Constructing *Mycobacterium tuberculosis* Drug Resistance Prediction Model

Po-han Li, Yu-Sheng Ting, and Kuan-Hao Chao

Abstract—The worldwide spread of antimicrobial resistance (AMR) bacteria is a serious global issue. Among them, Tuberculosis (TB) caused by *Mycobacterium tuberculosis* is the major global health problem caused from a single pathogen which listed in the 10 causes of death from 2000 to 2016. Constructing TB antidrug resistance prediction model is crucial because drug susceptibility testing in lab is time-consuming. Furthermore, it could help doctor in early diagnosis. However, current prediction sensitivity and specificity of Isoniazid(INH), Rifampicin(RMP) and Ethambutol(EMB) TB drugs are not good enough. To address the need, we utilize protein information of each assembled *Mycobacterium tuberculosis* genome in PATRIC database to develop the state-of-the-art drug resistance model with machine learning approach. Our results show that with enough amount of data, the trained model with specific hyperparameters can outperform the previous studies, which may make the future diagnosis much faster.

Index Terms—*Mycobacterium tuberculosis*, machine learning, prediction model

1 Introduction

THE worldwide spread of antimicrobial resistance (AMR) bacteria turns out untreatable infections and humanity health crisis. It is a serious global problem which is not limited to national borders and can affect peoples around the world without discrimination [1], [2], [3]. The World Health Organization (WHO) has stated the threat to the ever-increasing AMR bacterial infections in a 2014 global report [4]. AMR not only jeopardize human health but also has a negative impact on the global economy. A 2009 European Union (EU)-based research about AMR identified that multidrug resistance caused 400,000 infections and 25,000 deaths each year; with regard to economic, it cost 900 million euros and 600 million days of lost productivity annually [5], [6]. Another European Union (EU)-based research made an estimation model based on collected EARS-Net data. The model reckoned that 671,689 infections and 33,110 deaths in EU, 2015 [7]. In the United States, 2013, it is estimated that about 2 million peoples acquire severe infections with AMR bacteria, and at least 23,000 people died each year; the economic cost ranges from $20 billion to 35 billion per year [8]. The instances above show the severity of AMR problem in recent years.

Among all these AMR bacteria, Tuberculosis (TB) caused by *Mycobacterium tuberculosis* is one of the major global health problem which attracts world's close attention [9], [10]. It is the leading cause from a single pathogen and listed in the top 10 causes of death from 2000 to 2016, which ranks higher than HIV/AIDS [10], [11]. Several meetings and conferences are held and the goal, eliminating TB epidemic by 2030, is set by all member of WHO and UN [10]. Although the global death percentage and the ranking of TB drop annually, there are still millions of people acquire TB disease each year. In 2017, it is estimated that 10.0 million infections and 1.6 million deaths worldwide [11]. The first modern TB chemotheraphy began in 1952. Afterwards, strains of *Mycobacterium tuberculosis* started to acquire resistance to various drugs gradually [12]. Due to the severe harmful impact that TB has brought, it is critical to establish a systematic treatment on TB and multidrug resistant TB (MDR-TB).

A TB treatment guidline book published by WHO clearly describes the disease symptom and recommended treatment [13]. The drug for TB Disease treatment is cautiously chosen in stepwise selection process considering efficacy, drug-resistance, patient-safty and cost [13], [14]. There are more than 20 drugs developed to for TB treatment. Among them, five first-line drugs, including Isoniazid(INH), Rifampin(RMP), Pyrazinamide(PZA), Ethambutol(EMB) and Streptomycin(STM), are used for the treatment of new patients. The daily or 3 times per week recommended dose of these drugs are avaialable in the book and WHO official website [13]. If patients have been previously affected and the result of drug susceptibility testing is resistant to at least Isoniazid(INH) and Rifampicin(RMP), then second-line drugs should be used for treatment instead. Second-line drugs, including aminoglycosides(WHO group 2), fluoroquinolones(WHO group 3), thioamides(WHO group 4), are used for MDR-TB treatment. Drugs that may be useful but have unproven efficacy are classified as third-line drugs [13], [15]. Of all these drugs, first-line drugs are the most important and widely-used.

The process of phenotypic drug-susceptibility testing is slow and expensive [16]. It is still difficult to identify resistant strains quickly and choose the appropriate second-line drugs because bacterial culture in lab takes time. Traditional diagnosis of tuberculosis takes about four to eight weeks from sample culturing to antibiotic susceptibility testing [17], [18]. Although the culturing time has been reduced into 72 hours [19], it is still time-consuming to get the drug resistance result. Therefore, constructing a precise

• *Y. Ting, P. Li, and K. Chao was with the Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan. E-mail: b05901026@ntu.edu.tw, b05901094@ntu.edu.tw, and b05901180@ntu.edu.tw*

Manuscript received June 25, 2019; revised June 25, 2019.

Mycobacterium tuberculosis drug resistance prediction model is crucial. The prediction model could help doctor makeing early diagnosis.

Current TB drug resistance prediction model mainly uses database comparison approach [20], [21], [22]. However, among the first-line drugs, the sensitivity and specificity of Pyrazinamide(PZA), Ethambutol(EMB) are not good enough to make the prediction. Moreover, the sample number for building the model is limited in previous work, which are 307 in PZA and 484 in EMB [22]. Some researches tried machine learning approach based on DNA sequencing data and have improved the result of Isoniazid(INH), Rifampicin(RMP) and Ethambutol(EMB) [23]. However, the result of PZA is still inaccurate and the method of constructing prediction model using protein data has not been tried.

To address the need, we utilized the *Mycobacterium tuberculosis* feature file in PATRIC database and try to imporve the prediction rate by machine learning approach.

2 METHODS

2.1 Database Selection

According to a review paper published in 2010 [24], there are many TB databases avaialable on the Internet. However, most database has limited data source and the format and file structure between these databases are mostly incompatible. Therefore, it makes data preprocessing a labor-intensive job. In this study we choose one of the largest bacterial database, PATRIC [25], as our raw data source. PATIRC, the Pathosystems Resource Integration Center, is a bacterial database providing integrated data and analysis tools. It welcomes clinical and biomedical researchers around the world to upload their assembled TB genome and antidrug resistance test result. It also provides FTP functionality, thus making it easy to access these great amount of data programmatically.

2.2 Data Preprocessing

We use PATRIC.features.tab gene annotation file provided by PATRIC database to build our resistance prediction machine learning model. There are many features including self-defined gene id, gene feature(CDS, tRNA, pseudogene), chromosome location, +/- strand, protein id etc. We select protein id as our main feature. The data have been organized into two types of tables, genome table and protein table. The identity of each genome is recorded in the first column of the genome table, and each boolean variable recorded in the second column indicates whether the according genome is resistant or susceptible to the target drug. A protein table of each genome records the numbers of the gene sequence fragment of the according proteins.

2.3 Machine Learning Models

Different models of different hyperparameters but of similiar structures are used to predict the resistance of a genome. All the models take a 9,071-dimension vector as the input, which is the protein table of the genome, and output a 2-dimension vector, each dimension in which represents the predicted probability of the resistance and susceptance respectively. One of the model structures is shown in Fig. 1. All the models have softmax function as the activation function of the last layer, ReLU function as the activation function of one of the layers, and use Adagrad as the optimizer, cross-entropy as the loss function.

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 100)	907200
dense_2 (Dense)	(None, 1000)	101000
dense_3 (Dense)	(None, 2000)	2002000
dense_4 (Dense)	(None, 2)	4002

Fig. 1. Model of four layers

The first comparison is of the difference among the number of layers of the models. The depth of the network of the models vary from 2 to 6. The second comparison is of the difference value of the learning rates. The learning rates of the models are 0.01, 0.001, 0.0001, 0.00001, 0.000001, respectively, and all of them are based on the 6-layer structure. The last comparsion is of the difference of the layer using ReLU function as the activation function, and all of them are based on the 4-layer structure.

2.4 Validation Split

To evaluate the accuracy of the prediction of the model in the training process, ten-percent of the data are used as a validation set, which are not used as the training data. The validation split is chosen so that the traing set and the validation set has the same resistance to susceptible ratio as the origin set.

2.5 Evaluation

The output of the model is a 2-dimension vector, the first dimension in which indicates the probability of the predicted resistance. A threshold is setted that if the probability is higher than the threshold value, then the genome would be classified as having drug resistance, and as having no drug resistance if otherwise. The threshold can be adjusted to obtain different sensitivity and specificity. Having different points of different threshold, the ROC curve (receiver operating characteristic curve) could be plotted, as shown in Fig.2. In this experiment, the threshold changes from 0% to 100%, and each time increases 5%. Connecting the 21 points with the origin and point (1,1), a ROC curve with 23 points can be obtained. It should be noticed that given the ROC curve and the resistance to susceptible ratio, TP, TN, FP, and FN can all be obtained by simple calculation.

AUC (Area under the Curve of ROC) is equal to the probability that the classifier will rank a randomly chosen instance higher than a randomly chosen negative one, which is used to evaluate the accuracy of the prediction of the model. The larger the AUC is, the more accurately the model can predict whether the genome has drug resistance.

3 RESULTS

In the task of predicting the resistance to PZA, in which there are 3,295 drug-susceptible genome and 378 drug-resistant genome as training data, the model with 4-layer

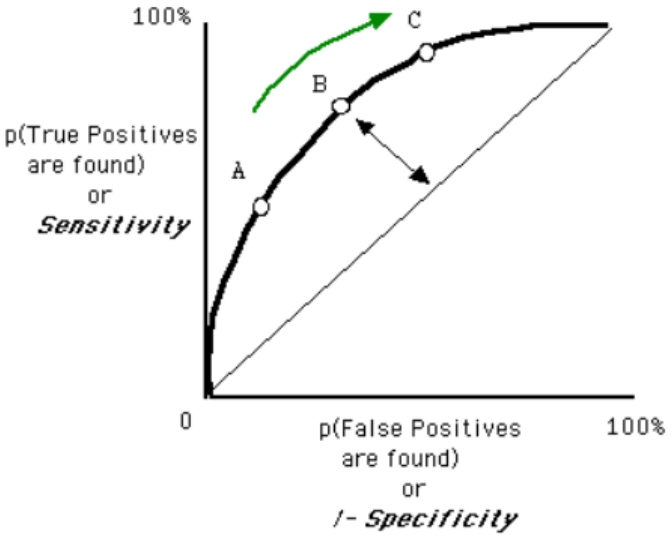

Fig. 2. As the threshold adjusted, the point moves in the ROC space.

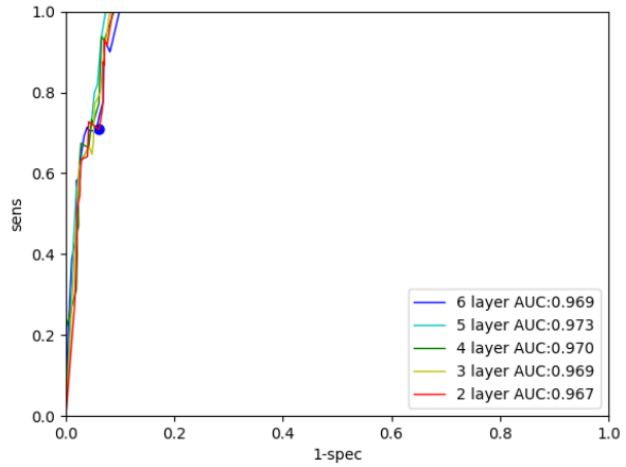

Fig. 3. Changing the depth of the model.

and 0.0001 as the learning rate works better after completing 500 epochs of training. The AUC of the model is 0.970.

3.1 Change in Depth

With the learning rate of 0.0001, 5 ROC curves can be obtained after completing 500 epochs as shown in Fig.3 by changing the depth of the models from 2 to 6. It can be noticed that all the models outperform the result of the previous study [26], which has the sensitivity of 70.9% and the specificity of 93.9%, as presented as the blue point.

3.2 Change in Learning Rate

With the 6-layer structure and the ReLU function as the activation function of the fifth layer, 5 ROC curves can be obtained after completing 500 epochs as shown in Fig.4 by setting the learning rate as 0.01, 0.001, 0.0001, 0.00001, 0.000001, respectively. It can be noticed that only the model trained with the learning rate of 0.0001 outperforms the

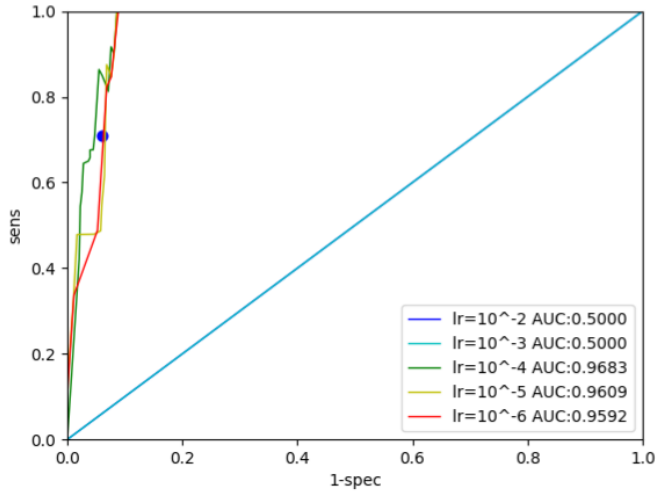

Fig. 4. Changing the learning rate of the model.

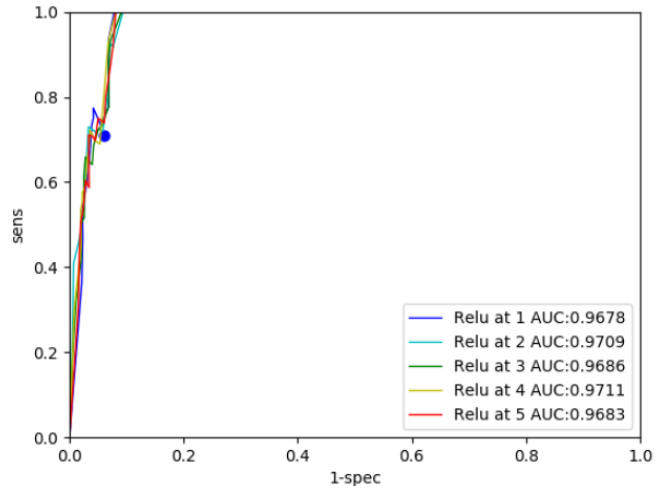

Fig. 5. Changing the position of the ReLU activation function.

result of the previous study [26], which has the sensitivity of 70.9% and the specificity of 93.9%, as presented as the blue point.

3.3 Change in the Layer Using ReLU Function as the Activation Function

With the 4-layer structure and the learning rate of 0.0001, 5 ROC curves can be obtained after completing 500 epochs as shown in Fig.5 by changing the layer using ReLU activation function from the first layer to the fifth layer. It can be noticed that all the models outperform the result of the previous study [26], which has the sensitivity of 70.9% and the specificity of 93.9%, as presented as the blue point.

3.4 The Result for EMB

The similiar experiment is done on EMB, in which there are 4,808 training data in total. With the learning rate of 0.0001, 5 ROC curves can be obtained after completing 500 epochs as shown in Fig.6 by changing the depth of the models from 2 to 6. It can be noticed that all the models except the one

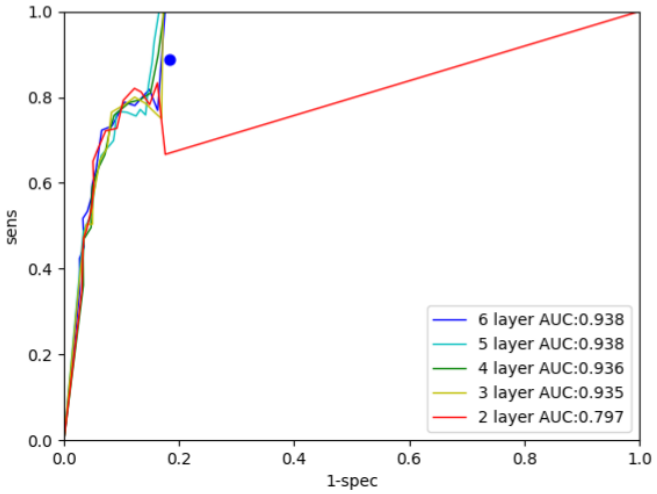

Fig. 6. Changing the depth of the model predicting the drug resistance to EMB.

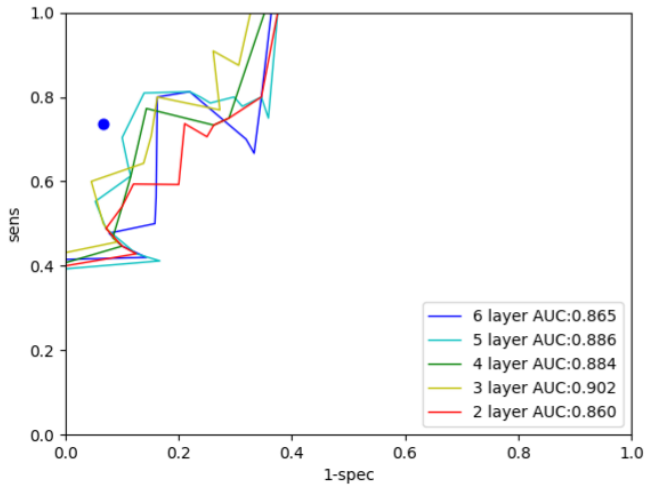

Fig. 7. Changing the depth of the model predicting the drug resistance to ETH.

with 2 layers outperform the result of the previous study [26], which has the sensitivity of 88.7% and the specificity of 81.7%, as presented as the blue point.

3.5 The Result for ETH

The similiar experiment is done on ETH, in which there are 571 training data in total. With the learning rate of 0.0001, 5 ROC curves can be obtained after completing 500 epochs as shown in Fig.7 by changing the depth of the models from 2 to 6. It can be noticed that none of the models outperforms the result of the previous study [26], which has the sensitivity of 73.6% and the specificity of 93.3%, as presented as the blue point.

3.5.1 Changing the Validation Percentage

The amount of the data is too small to train the network in the case of ETH. Thus, to have more training data, the validation set has been reduce to 5% of the total data, and the data used to train the model increase to 95% of the total

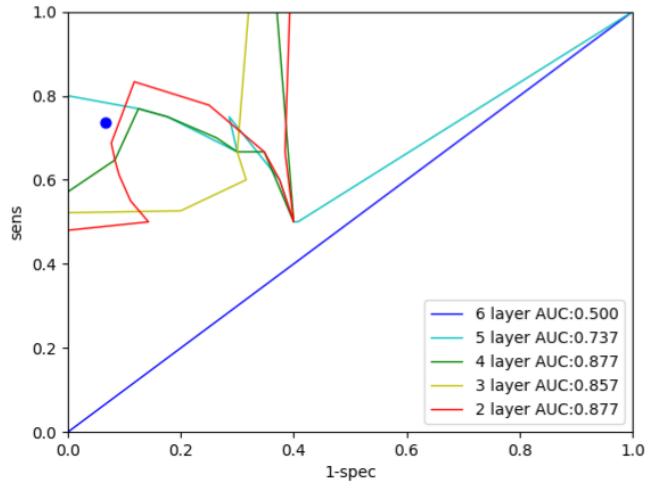

Fig. 8. Changing the validation split percentage for the training on ETH

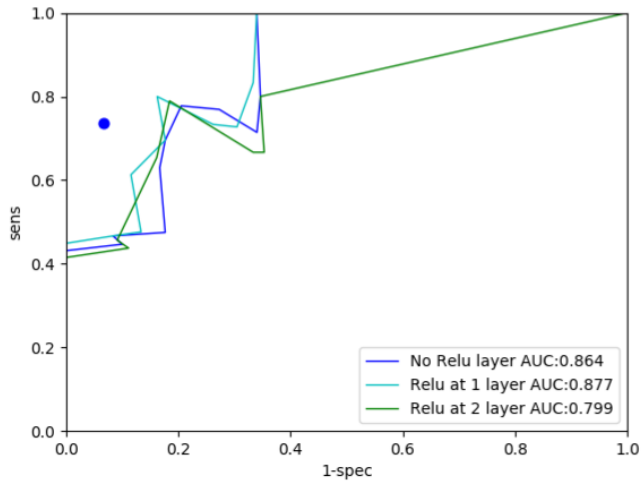

Fig. 9. Changing the model generalization ability for the training on ETH

data. 5 ROC curves can be obtained after completing 500 epochs as shown in Fig.8 by changing the depth of the models from 2 to 6. It can be noticed that there is some improvement because one of the model outperforms the result of the previous study [26], which has the sensitivity of 73.6% and the specificity of 93.3%, as presented as the blue point.

3.5.2 Changing the Model Generalization Ability

Since using ReLU as the activation function can make the model have the better generalization ability [27], the linear activation function is replaced by the ReLU function in different layers. With the 3-layer structure and the learning rate of 0.0001, 3 ROC curves can be obtained after completing 500 epochs as shown in Fig.9 by using linear function in all layers and using ReLU activation function in the first and the second layer, respectively. It can be noticed that still none of the models outperforms the result of the previous study [26], which has the sensitivity of 73.6% and the specificity of 93.3%, as presented as the blue point.

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 100)	561500
dense_2 (Dense)	(None, 100)	10100
dense_3 (Dense)	(None, 200)	20200
dense_4 (Dense)	(None, 300)	60300
dense_5 (Dense)	(None, 400)	120400
dense_6 (Dense)	(None, 500)	200500
dense_7 (Dense)	(None, 400)	200400
dense_8 (Dense)	(None, 300)	120300
dense_9 (Dense)	(None, 200)	60200
dense_10 (Dense)	(None, 100)	20100
dense_11 (Dense)	(None, 2)	202

Fig. 10. 11-layer model used for the training on ETH

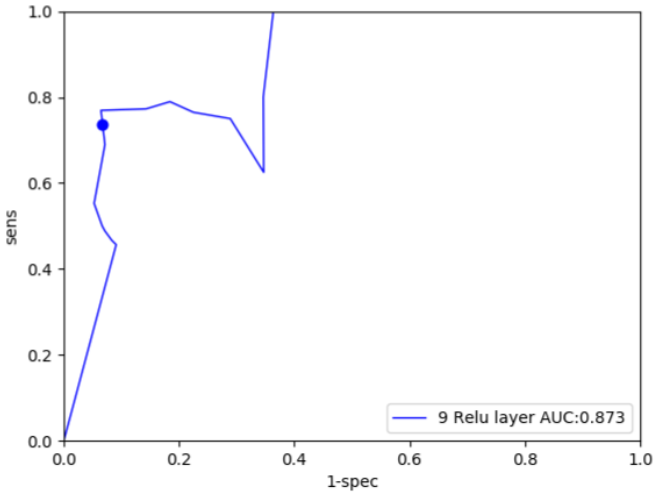

Fig. 11. Using the 11-layer model to predict the drug resistance to ETH

3.5.3 Very deep model on ETH

Since the deeper network has the better generalization ability [28], the 11-layer model shown in Fig.10 is adopted. With the ReLU function used as the activation function in the ninth layer, the ROC curve can be obtained after completing 5,000 epochs as shown in Fig.11. It can be noticed that the model is almost as accurate as the result of the previous study [26], which has the sensitivity of 73.6% and the specificity of 93.3%, as presented as the blue point.

4 Discussion

4.1 The Deeper, the Better?

When evaluating the ability of fitting curves of a model, we say the deeper the model is, the better it is. The reason is that theoretically, when a model gets deeper, the set of functions it can fit expands exponentially. However, the experiment shows that the model with five layers is better than that with six layers. It should be noticed that with the same width, deeper model means more parameters, which could make it take more time to fit the training data. If the target function of predicting resistance can be fit by two models, the deeper model needs more time to fit the function. Thus, as both models completing more than 500 epochs, it is possible that the 5-layer model is better than the 6-layer model.

4.2 Learning Rate

If the learning rate is too large, it is likely that the updated parameters during the traning process keeps skipping the local minimum. For example, in the training of PZA, the AUC of the learning rate of both 10^{-2} and 10^{-3} are 0.5, as shown in Fig.4, which indicates the models work as well as randomly guessing. Thus, it can be verified that too large learning rates may actually make the model learn nothing.

However, if the learning rate is too small, the training could be very slow, which can also be verified in the training of PZA by noticing that the AUC of the learning rate of 10^{-4} is larger than that of 10^{-5} and that the AUC of the learning rate of 10^{-5} is larger than that of 10^{-6}.

Therefore, the appropriate learning rate should be used to train the model.

4.3 The Position of ReLU

In the prediction of the drug resistance based on the number of specific proteins, the effect may be signifcant only when the number of the proteins exceeds some threshold. Therefore, some negative values in the network can be converted to zero by ReLU function since the values are useless. The ReLU activation function is used to make the input value of the next layer non-negative. Using ReLU function as the activation function in different layers means making the value non-negative after different levels of processing. However, in the experiment, the results of using ReLU activation function in different layers don't vary a lot, which means that the phenotype of drug resistance may depend on the existence of the specific proteins rather than the number of the proteins.

4.4 The Preformance of Predicting the Drug Resistance of ETH

There are two possible reasons that the performance is bad in the task of ETH. One is the lack of data, the other is that the function of predicting in ETH is so complicate that the model can't learn from the existent training data.

To deal with the first problem, the experiment of changing the validation split percentage. The results are shown in Fig.8. Although the performance is better than the results using the original validation split percentage, the ROC curves seem unreasonable because of the irregular polylines. It is because the small amount of testing data will make the testing accuracy of no valuable reference.

The second problem is tried to solved by adding ReLU activation function and using very deep model instead. The former method doesn't significantly improve the result but have slightly larger AUC. Although using the latter method, the result can be almost as good as the previous study, the AUC doesn't improve much.

It can be concluded that the failure of the prediction of the drug resistance of ETH is due to the lack of training data.

5 Conclusion

The depth of the model and the learning rate used to train the model should both chosen appropriately. In this task, the best model is of about 5 layers, and the learning rate is about 0.0001. In which layer the ReLU activation function is used doesn't matter.

The performance on predicting the drug resistance to PZA and EMB is good because of the huge dataset. However, the performance on ETH is bad due to the lack of dataset.

Acknowledgments

- All authors contributed equally.
- This work is the final project of Electrical Engineering Lab - Biomedical Engineering. It is supported by the Department of Electrical Engineering, National Taiwan University.
- We thank Prof. Pai-Chi Li for providing many great advice.
- We thank teaching assistant Shang-Ru Li, and Shih-Jou Chung for editing the manuscript.

References

[1] Christina S Thornton, Margot E Grinwis, Christopher D Sibley, Michael D Parkins, Harvey R Rabin, and Michael G Surette. Antibiotic susceptibility and molecular mechanisms of macrolide resistance in streptococci isolated from adult cystic fibrosis patients. *Journal of medical microbiology*, 64(11):1375–1386, 2015.

[2] Jessica MA Blair, Mark A Webber, Alison J Baylay, David O Ogbolu, and Laura JV Piddock. Molecular mechanisms of antibiotic resistance. *Nature reviews microbiology*, 13(1):42, 2015.

[3] Harold C Neu. The crisis in antibiotic resistance. *Science*, 257(5073):1064–1073, 1992.

[4] World Health Organization et al. *Antimicrobial resistance: global report on surveillance*. World Health Organization, 2014.

[5] Anthony R White, BSAC Working Party on The Urgent Need: Regenerating Antibacterial Drug Discovery, Development, Martin Blaser, Otto Carrs, Gail Cassell, Neil Fishman, Robert Guidos, Stuart Levy, John Powers, Ragnar Norrby, et al. Effective antibacterials: at what cost? the economics of antibacterial resistance and its control. *Journal of antimicrobial chemotherapy*, 66(9):1948–1953, 2011.

[6] EMEA ECDC. The bacterial challenge: time to react. *Stockholm: European Center for Disease Prevention and Control*, 2009.

[7] Alessandro Cassini, Liselotte Diaz Högberg, Diamantis Plachouras, Annalisa Quattrocchi, Ana Hoxha, Gunnar Skov Simonsen, Mélanie Colomb-Cotinat, Mirjam E Kretzschmar, Brecht Devleesschauwer, Michele Cecchini, et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the eu and the european economic area in 2015: a population-level modelling analysis. *The Lancet Infectious Diseases*, 19(1):56–66, 2019.

[8] Centres for Disease Control and Prevention (US). *Antibiotic resistance threats in the United States, 2013*. Centres for Disease Control and Prevention, US Department of Health and . . . , 2013.

[9] World Health Organization. *World health statistics 2016: monitoring health for the SDGs sustainable development goals*. World Health Organization, 2016.

[10] World Health Organization et al. *Global tuberculosis report 2018*. World Health Organization, 2018.

[11] World Health Organization et al. Global health estimates 2016: Deaths by cause, age, sex, by country and by region, 2000-2016. geneva. *World Health Organization*, 2018.

[12] Michael D Iseman. Treatment of multidrug-resistant tuberculosis. *New England Journal of Medicine*, 329(11):784–791, 1993.

[13] World Health Organization and Stop TB Initiative (World Health Organization). *Treatment of tuberculosis: guidelines*. World Health Organization, 2010.

[14] José A Caminero, Giovanni Sotgiu, Alimuddin Zumla, and Giovanni Battista Migliori. Best drug treatment for multidrug-resistant and extensively drug-resistant tuberculosis. *The Lancet infectious diseases*, 10(9):621–629, 2010.

[15] World Health Organization et al. *WHO treatment guidelines for drug-resistant tuberculosis*. World Health Organization, 2016.

[16] Timothy M Walker, Thomas A Kohl, Shaheed V Omar, Jessica Hedge, Carlos Del Ojo Elias, Phelim Bradley, Zamin Iqbal, Silke Feuerriegel, Katherine E Niehaus, Daniel J Wilson, et al. Whole-genome sequencing for prediction of mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. *The Lancet infectious diseases*, 15(10):1193–1202, 2015.

[17] H Adler, C Straub, and R Frei. Comparison of bact/alert 3d, lowenstein-jensen medium and middlebrook 7h10/7h11 biplate for recovering mycobacteria from clinical specimens. *European Journal of Clinical Microbiology and Infectious Diseases*, 24(7):499–500, 2005.

[18] Z Samra, L Kaufman, J Bechor, and J Bahar. Comparative study of three culture systems for optimal recovery of mycobacteria from different clinical specimens. *European Journal of Clinical Microbiology and Infectious Diseases*, 19(10):750–754, 2000.

[19] Ramzi Ghodbane, Didier Raoult, and Michel Drancourt. Dramatic reduction of culture time of mycobacterium tuberculosis. *Scientific reports*, 4:4236, 2014.

[20] Viola Schleusener, Claudio U Köser, Patrick Beckert, Stefan Niemann, and Silke Feuerriegel. Mycobacterium tuberculosis resistance prediction and lineage classification from genome sequencing: comparison of automated analysis tools. *Scientific reports*, 7:46327, 2017.

[21] Ted Cohen and Megan Murray. Modeling epidemics of multidrug-resistant m. tuberculosis of heterogeneous fitness. *Nature medicine*, 10(10):1117, 2004.

[22] Francesc Coll, Ruth McNerney, Mark D Preston, José Afonso Guerra-Assunção, Andrew Warry, Grant Hill-Cawthorne, Kim Mallard, Mridul Nair, Anabela Miranda, Adriana Alves, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome medicine*, 7(1):51, 2015.

[23] Yang Yang, Katherine E Niehaus, Timothy M Walker, Zamin Iqbal, A Sarah Walker, Daniel J Wilson, Tim EA Peto, Derrick W Crook, E Grace Smith, Tingting Zhu, et al. Machine learning for classifying tuberculosis drug-resistance from dna sequencing data. *Bioinformatics*, 34(10):1666–1671, 2017.

[24] James E Galagan, Peter Sisk, Christian Stolte, Brian Weiner, Michael Koehrsen, Farrell Wymore, TBK Reddy, Jeremy D Zucker, Reinhard Engels, Marcel Gellesch, et al. Tb database 2010: overview and update. *Tuberculosis*, 90(4):225–235, 2010.

[25] Alice R Wattam, David Abraham, Oral Dalay, Terry L Disz, Timothy Driscoll, Joseph L Gabbard, Joseph J Gillespie, Roger Gough, Deborah Hix, Ronald Kenyon, et al. Patric, the bacterial bioinformatics database and analysis resource. *Nucleic acids research*, 42(D1):D581–D591, 2013.

[26] Francesc Coll, Ruth McNerney, Mark D Preston, José Afonso Guerra-Assunção, Andrew Warry, Grant Hill-Cawthorne, Kim Mallard, Mridul Nair, Anabela Miranda, Adriana Alves, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome medicine*, 7(1):51, 2015.

[27] Roman Novak, Yasaman Bahri, Daniel A Abolafia, Jeffrey Pennington, and Jascha Sohl-Dickstein. Sensitivity and generalization in neural networks: an empirical study. *arXiv preprint arXiv:1802.08760*, 2018.

[28] Lei Wu, Zhanxing Zhu, et al. Towards understanding generalization of deep learning: Perspective of loss landscapes. *arXiv preprint arXiv:1706.10239*, 2017.

Yu-Shung Ting is an undergraduate student in the Department of Electrical Engineering at the National Taiwan University (NTU) and will be graduating in 2021 with a BS in Electrical Engineering. His research projects focus on machine learning on network optimization problems.

Po-Han Li is an undergraduate student in the Department of Electrical Engineering at the National Taiwan University (NTU) and will be graduating in 2020 with a BS in Electrical Engineering. His research projects focus on reinforcement learning on network optimization problems

Kuan-Hao Chao is an undergraduate student in the Department of Electrical Engineering at the National Taiwan University (NTU) and will be graduating in 2020 with a BS in Electrical Engineering. His research projects mainly focus on developing bioinformatics online systems and Bioconductor R package.