Peoplé's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

University of El-chahid Cheikh Larbi Tebessi –Tebessa

Exact sciences and natural and life science faculty

Department of Applied Biology

Field: Natural and Life Sciences

Thesis

Presented with a view to obtaining the Master's degree

In: Biological sciences

Speciality: Molecular and cellular Biology

Entitled :

Antibiotics classification based on efficiency against Bacteria using machine learning

Prepared by:

Radia Abdelhamid & Kelthoum Boughanem

In front the jury:

President: Dr. Amar BENLAKHAL	Larbi Tebessi, Tebessa University
Supervisor: Dr. Ayoub YAHIAOUI	Larbi Tebessi, Tebessa University
Examiner: Dr. Abderrazak KHEDIRI	Larbi Tebessi, Tebessa University

Year: 2022-2023


بسم الله الرحمن الرحيم

Dedication

This sincere study is dedicated with heartfelt gratitude:

To myself, for the unwavering determination, unwavering faith, inner strength, and relentless pursuit of personal growth. Thank you for being a constant companion on this journey of self-discovery.

To my loving parents, whose unwavering support and endless prayers have been a source of inspiration and encouragement throughout my professional journey. Words cannot express the depth of my gratitude and love for you. I have strived to make you proud and bring smiles to the faces of our loved ones.

To all those who believed in me and offered their prayers for my success, and to those who will have the opportunity to read this modest work.

Abdelhamid Radia

"**The best project you will ever work on is you**" -Sonny Franco-

Dedication

To the piece of my soul who's Always been there for me thank you boo

To Mr Benlakhal Amar for your benevolence , kindness, valuable advice, encouragement, patience ,sharing of knowledge ,you have over the years become a true mentor and you will remain so thank you very much .

Ma binome Abdelhamid Radia for her help, patience, advices, and treatment thnx for being here….thnx for everything.

Ma dear besties chaima nadia ibtihel who have done me a favor and contributed directly or indirectly to accomplish this work.

Also to the head of the central laboratory service of the EHS Khaldi Abdelaziz Mrs Sekiou Sabrina and laboratory equipment « Ami Mourad, Lazher,Aniss,fadhila,cherifa,hanane,nahed,bilel,rabie,malek,nawel,rokaia »

Boughanem kelthoum

Acknowledgments

Above all, all thanks and praises go to Allah, the Almighty, for the completion of this work.

We would like to express our heartfelt gratitude to our supervisor, Dr. Yahiaoui Ayoub, for his invaluable expert guidance, consistent support, insightful feedback and suggestions, and valued care throughout all phases of our research journey.

We would like to extend our sincere thanks to Dr. Benlakhal Amar that was always willing to help us and was there for us throughout the last few years

We are also grateful to all the members of the jury who agreed to examine and evaluate our work.

Abstract

_ Antibiotic resistance is a growing global problem that occurs when bacteria evolve to resist the effects of antibiotics. This makes it more difficult to treat infections and illnesses caused by these resistant bacteria. This is a serious concern because it can result in longer hospital stays, higher healthcare costs, and increased mortality rates., making it crucial to identify and classify the efficiency of antibiotics. Weka; is a data mining tool, based on machine learning algorithms that are used to analyze the data and make predictions about the effectiveness of antibiotics. One of the most important aspects of this tool is the use of classification algorithms. These algorithms are designed to classify data into different categories based on specific characteristics.

In our study Antibiotic resistance datasets are classified by six classifier algorithms. these algorithms are Naïve Bayes, J48, SMO, Random forest, Random tree, and Rep tree. To evaluate the effectiveness of classification strategies for determining accuracy and predicting class labels, the algorithm is applied to the dataset using a combination of stratified 10-fold testing and a 66/33 data split. This approach allows for a thorough assessment of the classification algorithm's performance. The results show that the performance of the J48 technique is significantly superior to the other five techniques for the classification of antibiotics data.

Keywords: machine learning, antibiotic resistance, J48, algorithms, weka, classification.

Résumé

Résumé

_ La résistance aux antibiotiques est un problème mondial croissant qui survient lorsque les bactéries évoluent pour résister aux effets des antibiotiques. Il devient alors plus difficile de traiter les infections et les maladies causées par ces bactéries résistantes. Il s'agit d'un problème grave car il peut entraîner des séjours hospitaliers plus longs, des coûts de santé plus élevés et des taux de mortalité plus importants, d'où la nécessité d'identifier et de classer l'efficacité des antibiotiques. Weka ; est un outil d'exploration de données, basé sur des algorithmes d'apprentissage automatique qui sont utilisés pour analyser les données et faire des prédictions sur l'efficacité des antibiotiques. L'un des aspects les plus importants de cet outil est l'utilisation d'algorithmes de classification. Ces algorithmes sont conçus pour classer les données dans différentes catégories en fonction de caractéristiques spécifiques.

Dans notre étude, les ensembles de données sur la résistance aux antibiotiques sont classés par six algorithmes de classification : Naïve Bayes, J48, SMO, Random forest, Random tree et Rep tree. Afin d'évaluer l'efficacité des stratégies de classification pour déterminer la précision et prédire les étiquettes de classe, l'algorithme est appliqué à l'ensemble de données à l'aide d'une combinaison de tests stratifiés 10 fois et d'une division des données 66/33. Cette approche permet une évaluation approfondie des performances de l'algorithme de classification. Les résultats montrent que les performances de la technique J48 sont nettement supérieures à celles des cinq autres techniques pour la classification des données relatives aux antibiotiques.

Mots clés: apprentissage automatique, résistance aux antibiotiques, J48, algorithmes, classification, weka.

الملخص

_ مقاومة المضادات الحيوية هي مشكلة عالمية متزايدة تحدث عندما تتطور البكتيريا لمقاومة تأثير المضادات الحيوية. وهذا يجعل من الصعب أكثر معالجة العدوى والأمراض التي تسببها هذه البكتيريا المقاومة. وهذا مصدر قلق خطير لأنه يمكن أن يؤدي إلى إطالة مدة الإقامة في المستشفى، وارتفاع تكاليف الرعاية الصحية، وزيادة معدلات الوفيات، مما يجعل من المهم للغاية تحديد وتصنيف كفاءة المضادات الحيوية. ويكا؛ هي أداة لاستخراج البيانات، تعتمد على خوارزميات تعلم الآلة التي تستخدم لتحليل البيانات وإجراء تنبؤات حول فعالية المضادات الحيوية. أحد أهم جوانب هذه الأداة هو إستخدام خوارزميات التصنيف. صممت هذه الخوارزميات لتصنيف البيانات إلى فئات مختلفة استنادا إلى خصائص محددة.

في دراستنا، تصنف مجموعات بيانات مقاومة المضادات الحيوية حسب ست خوارزميات مصنفة. هذه الخوارزميات هي Naïve Bayes، J48، SMO، Random forest، Random tree، و Rep tree. لتقييم فعالية إستراتيجيات التصنيف لتحديد الدقة والتنبؤ تسميات الفئة، يتم تطبيق الخوارزمية على مجموعة البيانات باستخدام مجموعة من الاختبارات المصفوفة 10 أضعاف وتقسيم البيانات 33/66. يسمح هذا النهج بإجراء تقييم شامل لأداء خوارزمية التصنيف. تظهر النتائج أن أداء تقنية J48 يتفوق بشكل كبير على التقنيات الخمسة الأخرى لتصنيف بيانات المضادات الحيوية.

كلمات مفتاحية: التعليم الالي، مقاومة المضادات الحيوية، J48، خوارزميات، تصنيف، ويكا.

Table of contents

Table of contents

Table of contents

List of Figures

List of Figures

List of Table

Abbreviation's list

AI Artificial intelligence

AMR Antimicrobial resistance

AUC Area Under Curve

CLI Command-line Interface

CSV Comma Separated Values

DHF Dihydrofolate

DHP Dihydropteroate

DNA Deoxyribonucleic acid

EHRs Electronic Health Records

FP False Positives

GNB Gram-negative bacteria

GNU General Public License

GUI Graphical User Interface

ID3 Iterative Dichotomiser 3

IR Information Retrieval

LIBSVM Library for Support Vector Machines

MAE Mean Absolute Error

Abbreviation's list

MDR Multi drug resistance

MRSA Methicillin-resistant Staphylococcus aureus

MSE Mean Squared Error

PRC Precision-Recall Plot

QP Quadratic Programming

RAE Relative Absolute Error

Rep tree Reduced Error Pruning Tree

RMSE Root Mean Squared Error

RNA Ribonucleic acid

ROC Receiver Operating Characteristics

RRSE Root Relative Squared Error

SMO Sequential Minimal Optimization

SVM Support Vector Machines

THF Tetrahydrofolate

TP True Positives

VRE Vancomycin-resistant Enterococci

WEKA Waikato Environment for Knowledge Analysis

WHO World Health Organization

General Introduction

In the healthcare organizations and medical diagnosis, artificial intelligence and machine learning techniques and algorithms can be implemented to treat some hazardous diseases. **(Kunjir et all. 2017)**. One increasingly relevant public health concern is antibiotic resistance **(Do Nascimento,2015).**

A 2014 World Health Organization report states that antibiotic resistance is rising to dangerously high levels worldwide. Novel mechanisms of resistance are continuously emerging and disseminating on a global scale. Antibiotic resistance is accelerated by the misuse and overuse of antibiotics, as well as poor infection prevention and control, The World Health Organization has already stated that if we do not act quickly, we are entering a post-antibiotic era in which common infections and minor infections can once again become fatal **(WHO, 2014).**

Accurate prediction of antibiotic resistance risk holds immense global significance as it empowers clinicians to make informed decisions regarding the selection of appropriate antibiotics. This, in turn, can contribute to reducing levels of antibiotic resistance, enhancing patient treatment outcomes, and ultimately lowering healthcare costs **(Do Nascimento, 2015).**

Artificial intelligence has proven to be a valuable asset in various scientific and practical fields, offering time and effort savings. In the specific context of combating antibiotic resistance, which poses a significant threat to humanity, there is a critical need for efficient collection, organization, and classification of data.

The main objective of this study is to utilize the Weka machine learning toolkit to classify antibiotics based on their effectiveness against bacteria. By employing Weka's advanced algorithms and tools, the study aims to compare the accuracy of various classification algorithms. The experimental findings aim to evaluate the performance of these algorithms and determine the most effective algorithm for classifying the provided antibiotic dataset.

Chapter 01

Antibiotics

Introduction

The development of antibiotics is widely regarded as one of the significant medical breakthroughs of the 20th century. These medications have brought about a revolution in the treatment of bacterial infections and have played a crucial role in saving numerous lives.

1. Definition

The Greek origins of the term "antibiotic" are "anti" for "against" and "bioticos" for "associated to life. This term was introduced into modern medicine in 1947 by the biochemist/microbiologist Selman Waksman **(Stachelek, 2021)**. He defined antibiotics as a chemical substance, produced by micro-organisms, which has the capacity to inhibit the growth of and even to destroy bacteria and other micro-organisms. The action of an antibiotic against micro-organisms is selective in nature, some organisms being affected and others not at all or only to a limited degree; each antibiotic is thus characterized by a specific antimicrobial spectrum. The selective action of an antibiotic is also manifested against microbial vs. host cells. …some antibiotics have remarkable chemotherapeutic potentialities and can be used for the control of various microbial infections in man and animal **(Sánchez & Demain, 2015).**

2. Mode of action of antibiotics

2.1. Bactericidal

All antimicrobial treatments which directly kill bacteria.

They had an impact on lipids, enzymes, protein production, and the cell wall.. By disrupting the cell wall structure of existing cells and inhibiting the formation of new cells, bactericidal substances cause bacterial cells to die off (irreversible cell death) **(ultra fresh, 2019)**.

2.2.Bacteriostatic

inhibit the growth and multiplications of bacterial cells, further growth and DNA replication by obstructing the metabolic mechanisms of the bacterial cell (does not cause cell death). It produce reversible results **(ultra fresh, 2019).**

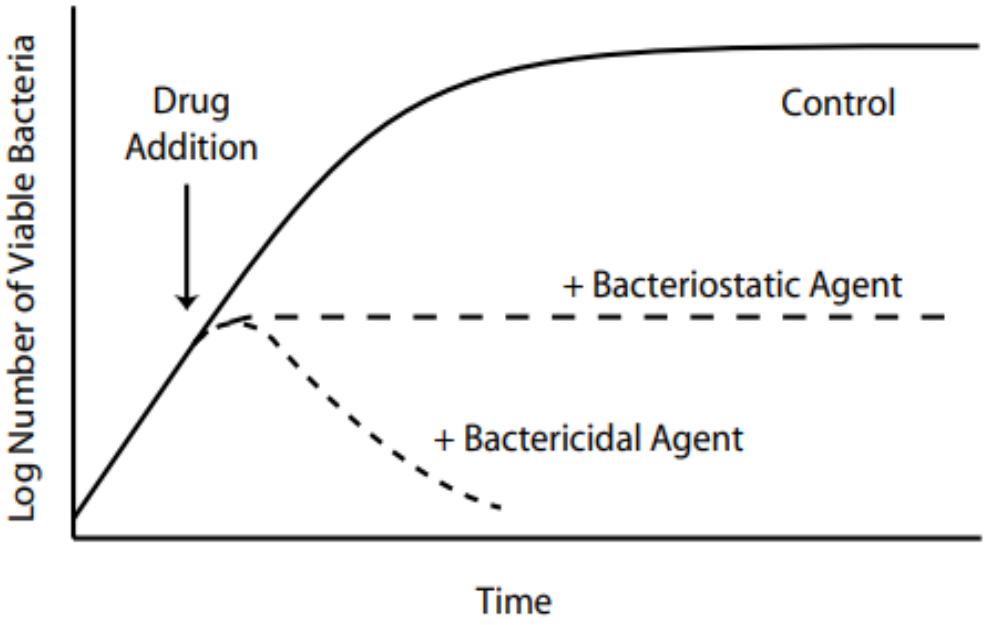


Figure 1 Effects of bacteriostatic versus bactericidal antibiotics on a logarithmically growing bacterial culture **(Walsh, 2003)**.

3. Mechanisms of action of antibiotics

For the growth and dividing of bacterial cells,organisms must synthesize or take up many types of biomolecules. Antimicrobial agents interfere with specific processes with bacterial cells such as : inhibitors of bacterial and fungal cell walls, inhibitors of cytoplasmic membranes, inhibitors of ribosome function (Protein synthesis) , inhibitors of nucleic acid synthesis and inhibitors for metabolites or growth factors needed in bacterial metabolism (competitive inhibitors) **(Neu & Gootz, 2001).**

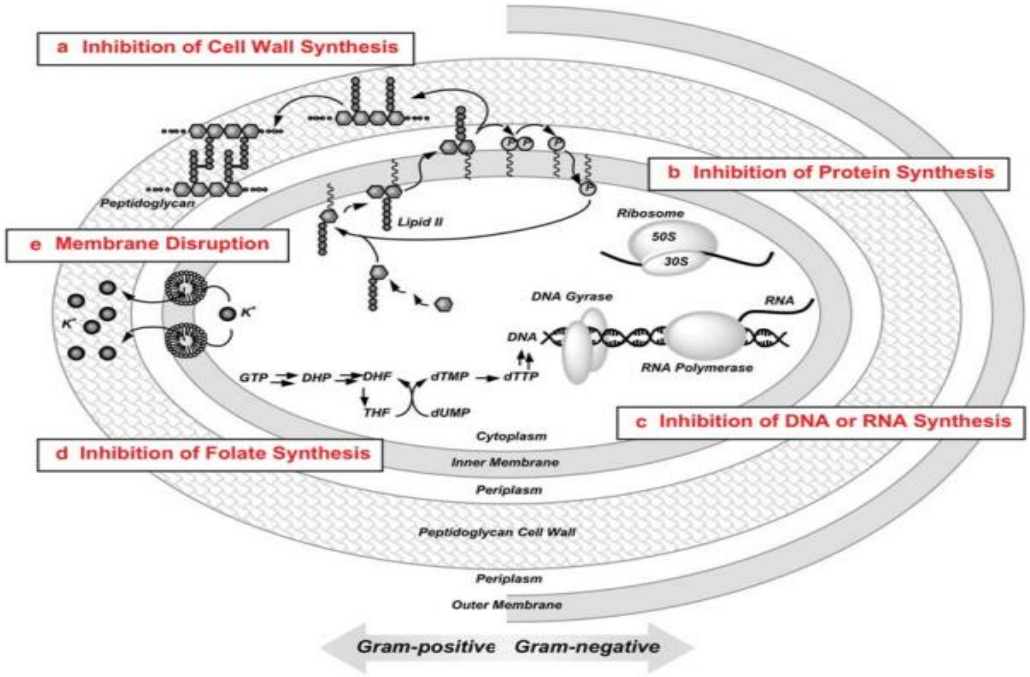


Figure 2 Five classes of bacterial machinery comprise the targets of the major classes of antibiotics: (a) cell wall biosynthesis, (b) protein biosynthesis, (c) macromolecular synthesis (DNA and RNA) and metabolism, (d) the folate biosynthetic pathway (interdicting the supply of deoxythymidylate for DNA synthesis), and (e) membrane function. DHP, dihydropteroate; DHF, dihydrofolate; THF, tetrahydrofolate **(Walsh & Wencewicz, 2016).**

3.1.Selective toxicity

3.1.1. Inhibitors of bacterial cell wall synthesis

Eukaryotic cells in humans lack peptidoglycan and a cell wall, distinguishing them from prokaryotic cells. This distinction provides an advantage for antibiotic therapy, as the therapy can specifically target the bacterial cell wall without affecting human cells **(Kaufman, 2011).**

Gram-negative bacteria (GNB) exhibit structural differences in their cell wall compared to Gram-positive bacteria. These differences affect the ability of chemical agents to penetrate and remain within the bacterial cells. Gram-negative bacteria possess an envelope composed of three main layers **(Exner et all., 2017):**

1. The outer membrane contains lipopolysaccharide/endotoxin, which can potentially be harmful.

2. The peptidoglycan cell wall consists of peptide chains that are partially cross-linked, and it is situated between the outer membrane

3. and the inner membrane is located inside the cell and forms the boundary between the cytoplasm and the cell wall

Gram-positive bacteria typically do not possess an outer membrane. The primary role of the outer membrane in Gram-negative bacteria is to act as a barrier that restricts the entry of certain drugs and antibiotics into the cell. This characteristic is a significant factor contributing to the inherent antibiotic resistance observed in Gram-negative bacteria.

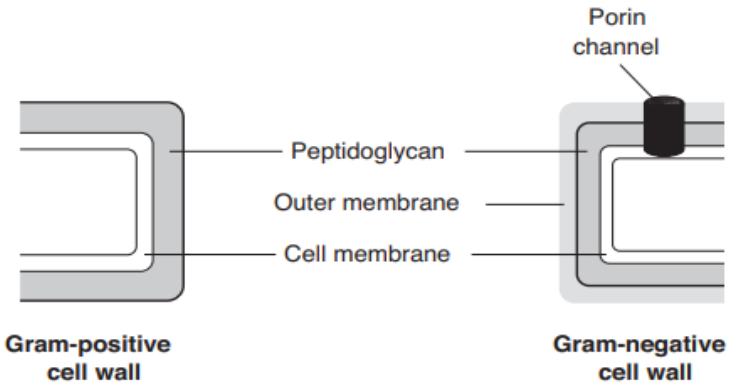


Figure 3 Cell Walls of Gram-Positive and Gram-Negative Organisms **(Gallagher & MacDougall, 2014).**

Antibiotics such as penicillins and cephalosporins, which belong to the class of beta-lactam antibiotics, work by disrupting the synthesis of peptidoglycan. These antibiotics are able to enter the bacterial cell and bind to specific enzymes called penicillin-binding proteins. As a result, the formation of a weakened or deformed cell wall occurs, leading to the swelling and eventual bursting of the bacterial cell. This mechanism of action ultimately results in the destruction of the bacterial cell, making these drugs bactericidal in nature **(Kaufman, 2011).**

3.1.2. Cell membrane inhibitors

Different classes of antibiotics target the cell membranes of bacteria, and their specificity is determined by the variations in the types of lipids present in the cell membranes of different microbial groups. For instance, Daptomycin works by depolarizing the calcium-dependent membrane of bacteria, which ultimately leads to the halt of macromolecular synthesis and

disruption of the cellular membrane. On the other hand, polymyxins exert their effect by binding effectively to the lipid portion of the lipopolysaccharide in the bacterial cell, resulting in the disintegration of the bacterial cell membrane **(Etebu & Arikekpar, 2016).**

3.1.3. Inhibitors of nucleic acid synthesis

Most compounds that directly bind to the DNA double helix tend to be highly toxic to mammalian cells. However, a select few that specifically interfere with enzymatic processes associated with DNA have sufficient selectivity to be used as antibacterial agents. Antibacterial quinolones, novobiocin, and rifampicin (rifampin) are examples of such compounds. Additionally, diaminopyrimidines, sulfonamides, 5-nitroimidazoles, and possibly nitrofurans also impact DNA synthesis and fall under this category **(Finch et all., 2010).**

3.1.4. Inhibitors of bacterial Protein synthesis

Protein synthesis is a crucial and complex biological process that involves the synthesis of specific proteins in living cells. It consists of two main steps: transcription and translation, which encompass initiation, elongation, termination, and recycling. Antibiotics that inhibit protein synthesis exploit the structural differences between bacterial ribosomes and eukaryotic ribosomes. These antibiotics can selectively impede bacterial growth by targeting either the 30S or 50S subunit of the bacterial ribosome, which is part of the 70S ribosome. By blocking bacterial protein synthesis, these antibiotics halt or slow down cell growth**. (Abushaheena et all., 2020)**

3.2.Competitive inhibitors

Many synthetic chemotherapeutic agents function as competitive inhibitors of vital metabolites or growth factors required for bacterial metabolism. As a result, these antimicrobial agents are often referred to as anti-metabolites or growth factor analogs, as their purpose is to specifically hinder an essential metabolic pathway in bacterial pathogens. Chemically, competitive inhibitors bear structural similarities to bacterial growth factors or metabolites, but they fail to fulfill their metabolic function within the bacterial cell. Some of these inhibitors exhibit bacteriostatic effects, while others are bactericidal. The selective toxicity of these agents is based on the fact that the targeted bacterial pathway does not occur in the host organism. For

example, sulfonamides and trimethoprim, when combined as sulfamethoxazole-trimethoprim, demonstrate this selective toxicity **(Chopra, 2002).**

4. Classification of antibiotics

Antibiotics have been categorized using various methods, but the most common classification is based on their molecular structure, mechanism of action, and spectrum of activity. Another approach to classifying antibiotics is based on the route of administration (**Fomnya et all., 2021).**

4.1.Classification based on molecular structure

Antibiotics can be classified based on their molecular structure. This classification system categorizes antibiotics according to the chemical composition and structural characteristics of the molecules. **(Fomnya et all. 2021).**

- **Beta-lactams:** Penicillins, Cephalosporins, Carbapenems, Monobactams
- **Tetracyclines and glycylcyclines:** Examples include tetracycline, tigecycline, doxycycline, minocycline, chlortetracycline, and oxytetracycline.
- **Chloramphenicol:** Chloramphenicol
- **Aminoglycosides:** Gentamicin, Amikacin, Tobramycin, Netilmicin, Streptomycin, Neomycin, Kanamycin
- **Quinolones:** Ciprofloxacin, Norfloxacin, Levofloxacin, Moxifloxacin, Gemifloxacin, Ofloxacin, Enrofloxacin
- **Macrolides and Ketolides:** Azithromycin, Telithromycin, Erythromycin, Clarithromycin
- **Lincosamides:** Lincomycin, Clindamycin, Pirlimycin
- **Streptogramins:** Quinupristin/Dalfopristin, Pristinamycin, Virginiamycin

- **Glycopeptides:** Vancomycin, Teichoplanin, Telavancin, ramoplanin, decaplanin
- **Sulfonamides:** Sulfadiazine, Sulfamethizole, Sulfamethoxazol, Sulfasalazine, Sulfisoxazole
- **Trimethoprim:** Trimethoprim
- **Polymixins:** Colistin, Polymixin B
- **Oxazolidinones:** Linezolid, Tedizolid
- **Lipopeptides:** Daptomycin, Surfactin, Mycosubtilin
- **Ansamycins:** Rifampicin, Rifamycin, Geldanamycin, Streptovaricin, Ansalactam .

4.2.Classification based on mechanism of action

The classification is based on the mechanism of action as follows **(Fomnya et all. 2021):**

- **Interference with cell wall synthesis**: beta lactams, glycopeptides
- **Inhibition of protein synthesis**: macrolides, aminoglycosides, tetracyclines
- **Interference with nucleic acid synthesis**: quinolones
- **Inhibition of metabolic pathways:** sulphonamides
- **Disorganizing the cell membrane structure or function**: lipopeptides and polymyxins.

4.3.Classification based on Origin

According to the definition provided by the Merriam-Webster dictionary, an antibiotic is described as a substance produced by or derived from a microorganism, or a semisynthetic derivative, that has the ability to inhibit or kill another microorganism when present in a diluted solution. This definition would include natural antibiotics like penicillin, which is derived from

a mold, as well as semisynthetic derivatives like ampicillin, which is derived from penicillin. However, it would exclude entirely synthetic agents such as sulfonamides and quinolones. Additionally, drugs like chloramphenicol, initially discovered as a product of soil bacteria but now produced entirely through chemical synthesis, would fall into a more ambiguous category **(Carlos & Cuevas, 2016).**

4.4.Classification based on spectrum of activity

The spectrum of activity is another basis for classifying antibiotics, as outlined in the study by **(Fomnya et all. 2021):**

- **Extended spectrum:** These antibiotics demonstrate efficacy against a range of bacterial, rickettsial, and protozoan organisms. An example is the tetracycline class of antibiotics.

- **Broad spectrum**: Antibiotics classified as broad spectrum are effective against both Gram-positive and Gram-negative bacterial organisms. Quinolones are an example of broad-spectrum antibiotics.

- **Narrow spectrum:** This category includes antibiotics that are specifically effective against either Gram-positive or Gram-negative bacterial organisms. Macrolides are an example of narrow-spectrum antibiotics.

4.5.Classification based on route of administration

Antimicrobial agents are also classified based on the mode of administration **(Fomnya et all. 2021):**

- **Oral:** These antibiotics are taken through the mouth and enter the body through the oral cavity.

- **Parenteral:** Antibiotics in this category are administered through injection, delivering them directly into the bloodstream or muscle.

- **Topical:** These antibiotics are applied directly to body surfaces, such as the skin or mucous membranes.

5. Complications of antibiotic therapy

Although antibiotics are undoubtedly one of the most beneficial discoveries of science, their use does carry risks. They can adversely affect patients by eliciting allergic reactions, causing direct toxicity, or altering the normal bacterial flora, leading to superinfections with other organisms. The main factor behind the rise in antibiotic resistance is the overuse of antibiotics. which can affect not only treated patients but other patients by transmission of resistant organisms **(Gallagher & Macdougall, 2023).**

Conclusion

As we mentioned in the first chapter, an antibiotic is a medicine to combat bacterial infections and eliminate them either by killing them completely or by inhibiting them.

However, the world of medicine was confronted with the problem of antibiotic resistance, as bacteria introduced mutations that were counterproductive to the effectiveness of antibiotics, and this is what we will discuss in the second chapter, which is titled Antibiotic Resistance.

Chapter 02

Antibiotic Resistance

Introduction

Antibiotic resistance poses a serious threat to global health, as it compromises our ability to treat infections effectively. It leads to prolonged illnesses, increased mortality rates, and higher healthcare costs. Additionally, the emergence of antibiotic-resistant bacteria complicates the management of common infections, making them more difficult to control and increasing the risk of severe complications.

1. History

The emergence of antimicrobial resistance has posed a global challenge. However, the discovery of penicillin by Alexander Fleming in 1928 marked the beginning of a golden era for antibiotics. Initially, penicillin had limited use but was widely prescribed for serious infections starting in the 1940s, ultimately saving millions of lives. The golden era relied on the strategy developed by Selman Waksman, involving the screening of natural product extracts to find new antibiotics. However, by the mid-1960s, the discovery of novel antibiotic sources became increasingly challenging, and the development of resistance further complicated the search for new antibiotics. This led to a new era called the medicinal chemistry era, where synthetic analogues of natural antibiotics were created to overcome these challenges. During this period, many new and improved antibiotics were discovered, addressing pharmacological aspects, antimicrobial spectrum, and resistance issues. However, the pool of antibiotic reservoir began to diminish by the early 1980s, resulting in a scarcity of new discoveries. The resistance era emerged in the 1980s, focusing on target-based discovery of antibacterial agents. Unfortunately, this model of drug discovery failed to generate a sufficient number of new antibiotics to combat the growing resistance problem. Each new class of antibiotics introduced was soon followed by the emergence of resistance. The timeline of antibiotic introductions and the subsequent development of resistance is depicted in Figure 4. **(Hashm, 2020)**.

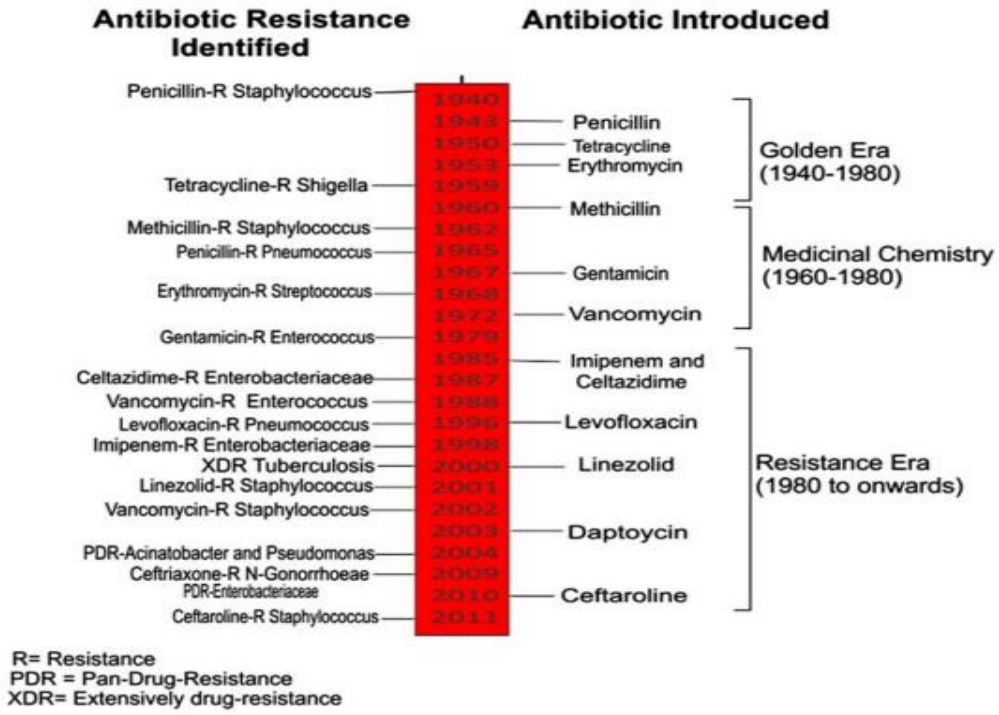


Figure 4 Timeline of development of antibiotics and their emerging resistance **(Hashm, 2020).**

2. Definition

Bacterial resistance to antibiotics refers to the phenomenon where bacteria become unaffected by the antimicrobial effects of a specific antibiotic. This can occur when the bacteria develop mechanisms to either neutralize the antibiotic or continue growing in its presence, rendering the antibiotic ineffective **(Chinedum, 2005).** The presence of antimicrobial-resistant pathogens in infections complicates treatment and raises the likelihood of disease transmission, severe illness, and mortality **(WHO, 2022).**

3. Types of antibiotic resistance

3.1.Natural (Intrinsic, Structural) resistance

Innate resistance is a basic form of resistance found in certain species, strains, or groups of bacteria. It is a natural trait that renders a microorganism insensitive to specific groups of antibiotics. Innate resistance can result from various factors such as the absence of specific receptors for the antibiotic, low affinity, impermeability of the cell wall, or the production of certain enzymes **(Chmiel & Marek, 2022).**

3.2.Acquired resistance

Acquired resistance refers to the development of resistance in bacteria as a result of changes in their genetic characteristics. This type of resistance occurs when bacteria that were previously susceptible to antibiotics become unaffected by them. Acquired resistance can be attributed to alterations in chromosome structures or the presence of extrachromosomal elements such as plasmids or transposons **(Cesur & Demiroz, 2013).**

- **Chromosomal resistance** occurs as a consequence of random mutations in the bacterial chromosome. These mutations can be induced by various physical and chemical factors.

- **Extrachromosomal resistance** is based on the presence of genetic materials outside the bacterial chromosome, which can be transmitted through mechanisms such as plasmids, transposons, and integrons **(Hasan & Al-Harmoosh, 2020).**

4. Multi drug resistance (MDR)

The emergence of Multiple Drug Resistant (MDR) microbes is a direct result of the widespread use of various available antibiotics. MDR refers to the ability of microorganisms to exhibit resistance against three or more classes of antibiotics, even at concentrations that would normally inhibit their growth. It is characterized by the insensitivity of microorganisms to antimicrobial medications, despite having been susceptible to those same drugs in the past. Over time, microorganisms can accumulate multiple resistance traits, making them immune to multiple categories of antibiotics. These microorganisms undergo mutations or adaptations that reduce or eliminate the effectiveness of antibiotics, allowing them to survive without being harmed. According to the World Health Organization, these resistant microorganisms possess the ability to withstand the effects of antimicrobial drugs, resulting in unsuccessful treatment and leading to the persistence and spread of infections. Consequently, the use of antimicrobials promotes the development of drug-resistant microorganisms and the dissemination of resistance genes within populations. Multidrug resistant microorganisms can also exhibit resistance to clinically important therapeutic agents, contributing to treatment failures in the management of infectious diseases **(Osagie & Olalekan, 2019).** The ESCAPE pathogens *(Enterococcus faecium, Staphylococcus aureus, Clostridium difficile, Acinetobacter species, P. aeruginosa, and Enterobacteriaceae)* represent the most significant emerging threats associated with antibiotic resistance. These organisms have the ability to evade the effects of antimicrobial

agents and are responsible for the majority of infections acquired in healthcare settings **(Kaye & Pogue, 2015).**

4.1.MDR Gram-positive bacteria

Gram-positive bacteria, including *Streptococcus pneumoniae, Staphylococcus aureus*, and *Enterococci*, are extensively studied pathogens commonly found in healthcare settings and known to have developed resistance to multiple drugs. Notably, there is significant concern regarding the emergence of *Vancomycin-resistant Enterococci* (VRE) and *methicillin-resistant Staphylococcus aureus* (MRSA), with some MRSA strains even acquiring resistance to vancomycin, leading to the development of VRSA. It is important to note that MRSA can also cause skin infections in individuals who are otherwise healthy, as evidenced by well-publicized cases of community-acquired infections **(Albrecht, 2018).**

4.2.MDR Gram-negative bacteria

Gram-negative bacteria inherently exhibit higher resistance to antibiotics compared to Gram-positive bacteria. This is primarily due to the outer membrane's low permeability and the presence of efflux pumps that expel antibiotics, limiting their entry and reducing intracellular concentrations. Moreover, Gram-negative bacteria, similar to Gram-positive bacteria, can acquire additional resistance mechanisms. *Pseudomonas aeruginosa*, a significant nosocomial pathogen, is particularly prone to developing multidrug resistance through a combination of intrinsic and acquired mechanisms. It poses a major concern for lung infections in cystic fibrosis patients. Other prominent nosocomial pathogens such as *Escherichia coli, Klebsiella pneumoniae, and Acinetobacter baumannii* are known to cause infections in various sites such as the urinary tract, lungs, bloodstream, bones, and joints. These bacteria are increasingly acquiring resistance to important antibiotics like cephalosporins and carbapenems, which are often considered as last-resort treatment options **(Albrecht, 2018).**

Table 1 Clinically important drug-resistant bacteria **(Saga & Yamaguchi, 2009)**.

Microorganisms	Drugs	Resistant bacteria	Mechanism of resistance
Staphylococcus aureus	B-lactam (methicillin)	MRSA	Production of an additional enzyme that avoids drug aureus binding (PBP2)
		VISA	Thickening of cell wall

	Vancomycin	(VISA)	Consequent changes in target
Enterococcus	Vancomycin	VRE	Consequent changes in target
Streptococcus pneumoniae	Penicillin Macrolide	PISP/PRSP Macrolide-resistant *S.pneumoniae*	Mutation in target (PBP) Modification of target (erm) Drug efflux pump (mef)
Haemophilus influenzae	Ampicillin	BLNAR	Mutation in target (PBP)
Pseudomonas aeruginosa	Multiple drugs	MDRP Metallo-B-lactamase-producing bacteria	Multiple factors including loss of porin, drug efflux pump, and drug-modifying enzyme Drug-degrading enzyme
Enterobacteriaceae (e.g., *Escherichia coli*)	B-lactam (carbapenem) Quinolone	ESBL-producing Bacteria Quinolone-resistant E. coli	Drug-degrading enzyme Mutation in target (gyrA, parC)
Gonococci	Quinolone	Quinolone-resistant gonococci	Mutation in target (gyrA, parC)

5. Mechanisms of antibiotic resistance

In order for an antimicrobial agent to effectively act against a specific microorganism, two requirements must be fulfilled. Firstly, the microorganism must possess a critical target that is vulnerable to the antibiotic's action even at low concentrations. Secondly, the antibiotic must be able to permeate the bacterial envelope and reach the target in an adequate amount **(Finch et all., 2010).**

5.1.Target Alteration (Modification)

Resistance to antimicrobials can occur through natural variations or acquired alterations in the target sites of the drugs, which hinder their binding or effectiveness. Such changes in target sites frequently arise from spontaneous mutations in specific bacterial genes located on the chromosome. When these mutations occur and the bacteria are exposed to the antimicrobial, selection pressures can lead to the emergence of resistance. It is important to note that this

mechanism of resistance is observed in both Gram-negative and Gram-positive bacteria **(Abebe et all., 2016).**

5.2.Reduced Permeability

Gram-negative bacteria develop acquired resistance to antibiotics by undergoing significant structural changes in their cell membrane, which result in a decrease in membrane permeability. This reduced permeability is primarily due to the presence of an outer membrane in Gram-negative species, which acts as a barrier to the entry of many antibiotics. In contrast to Gram-positive bacteria, Gram-negative bacteria inherently exhibit lower permeability to a wide range of antibiotics **(Abebe et all., 2016).**

5.3.Antibiotic Inactivation

Certain bacteria have the ability to produce modifying enzymes located either on or near their cell surface. These enzymes specifically target and deactivate the drugs, rendering them ineffective **(Abebe et all., 2016).**

5.4.Efflux pumps

Efflux pumps found in bacteria play an active role in transporting numerous antibiotics out of the cell. These pumps significantly contribute to the inherent resistance of Gram-negative bacteria to various drugs, including those commonly used to treat Gram-positive bacterial infections **(Abebe et all., 2016).**

5.5.Using an alternative metabolic pathway

In contrast to certain alterations in the bacterial target, another mechanism of resistance involves the acquisition of a new pathway that circumvents the need for the drug to exert its effect. This type of resistance has been observed in sulfonamide and trimethoprim. Bacteria can acquire the ability to obtain folate from the environment instead of synthesizing it, thereby bypassing the action of these drugs **(Cesur & Demiroz, 2013).**

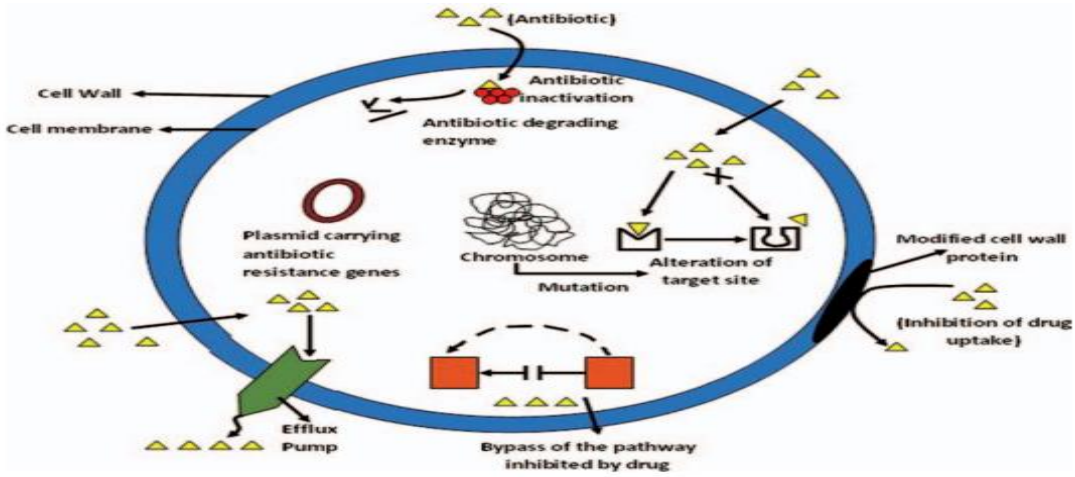


Figure 5 Mechanisms of antibiotic resistance **(Singh, 2014).**

6. Factors Affecting Resistances of Antibiotics

The emergence of antimicrobial resistance (AMR) can be attributed to several factors, which include the following **(Chis et all., 2022):**

- Biological factors, mainly bacterial evolution and genetic mutations, play a significant role.

- The excessive and abusive use of antibiotics is a contributing factor.

- Antibiotics are extensively used in agriculture, including in animal or fish feed, water for infection prevention, and treatment of sick animals.

- Increasing population income directly leads to higher antibiotic consumption and indirectly through the consumption of contaminated meat.

- The ability to travel or transport consumer goods facilitates the spread of microorganisms.

- Insufficient information about antimicrobial resistance (AMR), including statistics on antibiotic consumption, hampers understanding.

- Limited public knowledge about the proper use of antibiotics and the risks of misuse.

- Authorities' lack of adequate measures, such as infection management and ensuring optimal conditions in healthcare facilities, contribute to the problem.

In addition to the factors mentioned above, there are several obstacles that discourage drug manufacturers from investing sufficient funds in developing new antibiotics. This results in a focus on drug classes used for treating chronic diseases. These obstacles include high research and development costs, lengthy authorization processes, the risk of antibiotics quickly becoming ineffective, and strict legislation and price controls **(Chis et all., 2022).**

The transmission of resistant bacteria between individuals and from nonhuman sources, such as food, is another significant factor contributing to the rise of antibiotic resistance. To combat these life-threatening infections, there are four key actions **(Kon & Rai, 2016):**

- ✓ Preventing infections and minimizing the spread of resistance
- ✓ Monitoring the presence of resistant bacteria
- ✓ Enhancing the appropriate use of existing antibiotics
- ✓ Facilitating the development of new antibiotics and the creation of new diagnostic tests for resistant bacteria.

conclusion

At the end of this chapter, we recall that antibiotic resistance is a mechanism developed by bacteria to reduce the effectiveness of a drug in eliminating it, either through natural resistance or through acquired methods of resistance, which led to the development of this into multiple drug resistance. It was necessary to find solutions for this. The problem and one of these solutions is collecting and analyzing data on the effectiveness of antibiotics against bacteria, and this is what will be studied in the next applied chapter.

Chapter 03

Experimental Part

Introduction

Medical decisions related to human life are, the most important factor in helping to quickly find a solution is time. In the modern era, everything depends on technological development and expert systems of artificial intelligence, because that saves time and effort, and there is no room for error in the results in large proportions. Therefore, we are in this The class will use one of the artificial intelligence tools, which is the Weka program, to classify medical data related to antibiotic resistance in several countries around the world.

1. Related work

Weka framework and machine learning techniques have been widely employed in the medical field to improve disease diagnosis, prognosis, treatment selection, and overall patient care. Several noteworthy studies in healthcare have utilized Weka and machine learning, demonstrating their benefits:

- In a study conducted in 2021 by Georgios FERETZAKIS et all., various classifiers such as JRip, Random Forest, MLP, and REP Tree in Weka were used to predict antimicrobial resistance of Acinetobacter Baumannii, Klebsiella Pneumoniae, and Pseudomonas Aeruginosa strains**. (FERETZAKIS et all., 2021)**

- In another study in 2021 by Zhi-Yuan Shi et all., Weka was utilized with classifiers like Decision tree, Simple Logistic, and Multilayer perceptron to apply machine learning techniques for auditing antimicrobial prophylaxis.**(Zhi-Yuan Shi et all. 2021)**

- Çelik Feyza and Karaduman Gül conducted a study in 2022 using Weka's BayesNet, SVM, kNN, J48, and RF algorithms to predict the safe use of antibiotics during pregnancy**. (Çelik & Karaduman, 2022)**

- In 2014, M. Venkat Das et all. used Weka's decision tree induction algorithm to classify subtypes of lung cancer based on biomarkers of non-small cell lung cancers. **(Das et all. 2014)**

- In a 2013 study by David S.K. et all., classification techniques were applied for predicting leukemia disease. The performance of algorithms like K-Nearest Neighbor,

Bayesian Network, Random tree, and J48 tree were compared, with the Bayesian algorithm demonstrating better classification accuracy. (**David et all 2013)**

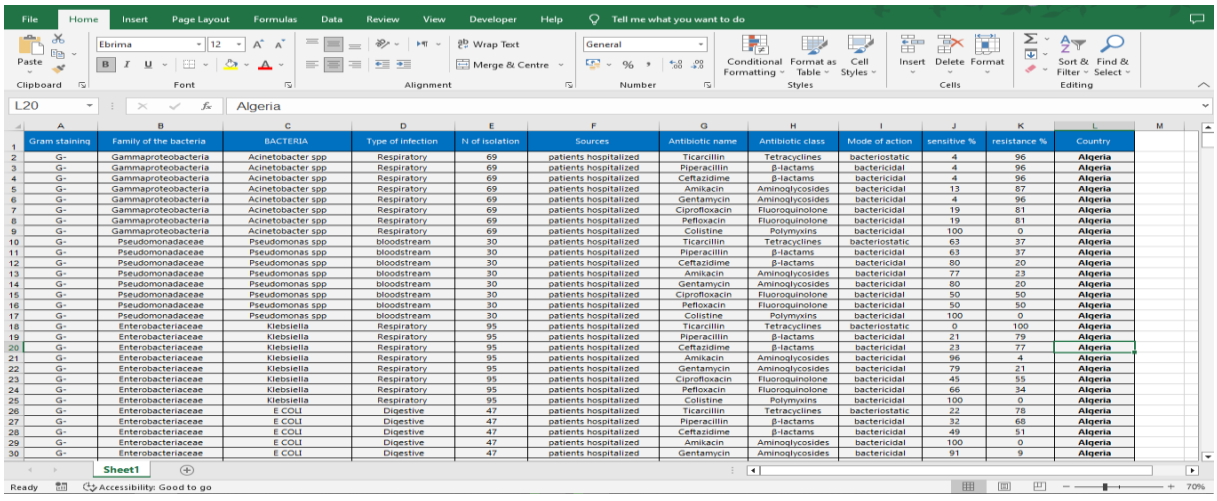
- Vijayarani S. and Sudha S. compared different classification function techniques for heart disease prediction in a 2013 study. They used algorithms such as Logistic, Multilayer Perception, and Sequential Minimal Optimization in Weka (**Vijayarani & Sudha, 2013)**

- In 2011, Yasodha P. and Kannan M. performed an analysis of a diabetic patient database using Weka. They classified the data and compared the outputs using algorithms like Bayes Network, REP Tree, J48, and Random Tree**. (Yasodha & Kannan, 2011)**

2. Materials and Methods

2.1.Dataset used

A dataset refers to a compilation of data, either in the form of a collection of information or a singular statistical data point, wherein each attribute represents a variable, and each instance has its own specific description. In Figure 6, an overview of an antibiotic resistance dataset is presented. The dataset utilized in our study comprises 12 attributes and 632 instances, which are used for classifying and determining the accuracy of antibiotic resistance.



Gram staining	Family of the bacteria	BACTERIA	Type of infection	N of isolation	Sources	Antibiotic name	Antibiotic class	Mode of action	sensitive %	resistance %	Country
G-	Gammaproteobacteria	Acinetobacter spp	Respiratory	69	patients hospitalized	Ticarcillin	Tetracyclines	bacteriostatic	4	96	Algeria
G-	Gammaproteobacteria	Acinetobacter spp	Respiratory	69	patients hospitalized	Piperacillin	β-lactams	bactericidal	4	96	Algeria
G-	Gammaproteobacteria	Acinetobacter spp	Respiratory	69	patients hospitalized	Ceftazidime	β-lactams	bactericidal	4	96	Algeria
G-	Gammaproteobacteria	Acinetobacter spp	Respiratory	69	patients hospitalized	Amikacin	Aminoglycosides	bactericidal	13	87	Algeria
G-	Gammaproteobacteria	Acinetobacter spp	Respiratory	69	patients hospitalized	Gentamycin	Aminoglycosides	bactericidal	4	96	Algeria
G-	Gammaproteobacteria	Acinetobacter spp	Respiratory	69	patients hospitalized	Ciprofloxacin	Fluoroquinolone	bactericidal	19	81	Algeria
G-	Gammaproteobacteria	Acinetobacter spp	Respiratory	69	patients hospitalized	Pefloxacin	Fluoroquinolone	bactericidal	19	81	Algeria
G-	Gammaproteobacteria	Acinetobacter spp	Respiratory	69	patients hospitalized	Colistine	Polymyxins	bactericidal	100	0	Algeria
G-	Pseudomonadaceae	Pseudomonas spp	bloodstream	30	patients hospitalized	Ticarcillin	Tetracyclines	bacteriostatic	63	37	Algeria
G-	Pseudomonadaceae	Pseudomonas spp	bloodstream	30	patients hospitalized	Piperacillin	β-lactams	bactericidal	63	37	Algeria
G-	Pseudomonadaceae	Pseudomonas spp	bloodstream	30	patients hospitalized	Ceftazidime	β-lactams	bactericidal	80	20	Algeria
G-	Pseudomonadaceae	Pseudomonas spp	bloodstream	30	patients hospitalized	Amikacin	Aminoglycosides	bactericidal	77	23	Algeria
G-	Pseudomonadaceae	Pseudomonas spp	bloodstream	30	patients hospitalized	Gentamycin	Aminoglycosides	bactericidal	80	20	Algeria
G-	Pseudomonadaceae	Pseudomonas spp	bloodstream	30	patients hospitalized	Ciprofloxacin	Fluoroquinolone	bactericidal	50	50	Algeria
G-	Pseudomonadaceae	Pseudomonas spp	bloodstream	30	patients hospitalized	Pefloxacin	Fluoroquinolone	bactericidal	50	50	Algeria
G-	Pseudomonadaceae	Pseudomonas spp	bloodstream	30	patients hospitalized	Colistine	Polymyxins	bactericidal	100	0	Algeria
G-	Enterobacteriaceae	Klebsiella	Respiratory	95	patients hospitalized	Ticarcillin	Tetracyclines	bacteriostatic	0	100	Algeria
G-	Enterobacteriaceae	Klebsiella	Respiratory	95	patients hospitalized	Piperacillin	β-lactams	bactericidal	21	79	Algeria
G-	Enterobacteriaceae	Klebsiella	Respiratory	95	patients hospitalized	Ceftazidime	β-lactams	bactericidal	23	77	Algeria
G-	Enterobacteriaceae	Klebsiella	Respiratory	95	patients hospitalized	Amikacin	Aminoglycosides	bactericidal	96	4	Algeria
G-	Enterobacteriaceae	Klebsiella	Respiratory	95	patients hospitalized	Gentamycin	Aminoglycosides	bactericidal	79	21	Algeria
G-	Enterobacteriaceae	Klebsiella	Respiratory	95	patients hospitalized	Ciprofloxacin	Fluoroquinolone	bactericidal	45	55	Algeria
G-	Enterobacteriaceae	Klebsiella	Respiratory	95	patients hospitalized	Pefloxacin	Fluoroquinolone	bactericidal	66	34	Algeria
G-	Enterobacteriaceae	Klebsiella	Respiratory	95	patients hospitalized	Colistine	Polymyxins	bactericidal	100	0	Algeria
G-	Enterobacteriaceae	E COLI	Digestive	47	patients hospitalized	Ticarcillin	Tetracyclines	bacteriostatic	22	78	Algeria
G-	Enterobacteriaceae	E COLI	Digestive	47	patients hospitalized	Piperacillin	β-lactams	bactericidal	32	68	Algeria
G-	Enterobacteriaceae	E COLI	Digestive	47	patients hospitalized	Ceftazidime	β-lactams	bactericidal	49	51	Algeria
G-	Enterobacteriaceae	E COLI	Digestive	47	patients hospitalized	Amikacin	Aminoglycosides	bactericidal	100	0	Algeria
G-	Enterobacteriaceae	E COLI	Digestive	47	patients hospitalized	Gentamycin	Aminoglycosides	bactericidal	91	9	Algeria

Figure 6 Screenshot view of antibiotic resistance Dataset

The dataset includes names of attributes and the explanation of these attributes shown in table 2

Table 2 Features and attributes descriptions of antibiotic resistance datasets

Attribute	Type	Description
Gram stain	Nominal	Gram staining result of the bacteria (Gram-positive, Gram-negative)
Family	Nominal	The family of the bacteria (e.g., Enterobacteriaceae, Staphylococcic)
Bacteria	Nominal	specific bacteria strain or species
Type of Infection	Nominal	The type of infection caused by the bacteria (e.g., Urinary tract infection, Respiratory tract infection)
Number of isolations	Numeric	Number of bacterial isolates
Sources	Nominal	The sources from which the bacteria were isolated (e.g., Hospital environment, Community)
Antibiotic Name	Nominal	The name of the antibiotic (e.g., Amoxicillin, Ciprofloxacin)
Antibiotic Class	Nominal	The class to which the antibiotic belongs (e.g., b-lactam, Fluoroquinolones)
Mode of action	Nominal	The specific mechanism by which the antibiotic targets and kills bacteria (bactericidal, bacteriostatic)
Percentage Sensitive	Numeric	The percentage of bacteria sensitive to the antibiotic
Percentage Resistance	Numeric	The percentage of bacteria resistant to the antibiotic
Country	Nominal	countries mentioned in the study

2.2. Classifiers

Classification is a technique employed to extract models that describe significant data classes or make predictions about future data. This process consists of two steps **(Josh & Shetty, 2015):**

- The learning or training step involves analyzing the data using a classification algorithm.

- The testing step utilizes the data for classification purposes and to assess the accuracy of the classification.

For the prediction of properties in antibiotic molecules, Weka provides several classifiers that can be utilized to generate association rules. In this study, the following classification algorithms were employed:

- Rep tree
- Naïve Bayes
- Decision tree (J48)
- Random tree
- Random forest
- Sequential minimal optimization (SMO)

a. J48 algorithm

J48, a decision tree algorithm, is closely associated with the ID3 algorithm. It is utilized to determine the target value of a new sample based on the various attribute values present in the available data. J48 offers additional features such as accounting for missing values, pruning decision trees, handling continuous attribute value ranges, and deriving rules. In the WEKA data mining tool, J48 is implemented and written in Java. The tool also provides multiple options for tree pruning. In a decision tree, the internal nodes represent different attributes, the branches indicate possible attribute values based on experimental results, and the terminal nodes indicate the final value of the dependent variable. This algorithm generates rules that represent the specific characteristics of the data. The objective is to gradually construct a decision tree that achieves a balance between flexibility and accuracy **(Al-Hatali, 2018).**

b. REPTree algorithm

Rep tree is one of the fast decision tree classifier algorithms. It constructs the decision tree using entropy and information gain of the attribute with reduced error pruning technique. It constructs multiple trees and selects the best tree from the generated list of trees. Rep tree prunes the tree using the back fitting method Rep tree algorithm sorts all numeric fields in the dataset only once at the start, and then it utilizes the sorted list to split the attributes at each tree node. It classifies the numeric attributes by minimising total variance. The non-numeric attributes classified with regular decision tree with reduced error pruning technique **(Hamsagayathri & Sampath, 2017).**

c. Random tree algorithm

Rep tree is a rapid decision tree classifier algorithm known for its speed. It constructs the decision tree by considering the entropy and information gain of the attributes, incorporating reduced error pruning. Multiple trees are built, and the best tree is selected from the generated set. The pruning of the tree is performed using the backfitting method. Unlike other algorithms, Rep tree sorts the numeric fields in the dataset only once at the beginning and utilizes the sorted list for attribute splitting at each node. Numeric attributes are classified by minimizing the total variance, while non-numeric attributes are classified using a regular decision tree with reduced error pruning technique **(Hamsagayathri & Sampath, 2017).**

d. Random forest algorithm

Random Forest is recognized as a highly accurate machine learning algorithm. It possesses the ability to handle a vast number of attributes without requiring feature selection. Additionally, it provides estimates of the significant attributes in the dataset. The algorithm demonstrates remarkable efficiency in estimating missing data while maintaining high accuracy. Random Forest is well-suited for processing large volumes of databases. Multiple trees are constructed, and the best tree is chosen based on the splitting criteria. Unlike Rep tree, Random Forest does not involve error pruning **(Hamsagayathri & Sampath, 2017).**

e. Naive Bayes algorithm

The Naive Bayes classifier is a Bayesian supervised classifier that assumes that all attributes are independent of each other given the output class. This assumption is quite strong and can lead to numerical instabilities and poor results when the probabilities approach 0 or 1 using a frequentist approach **(Khanwalkar & Soni, 2020).**

f. Sequential minimal optimization (SMO) algorithm

The Sequential Minimal Optimization (SMO) algorithm is employed to solve the quadratic programming (QP) problem encountered in training support-vector machines (SVM). It is a widely utilized tool for SVM training, support, and is commonly introduced through the LIBSVM tool. The introduction of the SMO algorithm in 1998 generated significant anticipation within the SVM community, as it offered a more straightforward and cost-effective approach compared to the complex and resource-intensive methods previously employed, which often required the use of third-party QP solvers **(Riad, 2021).**

The selection of the most suitable algorithm depends on specific characteristics of the dataset, such as the nature of the features and the size of the dataset. Additionally, the desired performance criteria for the classification task also play a role in determining the appropriate algorithm.

The study implemented in WEKA and dataset of bacterial resistance to antibiotics have been used.

2.3.WEKA (version 3.8.6)

The Weka, also known as the woodhen (Gallirallus australis), is a native bird species found exclusively in New Zealand. However, in the context of machine learning, Weka refers to the Waikato Environment for Knowledge Analysis. Developed at the University of Waikato, New Zealand, Weka is a widely used suite of machine learning software written in Java. It encompasses a range of algorithms and visualization tools for data analysis and predictive modeling. Weka offers user-friendly graphical interfaces, making its functionality easily accessible. It is an open-source platform-independent tool, freely available for use in data mining and machine learning applications.

The graphical user interface (GUI) Chooser, as depicted in the figure, includes four buttons **(Singhal & Jena, 2013):**

- **Explorer**: Provides a platform for data exploration within WEKA.
- **Experimenter:** Facilitates the execution of experiments and statistical tests for comparing learning schemes.
- **Knowledge Flow:** Offers similar features as the Explorer but with a drag-and-drop interface, providing the added advantage of supporting incremental learning.
- **Simple CLI:** Provides a straightforward command-line interface for executing WEKA commands, particularly useful for operating systems lacking their own command-line interface. This Java-based version of WEKA (Weka 3) finds applications in various domains, especially in education and research. WEKA offers several advantages.

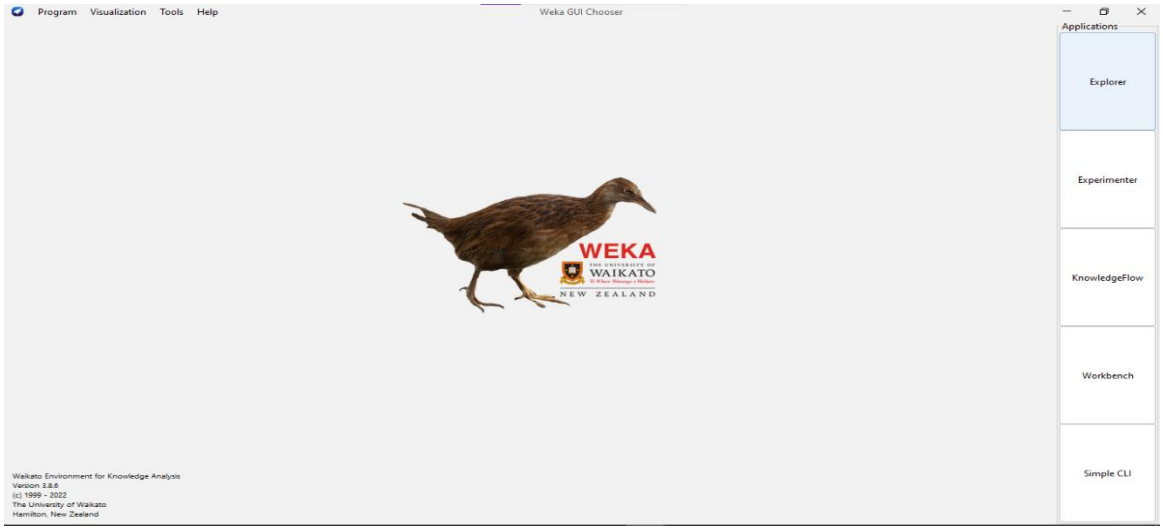


Figure 7 The Weka explorer

- Weka is available for free under the GNU General Public License.
- It is highly portable as it is implemented entirely in Java, enabling it to run on various architectures.
- Weka offers a wide range of data preprocessing and modeling techniques.
- Its user-friendly graphical interface makes it easy to use.

- Weka supports numerous standard data mining tasks, including data preprocessing, clustering, classification, regression, visualization, and feature selection.

2.4.Data Preparation and Processing

The databases that are commonly used tend to be prone to issues such as noisy, missing, and inconsistent data. This is primarily due to their large size, complexity, and the fact that they often originate from multiple heterogeneous sources. To address these problems, the target dataset selected during the selection step undergoes preprocessing **(Gupta, 2011).**

- The data is initially stored in Excel format and then converted into the Comma Separated Values (CSV) file format.

- The converted information is subsequently uploaded to the WEKA Application, where the raw data is preprocessed into a more comprehensible file format.

3. Results and discussion

3.1.Pre-processing on WEKA Explorer

The preprocessing of the antibiotic classification dataset is performed using the WEKA Explorer. The Explorer interface provides an overview of the dataset, including the current relation of attributes, the number of instances (632), and the number of attributes (12). To preprocess the dataset, a filter is applied to all attributes related to antibiotic resistance. Specifically, the "supervised->attributes->Add Classification" filter is chosen and applied to the attributes. Additionally, the preprocess tab allows for the visualization of all attributes after applying the filter.

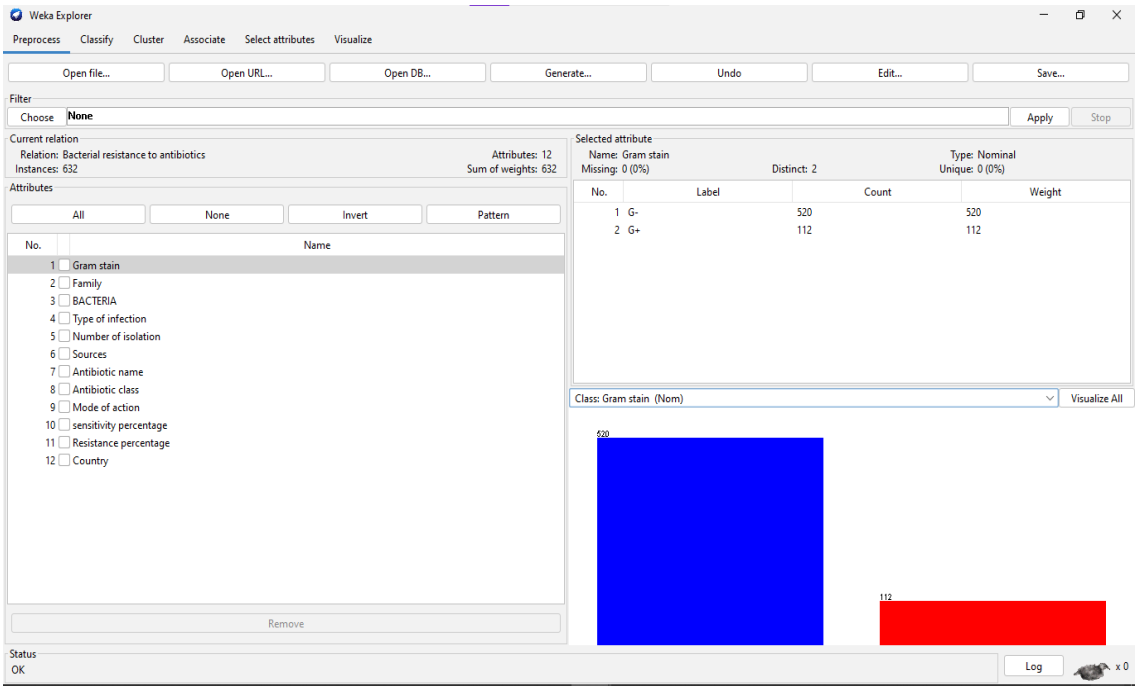


Figure 8 Data pre-processing of antibiotic resistance dataset

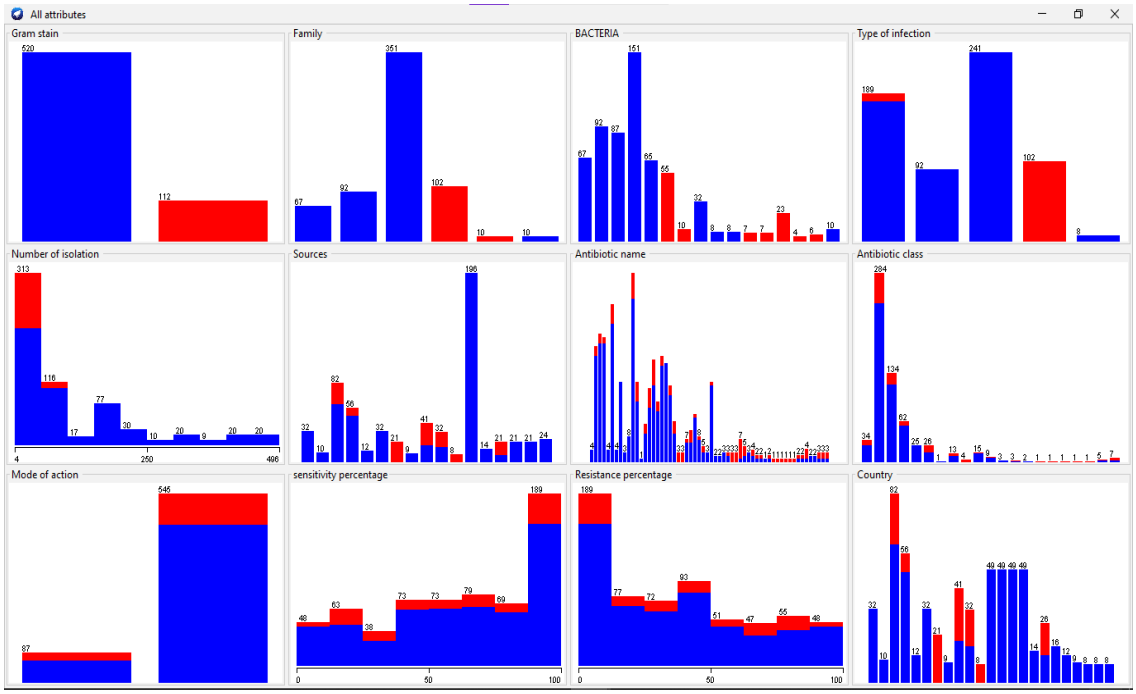


Figure 9 Visualization of the attributes of antibiotic resistance dataset.

The visualizations depicted in the aforementioned figures display the attribute names and their corresponding values based on the dataset. The number of graphs represents the number of

categories within the attributes, while the variation in values corresponds to the number of instances in each respective category.

3.2.Classify attributes on WEKA Explorer

The antibiotic resistance datasets are subjected to classification using six different classifier algorithms: Naïve Bayes, J48, SMO, Random Forest, Random Tree, and Rep Tree. To evaluate the effectiveness of these classification strategies in predicting class labels and determining accuracy, the algorithms are applied to the dataset using a combination of stratified 10-fold testing and a 66/33 data split. This comprehensive approach enables a thorough assessment of the performance of the classification algorithms.

The performance of the discussed classifiers is evaluated using the following parameters **(kumar & Sahoo 2012) :**

- **Kappa statistic:** The Kappa statistic measures the agreement between two sets of categorized data. It ranges from 0 to 1, where higher values indicate stronger agreement. Values between 0.40 and 0.59 are considered moderate, 0.60 to 0.79 are substantial, and above 0.80 are outstanding.

- **Mean Absolute Error (MAE):** MAE calculates the average of absolute errors between predicted and actual values. It measures how close the predicted values are to the actual values.

- **Root Mean Square Error (RMSE):** RMSE is the square root of the average of squared errors between predicted and actual values. It quantifies the differences between predicted and observed values. Smaller values of RMSE indicate better accuracy.

- **ROC Area:** ROC (Receiver Operating Characteristic) area provides a comparison between predicted and actual target values in a classification task. It evaluates the performance of a model across various classification thresholds. The ROC area ranges from 0 to 1, where a higher value indicates better model performance

The performance of the classifiers is also evaluated using additional parameters, including **(Feretzakis et all..2020):**

- **MMC (Matthews Correlation Coefficient):** MMC is a correlation coefficient derived from the four values of the confusion matrix. It measures the quality of the classification results.
- **True Positive Rate (TP Rate):** TP Rate represents the percentage of instances that are correctly classified as belonging to a specific class.
- **False Positive Rate (FP Rate):** FP Rate represents the percentage of instances that are incorrectly classified as belonging to a specific class.
- **Precision:** Precision calculates the percentage of instances that truly belong to a particular class divided by the total number of instances classified as belonging to that class.
- **Recall:** Recall calculates the percentage of instances classified as a particular class divided by the actual total number of instances.
- **F-Measure:** The F-Measure is a broad measure of the model's quality, taking into account both precision and recall.

On the first part, correctly and incorrectly classified Instances will be partitioned in numeric and percentage value and Kappa statistic. In the next part, mean absolute error, root mean squared error, relative absolute error will be consider as parameters for evaluation. The results of the simulation are shown in Tables 3 and 4 below. Table 3 mainly summarizes the result based on accuracy and time taken for each simulation. Table 4 shows the result based on error during the simulation.

Table 3 Performance accuracy

Algorithms	Correctly Classified Instances % (value)	Incorrectly Classified Instances % (value)	Time taken to build model (second)	Kappa statistic
J 48	98.73	1.27	0.07	0.9864
Random tree	55.85	44.15	0.05	0.5258
Random Forest	65.51	34.49	0.73	0.6304
Rep tree	97.31	2.69	0.04	0.9712
SMO	69.78	30.22	3.64	0.6762
Naïve Bayes	68.99	31.01	0.03	0.668

From table 3, it can be stated that the percentage of correctly classified instances of J48 classifier (98.73 % i.e. 624 instances) is comparatively larger than other mentioned classifier and it is approximately closer to the Rep tree classifier (97.31 % i.e. 615 instances). As a result, the percentage of incorrectly classified instances of J48 (1.27 % i.e. 8 instances) is comparatively less than the others classifier where Random tree gives highest incorrect classification instances (44.15 % i.e. 279 instances). Here, the total number of instances is 632. naïve bayes classifier takes less time to build model.

Table 4 MSE, RMSE, RAE calculation

Classifier	Mean absolute error	Root mean squared error	Relative absolute error %
J48	0.0015	0.0285	1.7892 %
Random tree	0.0395	0.1902	48.6049 %
Random Forest	0.0403	0.1322	49.5518 %
Rep tree	0.9712	0.0354	2.8904 %
SMO	0.0796	0.1964	97.9357 %
Naïve Bayes	0.0303	0.1321	37.262 %

From table 4, it can be seen that Rep tree has largest value of MAE (0.9712). Random tree and Naïve Bayes have comparatively less MAE and approximately equal value (0.0395 and 0.0303). SMO has large percentage of relative absolute error (97.9357%).

3.3. Performance Parameters

To evaluate the experimental results, the proposed work utilizes standard information retrieval (IR) performance measures. These measures include accuracy, precision, recall, F-measure, and ROC area. These performance measures are calculated for each of the six classification algorithms.

Table 5 TP, FP, Precision and Recall calculation

Classifier	TP RATE	FP RATE	PRECISION	RECALL
J48	0.987	0.000	0.989	0.986
Random tree	0.513	0.036	0.530	0.513
Random Forest	0.663	0.024	0.666	0.663
Rep tree	0.973	0.001	0.975	0.972
SMO	0.698	0.020	0.689	0.689
Naïve Bayes	0.690	0.022	0.685	0.690

From table 5, it can be seen that J48 has highest True Positive rate (0.987) and Random tree has lowest True Positive rate (0.513). Also, J48 and Rep tree have good Precision value (0.989 and 0.975 respectively). Random tree has largest False positive rate (0.036). Table 6 illustrates that J48 has highest MCC, on the contrary Random tree has low MCC value. J48 and Rep tree covers the same highest percentage of ROC area (99.8%).

Table 6 F-measurement, MCC and ROC AREA calculation

Classifier	F-MEASURE	MCC	ROC AREA
J48	0.986	0.986	0.998
Random tree	0.516	0.484	0.745
Random forest	0.664	0.640	0.956
Rep tree	0.972	0.972	0.998
SMO	0.692	0.673	0.957
Naïve bayes	0.680	0.662	0.950

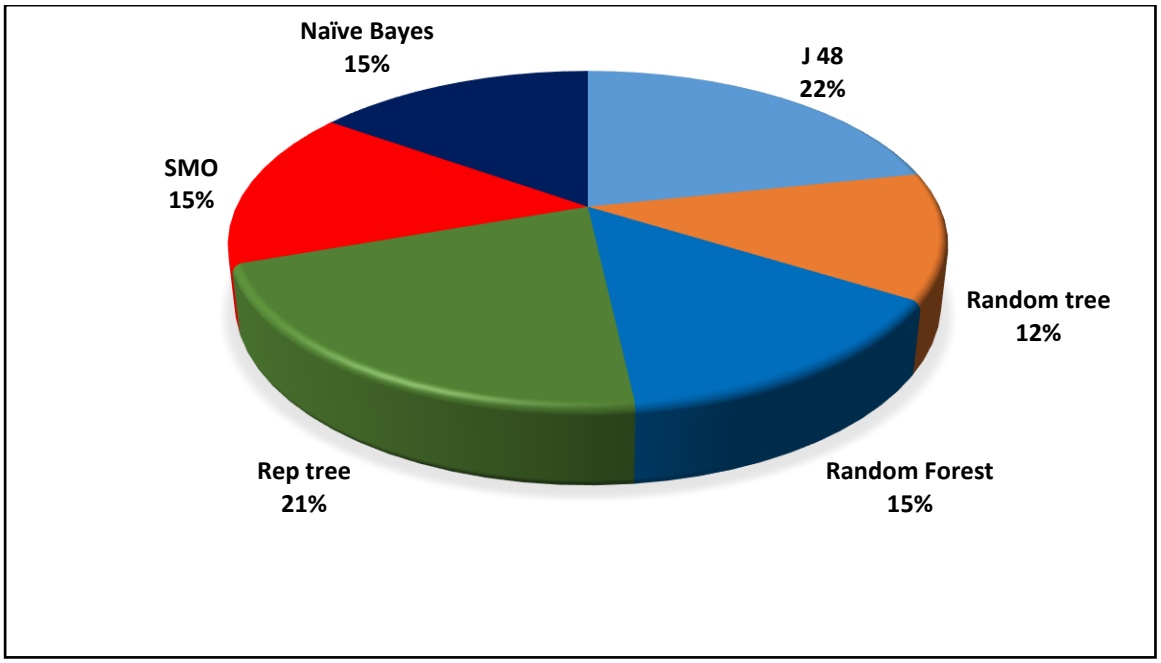


Figure 10 Accuracy on antibiotic resistance

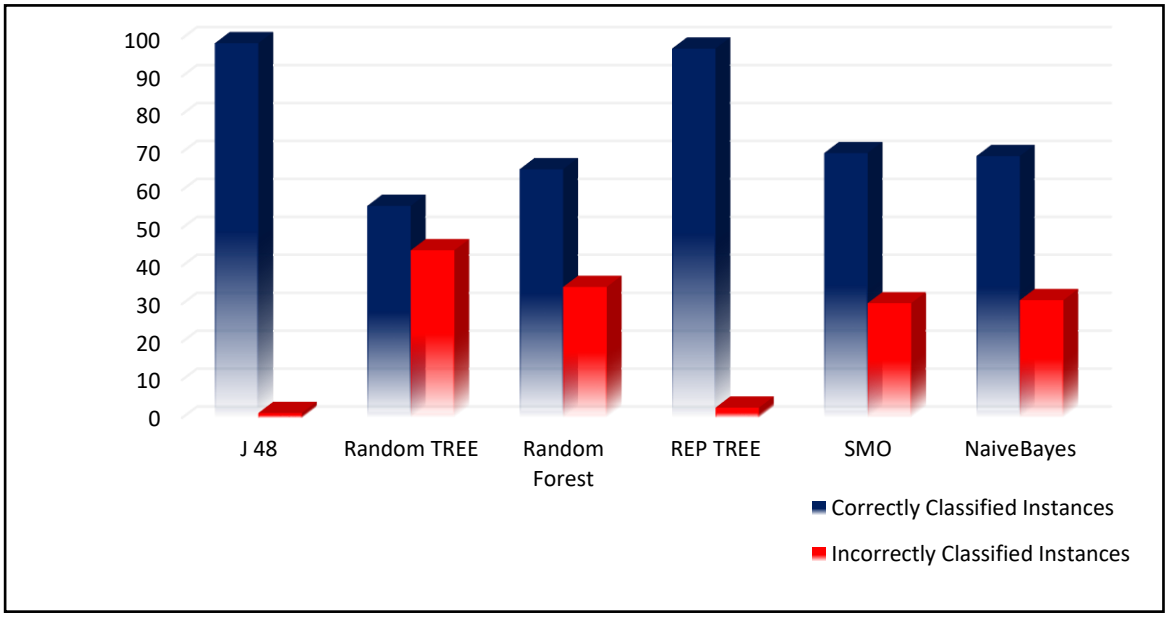


Figure 11 Performance comparison of algorithms

Figure 10 shows the accuracy on antibiotic resistance.

Figure 11 describes classifiers performance based on correctly and incorrectly classified instances. The highest percentage of correctly classified instances is observed with J48 classifier, the lowest percentage amounts to Random tree classifier.

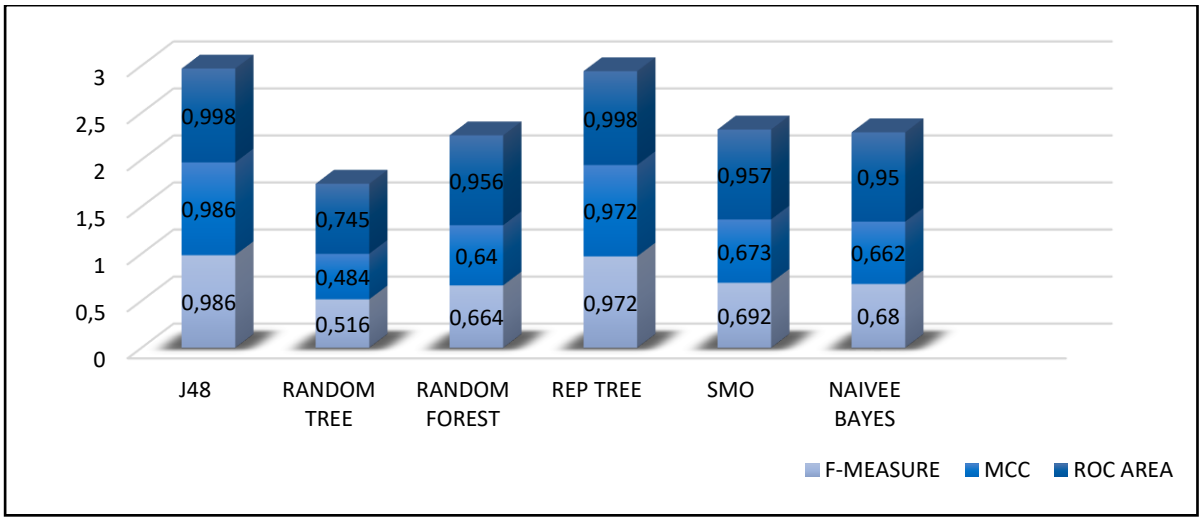


Figure 12 F- measurement, MCC and ROC AREA

Figure 12 shows the classifiers performance comparison based on MCC, ROC AREA and F-measure. F-measure of all algorithms is between 0, 98 and 0, 51. The good f-measure is observed with J48 classifier.

Conclusion

At the end of this applied chapter, and as we mentioned in the introduction to the chapter, time is one of the most important factors that help reduce the risk of any health threat to humanity. In this third chapter, we used six algorithms for classification and prediction, and each algorithm has a specific mechanism for analyzing and classifying data. Results in a record time that only took a few minutes and record time, j48 in our study was the best algorithm, and this does not necessarily have to be the best, but according to the data provided, the best way to obtain effective results with a rate of more than 95% was the j48 algorithm in terms of correct classification And the accuracy of the classification, and through the most important factor for evaluating the performance of the algorithms, which is the roc area, where the percentage of performance in general was 99.8%.

General conclusion

Weka and machine learning are powerful tools for antibiotic classification that can help researchers identify new antibiotics and optimize existing ones. Machine learning algorithms can learn from data and make predictions or decisions without being explicitly programmed, making them ideal for antibiotic classification. With the increasing threat of antibiotic resistance, the use of Weka and machine learning for antibiotic research is becoming increasingly important.

In this research work, the frequently used classification techniques J48, Random Tree , Random Forest , REP Tree , SMO, and Naive Bayes are analyzed and J48 performed the highest accuracy This study also observed Kappa statistic, Mean absolute error, F measure, MCC, ROC Area, Relative absolute error, FP rate, TP rate, Root mean squared error, Precision, Recall and It has also been comprised that J48 classifier has the highest percentage of ROC Area .

Based on the analysis of the parameters considered, the study showed that J48 outperformed the other techniques in classifying antibiotic resistance data. However, further improvements can be made by expanding the dataset with a larger number of attributes and implementing advanced feature selection methods. The accurate classification of antibiotic is crucial in developing targeted treatment strategies and addressing the emerging problem of antibiotic resistance.

To combat antibiotic resistance, there are several strategies that can be employed. One approach is to reduce the use of antibiotics in both humans and animals. This can be achieved by improving diagnostic tests, promoting better hygiene practices, and developing alternative therapies. Another strategy is to develop new antibiotics that are effective against resistant bacteria. This can be achieved by investing in research and development, as well as collaborating with industry and academic partners. Finally, it is important to educate healthcare professionals and the general public about the risks of antibiotic resistance, and how to use antibiotics appropriately.

References

Abebe E; Tegegne B (2016): A Review on Molecular Mechanisms of Bacterial Resistance to Antibiotics, Journal of Applied Sciences, Vol8, no5, p 301-310.

Abushaheena M; uzaheeda. (2020): Antimicrobial resistance, mechanisms and its clinical significance, DM Journal, VOL 66, NO 6, P 4.

Albrecht, L.M. (2018): Antibiotic Resistance: Selection in the Presence of Metals and Antimicrobials. P-P14-15.

Al-Hatali A. (2018): A Comparative Study of the Efficient Data Mining Algorithm to Find the Most Influenced Factor on Price Variation in Oman Fish Markets. Sakarya university journal of computer and information sciences, Vol 1, no 2, P4.

Al-Thani R; Al-Ali F. (2012): Incidences and antimicrobial susceptibility profile of Staphylococcus species isolated from animals in different Qatari farms, African Journal of Microbiology Research, Vol 6 , NO 48, P 7456.

Benko T; Boldizar M. (2015): Incidence of bacterial pathogens in equine uterine swabs, their antibiotic resistance patterns, and selected reproductive indices in English thoroughbred mares during the foal heat cycle, Veterinarni Medicina, VOL 11, P 617.

Bywater R ; Deluyker H .(2004): A European survey of antimicrobial susceptibility among zoonotic and commensal bacteria isolated from food-producing animals, Journal of Antimicrobial Chemotherapy, VOL 54, P 747-749.

Carlos F; Cuevas A. (2016): Antibiotics and Antibiotic Resistance in the Environment , CRC Press ,Mexico, ISBN 9781315679419, P12.

Çelik F ; Karaduman G (2022), In silico QSAR modeling to predict the safe use of antibiotics during pregnancy, taylor & francis, doi.org/10.1080/01480545.2022.2113888.

Cesur S ; Demiroz A .(2013): ANTIBIOTICS AND THE MECHANISMS OF RESISTANCE TO ANTIBIOTICS, Medical Journal of Islamic World Academy of Sciences ,Vol 21, no 4.

Chinedum I. (2005): Microbial resistance to antibiotics, African Journal of Biotechnology Vol 4, no 13, p 1606-1611.

References

Chis A ; Rus L .(2022): Microbial Resistance to Antibiotics and Effective Antibiotherapy, Biomedicines MDPI, Vol10, no 1121, P6.

Chmiel R; Marek A. (2022): Antibiotic Resistance in Bacteria, MDPI , Vol 11, no1079, P2.

Chopra I. (2002): Antibiotics Encyclopedia of Life Science, Natural Publishing Group, Vol 26, P6.

Dass et all, (2014): Classification of Lung cancer subtypes by Data Mining technique, IEEE, DOI: 10.1109/CIEC.2014.6959151

David S, e all (2013) Comparative Analysis of Data Mining Tools and Classification Techniques using WEKA in Medical Bioinformatics, Computer Engineering and Intelligent Systems, 4(13):28-38,2013.

Do Nascimento J. (2015): Resist: an intelligent system to predict antibiotic resistance, Dissertation master, Specialization in Information Systems, Department of Informatics, university of Lisbon, Portugal.

Etebu E; Arikekpar I, (2016): Antibiotics: Classification and mechanisms of action with emphasis on molecular perspectives, International Journal of Applied Microbiology Biotechnology Research, Vol 4, no 6, P97.

Exner M ; Bhattacharya S ; Christiansen B . (2017): Antibiotic resistance: What is so special about multidrug-resistant Gram-negative bacteria, GMS Hygiene and Infection Control, Vol 12, P2.

Fahmey S. (2013): Early-onset sepsis in a neonatal intensive care unit in Beni Suef, Egypt: bacterial isolates and antibiotic resistance pattern, Korean J Pediatr, Vol 56. no 8, p 235.

Feretzakis et all, (2020): Using Machine Learning Techniques to Aid Empirical Antibiotic Therapy Decisions in the Intensive Care Unit of a General Hospital in Greece Antibiotics, MDPI, vol 9, no 50

Finch et all, (2010): Antibiotic and Chemotherapy.Anti-infective agents and their use in therapy.9th ed. British. saunders elsevier, ISBN: 978-0-7020-4064-1, P25.

Finch, R; Greenwood D. (2010): Antibiotic and Chemotherapy: Anti-infective agents and their use in therapy, the British Library, UK, ISBN: 978-0-7020-4064-1, P 16.

References

Fomnya H; Ngulde S (2021): Antibiotics: Classifications and mechanism of resistance, International Journal of Applied Microbiology Biotechnology Research, Vol 9, no 3.

GALES AC; CASTANHEIRA M. (2012): Antimicrobial Resistance among Gram-Negative Bacilli Isolated from Latin America: Results from SENTRY Antimicrobial Surveillance Program (2008-2010), JMI Laboratories, North Liberty, IA, USA, p 1186.

Gallagher J; MacDougall C. (2014): Antibiotics Simplified, Jones & Bartlett Learning, USA, ISBN 978-1-284-02539-2, P 7.

Gallagher J; Macdougall C. (2023): Antibiotics simplified, 5th ed, JONES & BARTLETT L E A R N I N G , Burlington, P38.

Geta K; Kibret M. (2022): Antibiotic Resistance Profiles of Bacteria Isolated from Hotspot Environments in Bahir Dar City, Northwestern Ethiopia, Journal of Multidisciplinary Healthcare, VOL 15, P 1407.

Gupta S. (2011): Performance analysis of various data mining classification techniques on healthcare data. International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 4.p160.

Hamsagayathri P; Sampath P. (2017): Performance analysis of breast cancer classification using decision tree classifiers, International Journal of Current Pharmaceutical Research, Vol 9, no 2, p 20-21.

Hannachi A; Bentchouala C. (2016): Multidrug-resistant bacteria isolated from patients hospitalized in Intensive Care Unit in University Hospital of Constantine, Algeria (2011 - 2015), African Journal of Microbiology Research, Vol 10, no 33, p 1333.

Hasan H; Al-Harmoosh R (2020): Mechanisms of Antibiotics Resistance in Bacteria, Systematic Reviews in Pharmacy, Vol 11, no 6, P817 823.

Hashm M. (2020) : Antibiotics and Antimicrobial Resistance Genes in the Environment , Elsevier, British, ISBN 978-0-12-818882-8.

Hayder T; ALASEDI K. (2021): A Comparative Study of Prevalence Antimicrobials Resistance Klebsiella pneumoniae among Different Pathogenic Bacteria Isolated from Patients with Urinary Tract Infection in Al-Najaf City, Iraq, Latin American Journal of Pharmacy, special NO 40, P 176.

References

Hemmati H; Khoshdoz S. (2020): Two years study of prevalence and antibiotic resistance pattern of Gram-negative bacteria isolated from surgical site infections in the North of Iran, BMC Research Notes, Vol 13 no 383, P 4.

Josh S; Shetty P . (2015): Performance Analysis of Different Classification Methods in Data Mining for Diabetes Dataset Using WEKA Tool, International Journal on Recent and Innovation Trends in Computing and Communication. Vol 3, No 3, P1168.

Kaufman G. (2011): Antibiotics: mode of action and mechanisms of resistance, Nursing standard (Royal College of Nursing (Great Britain), Vol 25, no 42, P51.

Kaye K; Pogue J. (2015): Infections Caused by Resistant Gram-Negative Bacteria: Epidemiology and Management, Pharmacotherapy, Vol35, no10, p 949–962.

Khanwalkar A; Soni R. (2020): Sequential Minimal Optimization for Predicting Diabetes at its Early Stage, Journal of critical reviews, Vol 8, no1, p 974.

Kon, K; Rai, M. (2016): Antibiotic Resistance. Mechanisms and New Antimicrobial Approaches. British. Elsevier. ISBN: 978-0-12-803642-6.

kumar.Y ; G. Sahoo .(2012): Analysis of Parametric & Non Parametric Classifiers for Classification Technique using WEKA . I.J. Information Technology and Computer Science, no, 7, vol 43-49 .p47.

Kunjir et all, (2017): Data Mining and Visualization for Prediction of Multiple Diseases in Healthcare, International Conference On Big Data Analytics and computational Intelligence (ICBDACI), IEEE, doi 978-1-5090-6399-4.

Kuznetsova M; Gizatullina J. (2020): Escherichia coli Isolated from Cases of Colibacillosis in Russian Poultry Farms (Perm Krai): Sensitivity to Antibiotics and Bacteriocins, MDPI, VOL 8 NO741, P 6.

Manikandan C; Amsath A. (2013): Antibiotic susceptibility of bacterial strains isolated from wound infection patients in Pattukkottai, Tamilnadu, India, International Journal of Current Microbiology and Applied Sciences, Vol 2 no 6, P198.

Neu H; Gootz, T. (2001): Antimicrobial Chemotherapy. In: Baron, S. (eds.)., “Medical Microbiology”. 5th ed. Galvestone. The University of Texax Medical Branch.

References

Osagie E; Olalekan S. (2019): Multiple Drug Resistance: A Fast-Growing Threat, Biomedical Journal of Scientific & Technical Research, Vol 21, no2. P15716.

Riad R. (2021): Analysis of WEKA data mining algorithms Bayes net, random forest, MLP and SMO for heart disease prediction system: A case study in Iraq International Journal of Electrical and Computer Engineering (IJECE), Vol 11, No 6 .P5232.

Saga T: Yamaguchi k. (2009): History of Antimicrobial Agents and Resistant Bacteria, the Journal of the Japan Medical Association, Vol 52, No 2, P 106.

Sánchez S; Demain A. (2015): Antibiotics Current: Innovations and Future Trends, Caister Academic Press, Norfolk, Britch, ISBN: 978-1-908230-55-3, P2.

Singhal S; Jena M. (2013): A Study on WEKA Tool for Data Preprocessing, Classification and Clustering, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol 2, P 251.

Singhal S; Jena M. (2013): A Study on WEKA Tool for Data Preprocessing, Classification and Clustering, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol 2, P 251.

Stachelek M. (2021): Overcoming bacterial resistance to antibiotics, sciendo, Vol 21, No1, P1.

Tessema B; Lippmann N. (2021): Antibiotic Resistance Patterns of Bacterial Isolates from Neonatal Sepsis Patients at University Hospital of Leipzig, Germany, MDPI, Vol 10 no 323, P12-13.

Ultra. Fresh. (2019): online. Bactericidal vs Bacteriostatic: What's the Difference. https://www.ultra-fresh.com/bactericidal-vs-bacteriostatic-what-is-the-difference/

Vijayarani S, Sudha S, (2013) Comparative Analysis of Classification Function Techniques for Heart Disease Prediction, International Journal of Innovative Research in Computer and Communication Engineering, 1(3): 735-741.

Walsh C. (2003): Antibiotics: Actions, origins, resistance, Elsevier Science, USA, ISBN 1-55581-254-6, P4.

Walsh C; Wencewicz T. (2016): Antibiotics: challenges, mechanisms, opportunities Washington DC, ASM Press, ISBN978-1555819309. P30.

References

World Health Organization (2022): AWaRe (Access, Watch, Reserve) antibiotic book, electronic version, ISBN 978-92-4-006238-2, P12.

World Health Organization's 2014 "Antimicrobial resistance - global report on surveillance2014" – http://www.who.int/drugresistance/documents/surveillancereport/en/ (Accessed online: November 2014)

Yasodha P, Kannan M, (2011) Analysis of Population of Diabetic Patient Database in WEKA Tool, International Journal of Science and Engineering Research, 2 (5), May

Yuan et all (2021) : Applying Machin Learning Techniques to the Audit of Antimicrobial Prophylaxis, MDPI,vol 12, no 2586.

Yuksel S; Ozturk B. (2006): Antibiotic resistance of urinary tract pathogens and evaluation of empirical treatment in Turkish children with urinary tract infections, International Journal of Antimicrobial Agents, Vol 28, p 414.

Zhang C; Liang Y. (2022): Distribution of bacterial keratitis and emerging resistance to antibiotics in China from 2001 to2004, Taylors Francis Online, Vol 2 no 3, p 578.