Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018)
IEEE Xplore Compliant Part Number: CFP18K74-ART; ISBN:978-1-5386-2842-3

Prediction of Antimicrobial Resistance for disease-causing agents using Machine Learning

Using Machine Learning algorithms to predict individuals susceptibility of developing AMR for drugs

Srajan Kulshrestha
Department of Computer Engineering
International Institute of Information Technology
Pune, India
srajan.kulshreshtha@gmail.com

Dinesh Nayar
Department of Computer Engineering
International Institute of Information Technology
Pune, India
dineshnyr@gmail.com

Sanjana Panda
Department of Computer Engineering
International Institute of Information Technology
Pune, India
sanjana.96panda@gmail.com

Dr. Vaishali Dohe
Associate Prof. Department of Microbiology
B. J. Medical College
Pune, India

Ashwini Jarali
Asst. Prof. Department of Computer Engineering
International Institute of Information Technology
Pune, India

***Abstract*—Antimicrobial resistance (AMR) occurs when disease-causing microorganisms are resistant towards prescribed drugs, nullifying its effect. As a consequence, there is a delay in recovery which worsens the patient's health. Antimicrobial resistance is identified as a global threat by the medical fraternity and various government bodies.**

Objective of the proposed system is to integrate technology with the field of bio-medical, in context with AMR. We applied various machine learning algorithms on datasets, to identify patterns and use them to predict resistance towards various drugs. This model would help in closing the gap between Doctors and Labs.

In this model, we used ML and data mining techniques to predict AMR for individual patients based on trends identified from datasets. For building the model we use results of Patients undergoing antibiotic susceptibility test as datasets.

***Keywords*—Machine Learning, Classification, Decision Tree, Association Rule, Apriori Algorithm, Data Mining, Pathogens, Drugs, Combination Therapy, Antimicrobial Resistance, CLSI Guidelines, Staphylococcus aureus (sau), Pseudomonas aeruginosa elastase (pae).**

I. INTRODUCTION

Pathogens are disease-causing microorganisms. An antimicrobial is that agent that prevents the multiplication effect of pathogens. Antimicrobial resistance (AMR) is a Darwinian selection process that pathogens adapt to survive. Pathogens show resistant towards the drugs they are exposed to. Hence, these drugs have no effect on the patient, delaying the healing process. Sometimes, this may be fatal to the patient's life. In a recent report, it is said that an estimated population equivalent to 10 million people will perish by 2050 because of AMR. [1]

One of the major triggers for AMR is the use, misuse, or overuse of antimicrobial drugs. Antimicrobial resistance has become a serious global threat. World Health Organization (WHO) and governments of various countries have also identified this threat.

Knowledge engineering methods can prove to become a powerful tool in finding unexpected patterns and hidden knowledge, and establishing new rules from large datasets.

There are antibiotic susceptibility tests that determine AMR for patients, but these tests take time to deliver results.

Doctors constantly deal with impatient Patients, who do not wait for elementary diagnostic results, they want instant results. Due to which there is a gap between doctors and clinics. Our project aims to bridge this gap, by creating a tool that predicts AMR in Patient. In this model, we will use ML algorithms to classify drugs for individual patients, into two groups Resistant and Sensitive. A Resistant classified drug means the patient has grown resistive towards the drug. Similarly, a Sensitive classified drug implies that the patient is responding towards the drug. We would be applying classification algorithms on the datasets, to predict resistance towards certain drugs.

II. LITERATURE SURVEY

Antibiotics have been used since the 1940s. Since then, deaths from several infections and illness have been significantly reduced. Various reasons contribute to the growing Antibiotic resistance. Over-prescription of antibiotics is a problem that has to be tackled [2]. 2010 study results declared India as the world's biggest consumer of antibiotics for human health at 10.7 units per person [3]. Patients often do not complete their entire antibiotic course which allows the strongest bacteria to survive [4]. A survey conducted in 2015 by WHO which involved multiple countries suggested that there was a ubiquitous public misconception about antibiotic usage and resistance. The results indicated that 42% do not know that they should stop taking antibiotics only when they complete the dosage as administered [5]. Once a patient has grown resistant towards a drug, there is a transfer of resistant determinants between microorganisms. This brings a change in the genome sequence of the patient. Interaction between humans, animals and agricultural host create a platform where resistant genes can be transferred thus habilitating the spread of resistance [6]. Antimicrobials are heavily used in animal food production industry for disease prevention, treatment, and growth promotion. But the large-scale use of antimicrobials in agriculture (livestock and fish farming) results in human exposure to antimicrobial-resistant bacteria via direct and indirect pathways. Poor infection prevention, control practices and unsanitary conditions in healthcare facilities are also responsible to further spread and increase of antimicrobial resistance. Also, as new, the rate at which bacteria are getting resistant to existing medicines is a lot faster than newer antibiotics are being developed [7]. Incorrect and excessive use of antibiotics, as well as poor infection control, has boosted antibiotic resistance. With proper steps, society can reduce and limit the spread of resistance. The World Health Organization has created various guidelines to help organizations tackle AMR [8]. To tackle the challenge of resistance and infections, antibiotic stewardship and hospital infection control have been deployed worldwide [9]. Persistent reconnaissance of local antimicrobial susceptibility patterns is a must for fighting rising antimicrobial resistance. WHONET is a compelling computerized microbiology research facility information administration and examination program that can give direction for empiric treatment of contaminations, alarm clinicians of patterns of antimicrobial resistance, direct drug-policy choices and preventive measures. The program encourages sharing of information among different hospitals by keeping a common format which can be collaborated for global or national antibiotic resistance surveillance [10]. In a paper published by the University of Athens proposed a framework in which data produced by various hospitals were integrated into a data warehouse and data mining approaches like Apriori algorithm was used to detect hidden and previously unknown patterns on large datasets [11]. A paper published in PLOS used decisions trees to find special relationships among variables and were used to establish new rules from datasets [12].

Association rule learning is knowledge engineering method which uses rules to find interesting relations between different variables in large datasets. Apriori is an algorithm for frequent itemset mining and association rule learning over large databases [13]. Support says how popular an item set is in the datasets. Confidence says how likely item A is purchased when item B is purchased. Lift says how likely item A is purchased when item B is purchased [14].

III. PROPOSED SYSTEM

We aim to make a model that takes patient details and predicts whether the person is sensitive or resistant to first line of drugs of treatment. In an ideal system, the model needs to know vast patient history for training and accurate prediction of future instances but in India there is no system for keeping a track of patients, therefore we use results of antibiotic susceptibility test as datasets. To increase the efficiency, association rules were discovered among various drugs for individual micro-organisms. And these patterns were used to predict results for further drugs.

This section provides the description of components of proposed model, as illustrated in “Fig. 1,” (A) Data Cleaning and Transformation, (B) Association Rules Generation, (C) Features Selection, (D) Model, (E) User Interface.

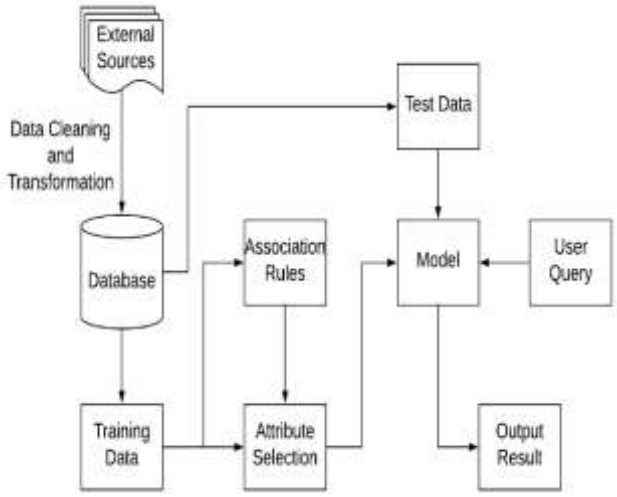

Fig. 1 Architecture Diagram

A. *Data Cleaning and Transformation*

Data from Antibiotic Susceptibility Test from various Patients were used. Data available was highly noisy data, hence cleaning was required. Mainly the data indicates sensitivity or resistivity of drugs. Various mathematical formulas were used with reference to CLSI guidelines to fill those empty data cells. Thereafter a transformed and cleaned data was created on which all further processing was carried out.

B. Generation of Association Rules

Patterns were generated amongst different drugs for individual micro-organisms using Apriori Algorithm. Hidden patterns among drugs were discovered. By this thousand of association rules were generated; filtering of patterns was done using support, confidence and lift as qualifying metrics.

C. Features Selection

Antecedents of identified association rules were used as attributes. In the system antecedents were the features and consequent was our target class. The qualifying metrics depends on the organism. Hence, for each organism a different Decision Tree was to be built.

D. Model

System works on data tailoring, association generation, and decision tree. Role of former two is explained now comes the decision tree.

The system was trained with data to build a decision tree classifier; it worked on entropy as split criteria. This trained model is now used to predict the target class for an unseen instance.

E. User Interface

Our system was highly client interactive in nature, end users being medical practitioner.

An interface is provided to feed details of any patient to the system such as his age, gender, site of infection, any prior known resistant drugs. These details are used as input to our model which in turn classify first line of drug for his/her treatment as resistant or sensitive as output on our robust dashboard.

Fig 2. Doctor Dashboard UI

IV. Mathematical Model

Decision tree algorithm works by recursive partitioning of data set into subsets. Each node of the tree is given a particular set of record T that is split by a specific test on feature.

Attributes are categorical in nature. Antibiotics can be split according to its nature. An antibiotic can belong to various subcategories i.e. Resistant, Sensitive, Intermediate or Not Determined. To split into entropy is used as criteria.

Let us consider following set of tuples:

S = {D, X, Y, F}

Where,

D = datasets

X = {basic patient information, site of infection, patient medical record}

Y = {pool of resistant drugs or sensitive drugs}

F = {data cleaning, apriori algorithm, decision classifier}

'D' is the training datasets with only essential features. 'X' is the input filled by the medical practitioner. 'Y' is the end result for any patient about resistivity or sensitivity of him to first line of drugs. 'F' is the various function used in the implementation of the system.

V. Results

Table I.

Organism	Features	Target Drug	Accuracy (%)
Sau	Penicillin, Clindamycin, Cefoxitin	Erythromycin	91.67
Sau	Clindamycin, Cefoxitin	Erythromycin	91
Sau	Clindamycin, Cefoxitin	Penicillin	85
Sau	Erythromycin, Cefoxitin	Penicillin	95.83
Pae	Aztreonam, Ceftazidime	Cefepime	96
Pae	Amikacin, Ceftazidime	Cefepime	92
Pae	Amikacin, Imipenem	Ceftazidime	76

As you can observe from the table, the average accuracy is around 85-90% for few microorganisms, the accuracy can be increased with more personalized Patient details.

VI. Conclusion

We believe that prediction of AMR can be a vital step to fighting AMR. It can also act as a tool to prevent AMR.

As stated earlier for building an ideal model to predict the resistivity or sensitivity of any drug in an individual requires vast data with minute details is needed. Such datasets require organizations to invest more in technology and build systems for the same.

Feedback loop can be used to increase discover new patterns and keep the system updated. As the patterns for AMR keeps changing a feedback loop will keep the model efficient.

VII. Acknowledgment

Our work wouldn't have been possible without the datasets is sourced from Microbiology Department, B. J. Medical College, Pune. We are deeply grateful of Doctor Renu Bharadwaj (HOD Microbiology Dept.) to believe in us and

give us suggestions and constant guidance with various aspects of AMR.

VIII. References

[1] Marlieke E. A. de Kraker, Andrew J. Stewardson, and Stephan Harbarth, "Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050?" *ncbi.nlm.nih.gov*, Nov 13, 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5127510/. [Accessed: Aug. 2, 2017]

[2] Andrew Duong and Michelle Jaelin, "6 Factors That Have Caused Antibiotic Resistance," *infection-control.tips*, Nov 18, 2015. [Online]. Available: https://infectioncontrol.tips/2015/11/18/6-factors-that-have-caused-antibiotic-resistance/. [Accessed: Aug. 10, 2017].

[3] Ramanan Laxminarayan and Ranjit Roy Chaudhury, "Antibiotic Resistance in India: Drivers and Opportunities for Action," *journals.plos.org*, March 2, 2016. [Online]. Available: http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001974. [Accessed:20 Aug, 2017]

[4] "CDC: 1 in 3 antibiotic prescriptions unnecessary," May 3, 2016. [Online]. Available: https://www.cdc.gov/media/releases/2016/p0503-unnecessary-prescriptions.html. [Accessed: Aug 30, 2017]

[5] WHO, "Combating Antimicrobial Resistance in India," April 30, 2014. [Online]. Available: http://www.searo.who.int/india/topics/antimicrobial_resistance/Combating_Antimicrobial_Resistance_in_India/en/. [Accessed: Sept. 10, 2017]

[6] WHO, "The world health report 2007 - A safer future: global public health security in the 21st century," *who.int*, 2007. [Online]. Available: http://www.who.int/whr/2007/en/. [Accessed: Sept. 20, 2017]

[7] Charles H. Brower, Siddhartha Mandal, Shivdeep Hayer, Mandeep Sran, Asima Zehra, Sunny J. Patel, Ravneet Kaur, Leena Chatterjee, Savita Mishra, B.R. Das, Parminder Singh, Randhir Singh, J.P.S. Gill, and Ramanan Laxminarayan, "The Prevalence of Extended-Spectrum Beta-Lactamase-Producing Multidrug-Resistant Escherichia coli in Poultry Chickens and Variation According to Farming Practices in Punjab, India," *ehp.niehs.nih.gov*, July 2017. [Online].Available:https://ehp.niehs.nih.gov/ehp292/. [Accessed: Sept. 30, 2017]

[8] WHO, "Antibiotic resistance," *who.int*, Feb 5, 2018. [Online]. Available: http://www.who.int/en/news-room/fact-sheets/detail/antibiotic-resistance. [Accessed: Oct. 10, 2017]

[9] Sujith J. Chandy, Joy Sarojini Michael, Balaji Veeraraghavan, O.C. Abraham, Sagar S. Bachhav, and Nilima A. Kshirsagar, "ICMR programme on Antibiotic Stewardship, Prevention of Infection & Control (ASPIC)," *ncbi.nlm.nih.gov*, Feb 2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4001333/. [Accessed: Oct 20, 2017].

[10] A Agarwal, K Kapila, and S Kumar, "WHONET Software for the Surveillance of Antimicrobial Susceptibility" *ncbi.nlm.nih.gov*, Jul 21, 2011. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4921382/. [Accessed: Oct 30, 2017].

[11] Eugenia G. Giannopoulou, Vasileios P. Kemerlis, and Michalis Polemis, "A Large Scale Data Mining Approach to Antibiotic Resistance Surveillance" Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07), June 15, 2007.

[12] Joana Rosado Coelho, João André Carriço, Daniel Knight, Jose-Luis Martínez, Ian Morrissey, Marco Rinaldo Oggioni and Ana Teresa Freitas, "The Use of Machine Learning Methodologies to Analyse Antibiotic and Biocide Susceptibility in Staphylococcus aureus" *journals.plos.org*, Feb. 19, 2013. [Online]. Available: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0055582#s2. [Accessed: Nov 10, 2017]

[13] Rakesh Agrawal and Ramakrishnan Srikant, "Fast algorithms for mining association rules." Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487-499, Santiago, Chile, September 1994.

[14] Annalyn Ng, "Association Rules and the Apriori Algorithm: A Tutorial," *kdnuggets.com*, April 2016. [Online]. Available: https://www.kdnuggets.com/2016/04/association-rules-apriori-algorithm-tutorial.html. [Accessed: Nov 20, 2017