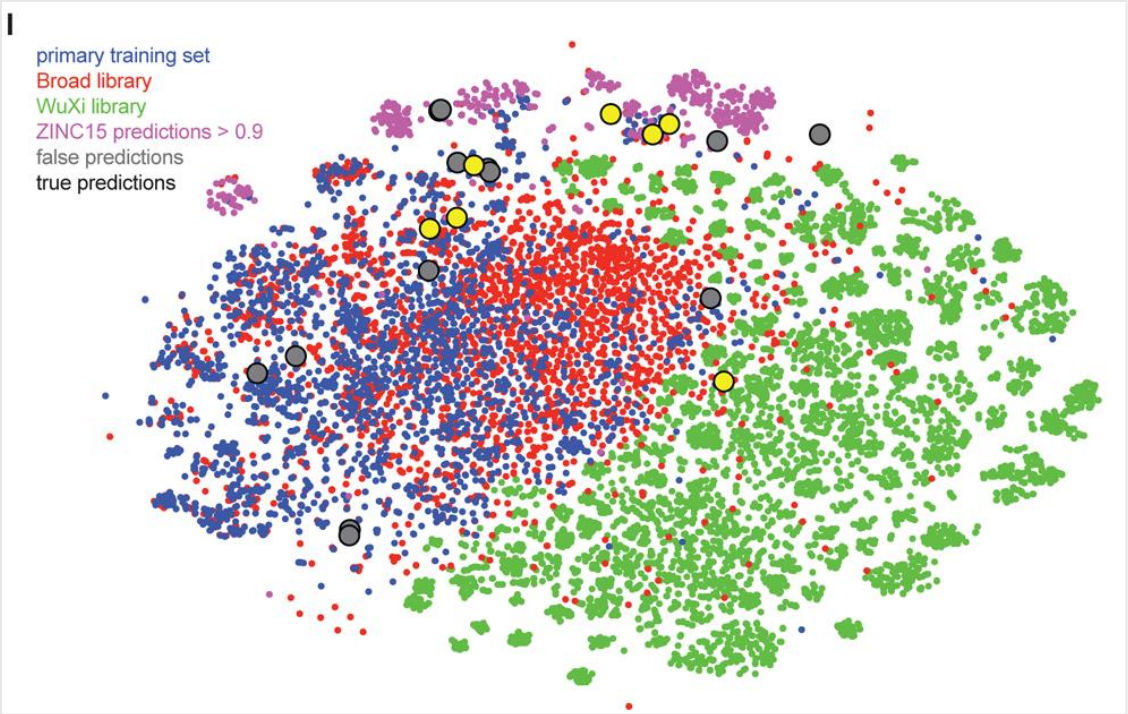


Figure 14: Predicting new antibiotic candidates from unprecedented chemical libraries. False positive predictions (grey), and true positive predictions (yellow) (14).

4.7. Halicin's mechanism of action and characteristics

The first groundbreaking antibiotic discussed in this article - Halicin - was discovered by the Massachusetts Institute of Technology (MIT) in February 2020, utilizing ML. Initially predicted to possess broad-spectrum antimicrobial properties through computational models, succeeding in-vitro experiments substantiated its effectiveness in inhibiting the growth of various strains of *Enterococcus faecalis* and *E. faecium*, with MIC values ranging from 4 to 8 μg/mL (15). Another investigation delved into its efficacy against *Mycobacterium tuberculosis* (MIC 16 μg/mL), CRE, *Clostridioides difficile*, and *pan-resistant Escherichia coli* with a MIC value of 2 μg/mL (14).

In a separate study published in December 2021, Halicin displayed promising activity against additional ESKAPE pathogens. Booq et al. reported varying doses of Halicin exhibiting potent antimicrobial activity against *S. aureus*, *E. coli*, *A. baumannii*, and notably, multidrug-resistant *A. baumannii* (**Table 1**). The strong Pearson correlation coefficients (R = 0,90 to 0,98) for their obtained MIC values, as well as for the inhibition zone assays, reinforced the initial in-silico hypothesis, suggesting Halicin's potential as a versatile antibacterial agent. Furthermore, the study meticulously examined the stability of Halicin across a dilution series. Remarkably, they observed a rise in MIC values over successive experiment cycles, hinting at Halicin's prospective instability after a week of production and storage at 4°C (13). The MICs derived from these three studies exhibited considerable disparities, spanning from 2 μg/mL for *E. coli* to 256 μg/mL for MDR *A. baumannii*. It is essential to consider that various microorganisms - gram-positive, gram-negative, and mycobacteria - possess distinct morphologies and chemical properties, influencing Halcin's efficacy. For instance, gram-positive microorganisms commonly exhibit poor cell-wall permeability. A common and well-known example of this phenomenon is *Enterococcus* developing resistance to aminoglycoside antibiotics like vancomycin. This resistance makes vancomycin ineffective as a standalone treatment for these infections at clinically acceptable doses. Additionally, the experimental settings and methodologies employed to derive these MIC values were different, contributing to the wide spectrum of MIC results. Stokes et al. experimented with an initial cell density of approximately ~10^6 CFU/mL and monitored the bacterial cell killing properties across various Halicin concentrations (14). In contrast, Booq's team conducted two separate serial dilutions with Halicin, ranging from 256 to 0.125 μg/mL. After adding them to a 96-well plate containing bacterial suspensions, one was only incubated overnight at 37 °C while the other serial dilution of Halicin was stored for about a week to observe the antimicrobial efficacy outcome after storage.

Table 1: The zone of inhibition diameters of Halicin. Halicin was effective against all the ATCC bacterial strains at all concentrations, while it was only effective at >128 µg/mL for the MDR *A. baumannii*. The data show a strong correlation coefficient between the MICs and their bacterial inhibitory activity. The results represent the mean (±SD) of n = 3 (13).

Bacterial Strain	Halicin Concentration (µg/mL)	Zone of Inhibition (mm) Mean ± SD (*n* = 3)	Correlation Coefficient (R)
S. aureus ATCC BAA-977	16	17 ± 1	0.90
	32	21 ± 0	
	64	27 ± 1	
	128	29 ± 1	
E. coli ATCC 25922	16	11 ± 1	0.98
	32	14 ± 1	
	64	16 ± 1	
	128	20 ± 0	
A. baumannii ATCC BAA-747	16	10 ± 1	0.98
	32	11 ± 1	
	64	14 ± 0	
	128	20 ± 0	
	256	25 ± 1	
A. baumannii MDR 3086	16	0	0.93
	32	0	
	4	11 ± 1	
	128	14 ± 1	
	256	22 ± 1	

Undoubtedly, Halicin shows promising potential, but thorough investigation of its pharmacokinetic and pharmacodynamic parameters are indispensable. Regulatory agencies must evaluate these aspects to determine Halicin's clinical relevance in the near future. For instance, the FDA has broken down its approval process to provide an overview, encompassing analysis of the target condition and available treatments, assessment of clinical data for benefits and risks, and strategies for risk management (43). Therefore, it's evident that designing new drugs requires scientists' comprehensive understanding of the drug's mechanism (11).

Looking into Halicin's mechanism of action, it functions as a c-Jun N-terminal protein kinase (JNK) inhibitor, disrupting the pH-dependent component of bacteria's membrane proton motive force (13,14). JNK belongs to the mitogen-activated protein kinases (MAPK) family, pivotal in signaling cascades regulating vital cell functions like proliferation, migration, apoptosis, and autophagy. JNKs, also called stress-activated protein kinases (SAPKs), trigger a signaling cascade via phosphorylation of serine and threonine residues. Halicin induces a bacterial stress response to this stimulus by upregulating genes controlling iron homeostasis while downregulating those responsible for motility (44). Consequently, the transmembrane electrochemical gradient is disrupted. This perturbation in iron concentration regulation across the bacterial cell membrane

significantly impacts pH regulation, as iron is crucial for essential cellular mechanisms including oxygen transport, requiring it to be in its ferrous, highly reactive, form. Elevated ferrous concentration, due to Halicin, leads to peroxidations, generating reactive oxygen species (ROS) (45). The accumulation of toxic lipid peroxides and compromised membrane integrity culminates in ferroptosis as seen in **Figure 15**, or iron-induced cell death, halting bacterial growth via this unconventional mechanism.

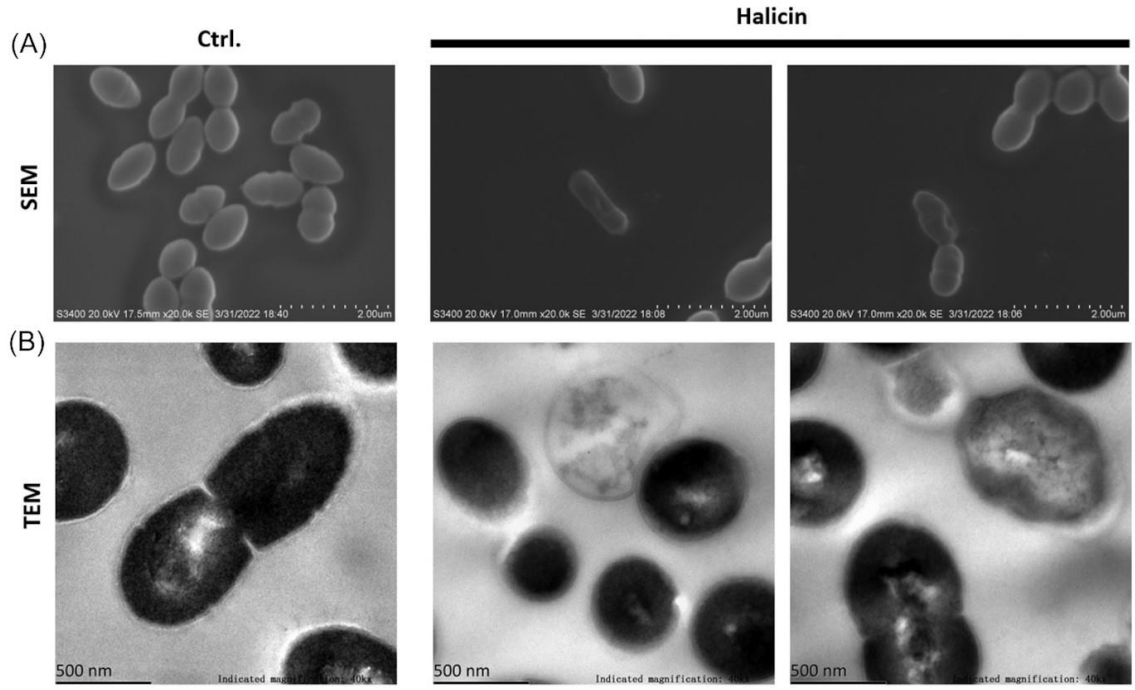


Figure 15: The ultrastructure images of *E. faecalis* ATCC 29212 under scanning electron microscopy (SEM) (A) and transmission electron microscopy (TEM) (B) after treatment with Halicin at concentration of 20 mg/mL and DMSO as control (15).

4.8. Synergy between Halicin and Doxycycline

A brief examination of clinical practice reveals treatment guidelines often incorporate multiple antibiotics to leverage diverse mechanisms simultaneously. Such was the premise of a study prompted by the detection of synergy between Halicin and the antibiotic Doxycycline (DOX) via the checkerboard method.

The study's findings demonstrated that even at sub-MIC concentrations, this combination effectively inhibited the emergence of biofilms and eradicated already formed ones (15). Despite promising in-vitro assay results against various gram-positive and gram-negative strains, the minimum bactericidal concentration (MBC) for *E. faecalis* and proved higher than optimal. Notably, Halicin exhibited a bacteriostatic effect rather than completely eradicating the biofilms formed by these common nosocomial urinary tract pathogens. However, higher concentrations of Halicin and DOX, up to four times the MIC (**Figure 16**), effectively eradicated biofilms and most viable cells.

In-vivo toxicity testing in mice revealed no significant cytotoxic properties associated with the Halicin-DOX combination, as histological samples displayed no notable changes nor biomarker alterations in liver, kidney, lung, or spleen. Moreover, in subcutaneous abscess models of infection, researchers confirmed the dose-dependent antibacterial effect alongside the discovery of anti-inflammatory properties, further underscoring the synergistic efficacy of the combination.

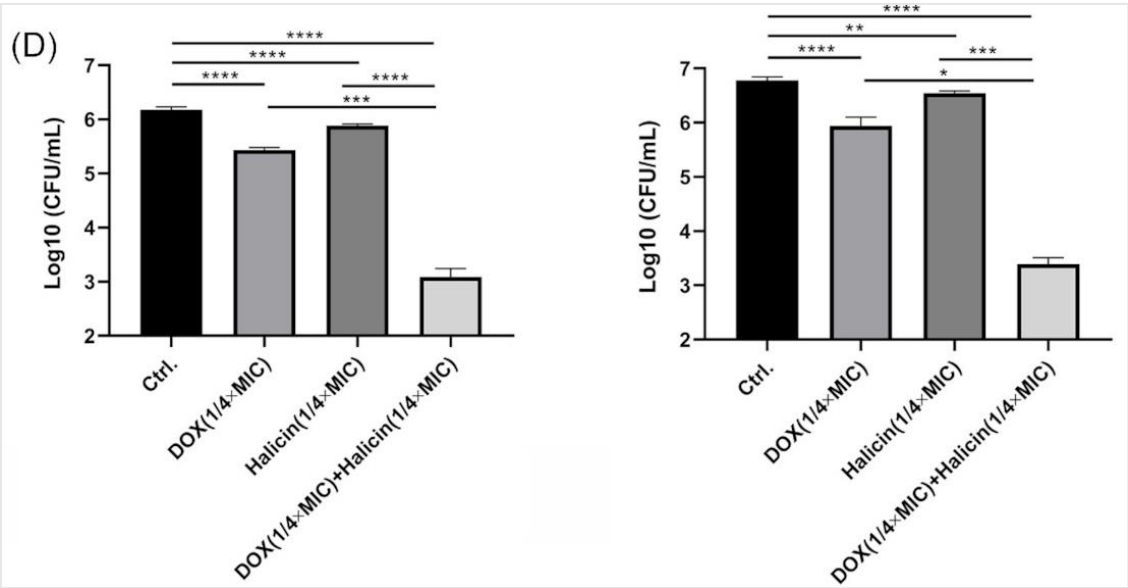


Figure 16: The combined antimicrobial effects between Halicin and DOX. (D) The viable cell count of *E. faecalis* and *E. faecium* after 12-h combined treatment of Halicin and DOX (15).

4.9. Abaucin's mechanism of action and characteristics

The second significant discovery in the field of antimicrobial molecules is Abaucin, a narrow-spectrum antibiotic identified using DL trained on an *A. baumannii* inhibition dataset. Originally designated as RS102895, it was renamed Abaucin because of its modest bactericidal activity against *A. BAUmannii*. The in-silico screening of a dataset containing 7500 molecules that inhibit this ESKAPE pathogens was completed in just a few hours. The initial predictions suggested 240 molecules, and to manage the data, researchers prioritized those with over 80% inhibition likelihood (17).

Abaucin exhibits two noteworthy effects that must be distinguished. Firstly, it disrupts the transport and integration of lipopolysaccharides (LPS) in the bacterial cell membrane by blocking LolE, part of an ATP-binding cassette (ABC) transporter complex in gram-negative microorganisms, as seen in **Figure 17**. *A. baumannii* is notoriously difficult to treat due to its ability to acquire and retain antibiotic resistance. The advantage of targeting a conserved protein like LolE increases the likelihood of prolonged efficacy. Additionally, narrow-spectrum antibiotics like Abaucin help reduce antibiotic resistance and minimize side effects, such as disruptions to the gut or skin microbiome.

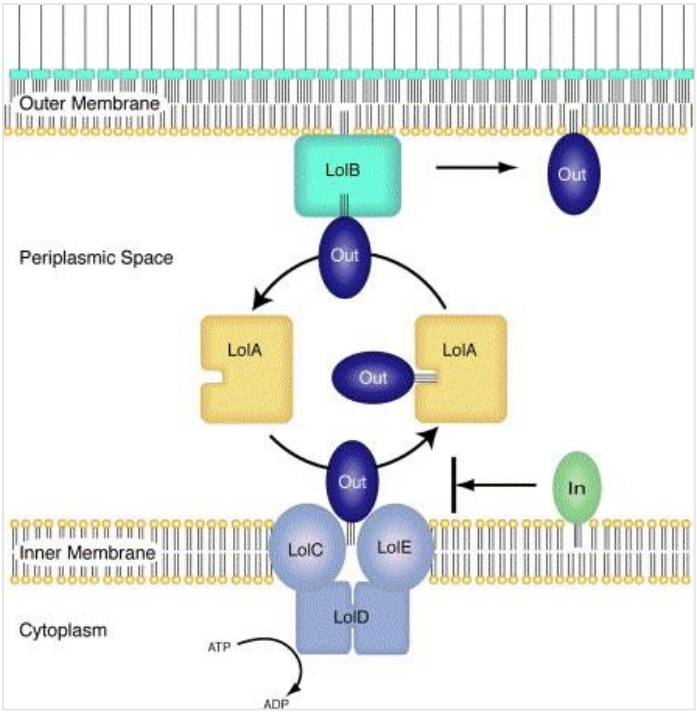


Figure 17: Localization of lipoproteins of Lol system. Arrows represent the transport of lipoproteins by the system. "In" and "Out" represent inner membrane-specific and outer membrane-specific lipoproteins, respectively. The ABC transporter LolCDE recognizes outer membrane-directed lipoproteins and releases them from the inner membrane, causing the formation of a complex between one molecule each of lipoprotein and LolA, a periplasmic carrier protein (46).

Secondly, Abaucin is a well-studied CCR2-selective chemokine receptor antagonist, which can help suppress infections and promote wound healing. This G-protein-coupled receptor participates in a potent proinflammatory signaling pathway that can be activated by LPS exposure. While inflammation is typically a defense mechanism, prolonged or intense responses can damage natural barriers, potentially allowing bacterial survival and easier access into the host (47–49). In a mouse wound infection model, Abaucin's in vivo efficacy was less impressive compared to other antibiotics, although it excelled in treating critical wounds caused by *pan-resistant A. baumannii* strains.

To determine the MIC, which was found to be 2 µg/mL, researchers evaluated Abaucin's activity against 41 clinical isolates from the Center for Disease Control and Prevention Antibiotic Resistance Isolate Bank. Abaucin overcame all resistance mechanisms present in these strains. Further experiments by Liu et al. revealed that Abaucin is most effective during the bacterial growth and division phases. Bacterial killing assays conducted in nutrient-depleted phosphate-buffered saline (PBS) confirmed that its mechanism of action does not involve physical disruption of the phospholipid bilayer, a method that, despite its efficacy, results in significant toxicity in humans due to its lack of specificity for bacterial cell walls (48).

Overall, the discovery of Abaucin opens new avenues for R&D&I in exploring novel chemical spaces, potentially leading to the discovery of more narrow-spectrum

antibacterial molecules. As data volumes continue to grow, machine learning will only improve in aiding such discoveries.

4.10. Zosurabalpin's mechanism of action and characteristics

Finally, the antibiotic Zosurabalpin, a complex macrocyclic peptide, represents a significant milestone in the emerging era of AI-driven R&D&I. Recently discovered in January 2024, it is the only molecule to have not just one, but two completed phase 1 clinical trials. The objective of these studies is to gather critical pharmacokinetic information, enabling more thorough analysis and prediction accuracy in the future. Prior preclinical pharmacokinetic studies indicated a favorable profile for Zosurabalpin's activity. In-vivo research demonstrated that plasma concentration increases are dose-dependent, and that the potential antibiotic's toxicity was not a concern. However, challenges have arisen regarding dosing strategies. Due to its short half-life and rapid renal and fecal clearance, dose management must be optimized to achieve effective therapeutic levels without significantly impacting patient compliance, which is crucial for successful infection treatment.

To address this, one clinical study observed the body's interaction with the administered intravenous (IV) medication throughout the exposure period. The selection criteria were strict, accepting only forty-eight patients hospitalized in critical intensive care units (ICU) with bacterial infections. To date, the primary outcome measures from this study have yet to be evaluated and published. The second study examined pharmacokinetic parameters, safety, and tolerability in healthy participants. All 124 enrolled volunteers received the IV medication, but they were divided into three groups. The first group consisted of elderly participants, evaluated separately due to metabolic degeneration and potential polytherapy. The second group received a single ascending dose of Zosurabalpin, while the third group was administered multiple doses of ascending concentrations. Both trials suggest that Zosurabalpin is safe and well-tolerated at doses up to 2000 mg per day.

Zosurabalpin is a charge-balanced molecule with decent solubility. This entirely synthetic and novel structure, comprising a tripeptide and a non-peptidic subunit, specifically targets the membrane of gram-negative bacteria by binding to the LptB2FGC complex located in the periplasm - between the inner and outer membranes (20). This specific binding site is shown in **Figure 18**. Zosurabalpin requires the presence of LPS to maintain the LptB2FGC complex in a substrate-bound conformation, thereby inhibiting the transport of LPS (19). They are crucial for the membrane's integrity and functionality. An accumulation of LPS inside the cell can reach toxic levels, ultimately inducing apoptosis. For safety reasons, the chemical properties of Zosurabalpin and its binding site with LPS were studied using the non-pathogenic equivalent *A. baylyi*, which aids in protecting researchers and managing hazardous waste (23).

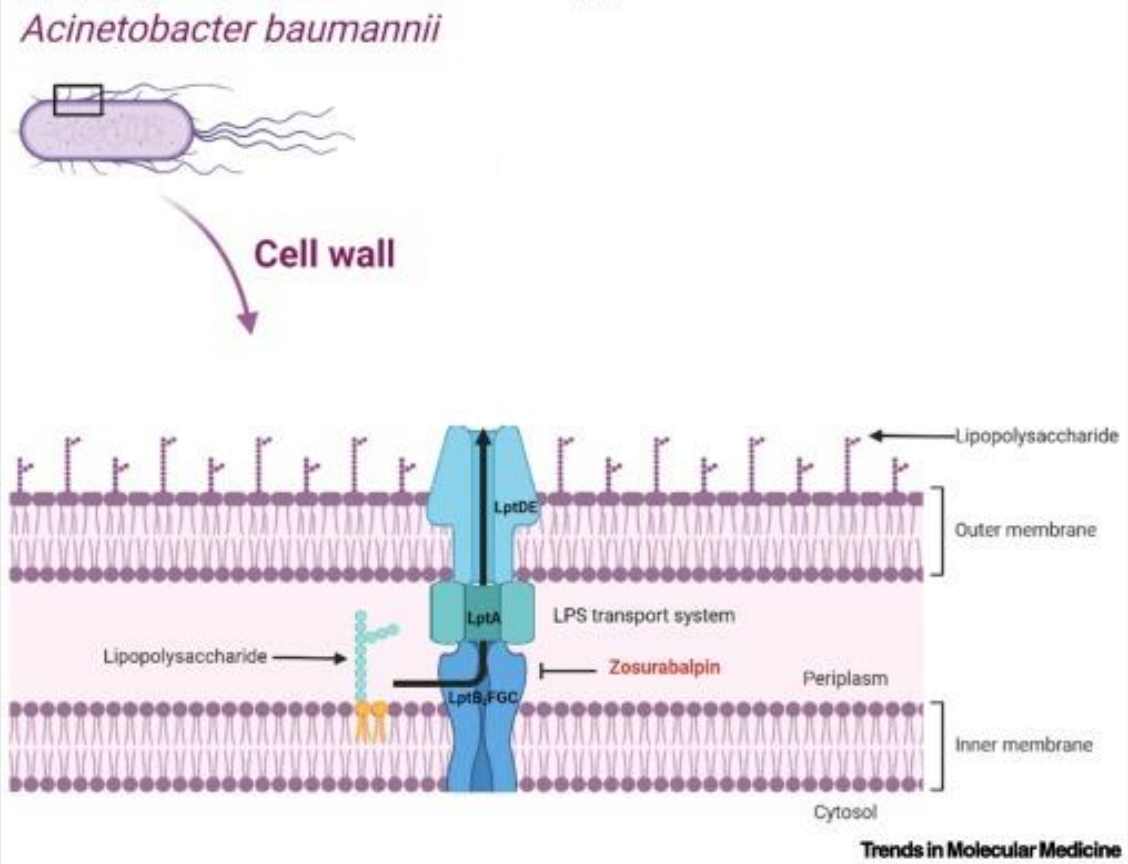


Figure 18: Mechanism of cellular inhibition by Zosurabalpin via disruption of LptB2FGC function in gram-negative *Acinetobacter baumannii*. Zosurabalpin ninhibition results in lethal disruption of LPS transport, culminating in bacterial cell death. Macrocyclic peptide inhibitors selectively target *Acinetobacter spp.* because the LptFG proteins they target are significantly differently from those of *Escherichia coli*, leading to ineffective peptide binding and a distinct LPS conformation in the latter (19).

Therapeutic options that "work around" the resistance mechanisms of MDR *A. baumannii* are scarce. The ABC transporter – Lpt - targeted by Zosurabalpin is a protein complex not coded by an evolutionarily conserved gene, resulting in sequence variation among different gram-negative microorganisms (50). This variability significantly reduces the likelihood of pre-existing resistance mechanisms (20). Since only gram-negative bacteria possess LPS, the selective inhibition of these essential functions offers promise for developing new therapies against the notorious *carbapenem-resistant A. baumannii*. Studies have also demonstrated that this antibiotic reduces dysbiosis in *E. coli* too.

Despite its potential, bacteria can detect environmental threats, such as increasing intracellular LPS concentrations, which may lead to a halt in their synthesis, rendering Zosurabalpin ineffective. However, this scenario is statistically unlikely due to the necessity of an extremely specific mutation, which would likely reduce bacterial virulence - a trait undesirable for the bacteria. Therefore, the risk of spontaneous mutation and loss of activity remains hypothetical until further clinical data is available (20).

The cost and availability of Zosurabalpin have been discussed in various articles. Although the antibiotic has not proven superior, researchers emphasize its potential because of the limited effective alternatives available (19). Parameters such as optimal dosing, synergistic effects with other antibiotics, and the development of new variations for other gram-negative bacteria while maintaining specificity need to be explored. Zosurabalpin's potential will be more comprehensively assessed as clinical outcomes increase and cost-effectiveness studies are conducted (9).

5. Conclusions

1. Neural network programs are a potent option to combat MDR microorganisms in R&D&I.

2. The next step to launch the use of AI, is increasing the quantity of available data.

3. Current lack of standardization in data, the need for robust validation of algorithms, and the difficulty in transferring models across different diseases highlights the necessity for developing general techniques and patterns in AI-driven drug discovery.

4. The “black box” phenomenon, meaning that the user does not fully comprehend the methodology behind the neural network doing the data transformations and pattern analyses, is a problem that still requires a solution.

5. The obtention of clinical information such as pharmacokinetics, dose–response curves, and potential cytotoxicity of new molecules remains critical and cannot be substituted by AI.

6. The fast-paced nature of AI research, combined with its lower labor requirements and cost efficiency, makes it a promising tool in both drug discovery and clinical practice.

7. Data protection and ethical standards need to be met, as well as other regulatory affairs concerning intellectual property rights.

8. AI can produce incorrect results if provided with erroneous data. This may occur due to accidental errors or inherent biases in the data.

6. Bibliography

1. United Nations Publications. Goal 3 | Ensure healthy lives and promote well-being for all at all ages. Available from: https://sdgs.un.org/goals/goal3#targets_and_indicators

2. United Nations Publications. Goal 12 | Ensure sustainable consumption and production patterns. Available from: https://sdgs.un.org/goals/goal12#targets_and_indicators

3. Toma A, Crisan O. Green pharmacy – a narrative review. Clujul Med. 2018 Oct;91(4):391–8.

4. Saderi H, Owlia P. Detection of multidrug resistant (MDR) and extremely drug resistant (XDR) P. Aeruginosa isolated from patients in Tehran, Iran. Iran J Pathol. 2015 Fall;10(4):265–71.

5. Schlander M, Hernandez-Villafuerte K, Cheng CY, Mestre-Ferrandiz J, Baumann M. How much does it cost to research and develop a new drug? A systematic review and assessment. PharmacoEconomics. 2021 Nov;39(11):1243–69.

6. Centers for Disease Control and Prevention. Antibiotic resistance threats in the United States, 2013 Apr. Available from: https://www.cdc.gov/drugresistance/pdf/ar-threats-2013-508.pdf

7. Yelin I, Kishony R. Antibiotic resistance. Cell. 2018 Feb;172(5):1136-1136.e1.

8. Akhondzadeh S. The importance of clinical trials in drug development. Avicenna J Med Biotechnol. 2016 Oct;8(4):151.

9. Guevara YAS, Santos MHC, Gomes FIR, Sheheryar S, Mesquita FP, Souza PFN. A historical, economic, and technical-scientific approach to the current crisis in the development of antibacterial drugs: Promising role of antibacterial peptides in this scenario. Microb Pathog. 2023 Jun;179:106108.

10. Oxford English Dictionary. “artificial intelligence (n.),” meaning. 2023. Available from: https://www.oed.com/dictionary/artificial-intelligence_n

11. You Y, Liu H, Zhu Y, Zheng H. Rational design of stapled antimicrobial peptides. Amino Acids. 2023 Apr;55(4):421–42.

12. Bess A, Berglind F, Mukhopadhyay S, Brylinski M, Griggs N, Cho T, et al. Artificial intelligence for the discovery of novel antimicrobial agents for emerging infectious diseases. Drug Discov Today. 2022 Apr;27(4):1099–107.

13. Booq RY, Tawfik EA, Alfassam HA, Alfahad AJ, Alyamani EJ. Assessment of the antibacterial efficacy of Halicin against pathogenic bacteria. Antibiotics. 2021 Dec;10(12):1480.

14. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A deep learning approach to antibiotic discovery. Cell. 2020 Feb;180(4):688-702.e13.

15. Hussain Z, Pengfei S, Yimin L, Shasha L, Zehao L, Yifan Y, et al. Study on antibacterial effect of Halicin (SU3327) against Enterococcus faecalis and Enterococcus faecium. Pathog Dis. 2022 Jan;80(1):ftac037.

16. National Center for Biotechnology Information. PubChem compound summary for CID 11837140, Halicin. Available from: https://pubchem.ncbi.nlm.nih.gov/compound/11837140

17. Liu G, Catacutan DB, Rathod K, Swanson K, Jin W, Mohammed JC, et al. Deep learning-guided discovery of an antibiotic targeting Acinetobacter baumannii. Nat Chem Biol. 2023 Nov;19(11):1342–50.

18. National Center for Biotechnology Information. PubChem compound summary for CID 10000456, 1'-[2-[4-(trifluoromethyl)phenyl]ethyl]spiro[1H-3,1-benzoxazine-4,4'-piperidine]-2-one. Available from: https://pubchem.ncbi.nlm.nih.gov/compound/10000456

19. Que W, Deng Z, Gao J. Clinical crusade: Zosurabalpin's charge against antibiotic resistance. Trends Mol Med. 2024 Mar;0(0).

20. Gugger MK, Hergenrother PJ. A new type of antibiotic targets a drug-resistant bacterium. Nature. 2024 Jan;625(7995):451–2.

21. Hoffmann-La Roche. A multicenter, single-dose, uncontrolled, open-label, one group study to investigate the pharmacokinetics of RO7223280 in critically Ill patients with bacterial infections. clinicaltrials.gov; 2024 Jan. Report No.: NCT05614895. Available from: https://clinicaltrials.gov/study/NCT05614895

22. Hoffmann-La Roche. A randomized, sponsor-open, adaptive, single- and multiple-ascending dose, placebo-controlled study to investigate the safety, tolerability, and pharmacokinetics of RO7223280 following intravenous administration in healthy participants. clinicaltrials.gov; 2023 Apr. Report No.: NCT04605718. Available from: https://clinicaltrials.gov/study/NCT04605718

23. National Center for Biotechnology Information. PubChem compound summary for CID 148636827, Zosurabalpin. Available from: https://pubchem.ncbi.nlm.nih.gov/compound/148636827

24. Ali T, Ahmed S, Aslam M. Artificial intelligence for antimicrobial resistance prediction: challenges and opportunities towards practical implementation. Antibiotics. 2023 Mar;12(3):523.

25. Lv J, Zhang S, Deng L. A review of artificial intelligence applications for antimicrobial resistance. Biosaf Health. 2021 Feb;3(1):22–31.

26. Wang L, Le X, Li L, Ju Y, Lin Z, Gu Q, et al. Discovering new agents active against methicillin-resistant Staphylococcus aureus with ligand-based approaches. J Chem Inf Model. 2014 Nov;54(11):3186–97.

27. Yuan Y, Shi C, Zhao H. Machine learning-enabled genome mining and bioactivity prediction of natural products. ACS Synth Biol. 2023 Sep;12(9):2650–62.

28. Tajrin N. What is a vector?. Analytics Vidhya. 2020 Jun. Available from: https://medium.com/analytics-vidhya/what-is-a-vector-5c86fc2b57c1

29. R V. What is an Array?. 2021 Aug. Available from: https://www.mygreatlearning.com/blog/what-is-an-array-learn-more-in-one-read/

30. Ganegedara T. Neural network with several input parameters and several output parameters. Stack Overflow. 2019 Dec. Available from: https://stackoverflow.com/q/59126070

31. Kerr LA, Goethel DR. Simulation modeling as a tool for synthesis of stock identification information - an overview | ScienceDirect Topics. 2013 Oct. Available from: https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/model-validation?__cf_chl_tk=V9ZcgCzpaQ8WbaX.FnYzjERZMzhtkWwjrrg7OWrnpyg-1716829474-0.0.1.1-1450

32. U.S. Environmental Protection Agency. SMILES Tutorial. 2016 Feb. Available from: https://archive.epa.gov/med/med_archive_03/web/html/smiles.html

33. Fabian B, Edlich T, Gaspar H, Segler M, Meyers J, Fiscato M, et al. Molecular representation learning with language models and domain-relevant auxiliary tasks. arXiv; 2020 Nov. Available from: https://arxiv.org/pdf/2011.13230v1

34. Dablander M. How to turn a SMILES string into a vector of molecular descriptors using RDKit | Oxford Protein Informatics Group. 2022 Jun. Available from: https://www.blopig.com/blog/2022/06/how-to-turn-a-molecule-into-a-vector-of-physicochemical-descriptors-using-rdkit/

35. George F. Medium. 2021 Sep. Basic molecular representation for machine learning. Available from: https://towardsdatascience.com/basic-molecular-representation-for-machine-learning-b6be52e9ff76

36. Wilimitis D. Medium. 2019 Feb. The kernel trick. Available from: https://towardsdatascience.com/the-kernel-trick-c98cdbcaeb3f

37. KNIME GmbH. Getting started guide. Available from: https://www.knime.com/getting-started-guide

38. KNIME GmbH. Analytics platform user guide. 2024 May. Available from: https://docs.knime.com/latest/analytics_platform_user_guide/index.html#introduction

39. Abdullah YI, Schuman JS, Shabsigh R, Caplan A, Al-Aswad LA. Ethics of artificial intelligence in medicine and ophthalmology. Asia-Pac J Ophthalmol Phila Pa. 2021 Jul;10(3):289–98.

40. Drummond R. MHAonline. 2021 Jul. What is a healthcare actuary and what do they do? Available from: https://www.mhaonline.com/blog/day-in-the-life-of-a-healthcare-actuary

41. Farhud DD, Zokaei S. Ethical issues of artificial intelligence in medicine and healthcare. Iran J Public Health. 2021 Nov;50(11):i–v.

42. Marchant J. Powerful antibiotics discovered using AI. Nature. 2020 Feb

43. Center for Drug Evaluation and Research. Development & approval process | Drugs. FDA; 2023 Jul. Available from: https://www.fda.gov/drugs/development-approval-process-drugs

44. Aburayan WS, Booq RY, BinSaleh NS, Alfassam HA, Bakr AA, Bukhary HA, et al. The delivery of the novel drug 'Halicin' using electrospun fibers for the treatment of pressure ulcer against pathogenic bacteria. Pharmaceutics. 2020 Dec;12(12):1189.

45. Guo R, Fang X, Shang K, Wen J, Ding K. Induction of ferroptosis: a new strategy for the control of bacterial infections. Microbiol Res. 2024 Jul;284:127728.

46. Narita S, Tokuda H. An ABC transporter mediating the membrane detachment of bacterial lipoproteins depending on their sorting signals. FEBS Lett. 2006 Feb;580(4):1164–70.

47. Cerri C, Genovesi S, Allegra M, Pistillo F, Püntener U, Guglielmotti A, et al. The chemokine CCL2 mediates the seizure-enhancing effects of systemic inflammation. J Neurosci. 2016 Mar;36(13):3777–88.

48. Stokes JM, Lopatkin AJ, Lobritz MA, Collins JJ. Bacterial metabolism and antibiotic efficacy. Cell Metab. 2019 Aug;30(2):251–9.

49. Mues N, Chu HW. Out-smarting the host: bacteria maneuvering the immune response to favor their survival. Front Immunol. 2020 May;11:819.

50. Putker F, Bos MP, Tommassen J. Transport of lipopolysaccharide to the gram-negative bacterial cell surface. FEMS Microbiol Rev. 2015 Nov;39(6):985–1002.