SUPERVISED IDENTIFICATION OF ACINETOBACTER BAUMANNI STRAINS USING ARTIFICIAL NEURAL NETWORK

MOHD IZZUDDIN MOHD TAMRIN[1*], MOHD HAFIDZ MAHAMAD MAIFIAH[2], MOHD ZULFAEZAL CHE AZEMIN[3], SHERZOD TURAEV[4], MOHAMED JALALDEEN MOHAMED RAZI[5]

[1]Kulliyyah of ICT, International Islamic University Malaysia, Gombak, Malaysia
[2]International Institute of Halal Research and Training, Gombak, Malaysia
[3]Kulliyyah of Allied Health, International Islamic University Malaysia, Gombak, Malaysia
[4]Department of Computer Science and Software Engineering, College of Information Technology, United Arab Emirates University, Abu Dhabi, United Arab Emirates
[5]Faculty of Commerce and Management Studies, University of Kelaniya, Kelaniya, Sri Lanka

**izzuddin@iium.edu.my*

(Received: 21st August 2019; Accepted: 5th October 2019; Published on-line: 30th October 2019)

ABSTRACT: In hospital environments around the world bacterial contamination is prevalence. One of the most commonly found bacteria is the Acinetobacter Baumannii. It can cause unitary tract, lung, abdominal and central nervous system infection. This bacteria is becoming more resistant to antibiotics. Thus, identification of the non-resistant from the resistant bacteria strain is of important for the correct course of treatments. We propose to use the artificial neural network (ANN) for supervised identification of this bacteria. The mass spectra generated from the liquid chromatography mass spectrometry (LCMS) were used as the features to train the ANN. However, due to the massive number of features, we applied the principle component analysis (PCA) to reduce the dimensions. Less than 1% of the original number of features were utilized. The hand out validation method confirmed that the accuracy, sensitivity and specificity are 0.75 respectively. In order to avoid selection biasness in the sampling, 5-fold cross validation was performed. In comparison, the average accuracy is close to 0.75 but the average sensitivity is slightly higher by 0.50.

KEY WORDS: *Acinetobacter Baumannii, Artificial Neural Network (ANN).*

1. INTRODUCTION

Acinetobacter Baumanni is a type of bacteria that is commonly found in hospital environments from the intensive care unit right to the long-term care unit (Thorne et al., 2019). This bacterium can easily colonize the human body with low immune system. It causes severe unitary tract, lung, abdominal and central nervous system infections. Left untreated it can lead to death. Usually these infections can be treated with multiple classes of antibiotic such as cefepime, ceftazidime and ciprofloxacin. However, A. Baumanni has demonstrated resistance against most classes of antibiotics (Trinh et al., 2019). Its genome has mutated to create defense

mechanism by developing pathogenicity assay. For example, modifying the structure and number of proteins that can pass through the outer cell membrane of the bacteria has tremendously reduced the effectiveness of relevant antibiotic to diffuse.

For this reason, identification of the strains of these bacteria is important to detect the one that is not resistant and the one that is resistant to the antibiotic. Therefore, the correct course of treatments can be effectively planned for the patients. The most commonly used techniques to identify bacteria are polymerase chain reaction (PCR) and phenotypical method (Jain et al., 2019). The former is highly sensitive but costly. However, the cost has reached economies of scale and become affordable. Alternatively, the phenotypical method observes the interaction between the bacteria and its environment. Due to its unique biochemical processes, the bacteria will either colonize the agar or remain inactive when cultured overnight inside a petri dish.

In order to increase the accuracy of the bacteria identification from the same species but different strains, the mass spectrometer can be employed to detect mass proteins from the bacteria (Chamberlain, Rubio & Garrett, 2019). Each bacterium is made up of different proteins and the correct bacteria can be identified if matched against the existing database. However, not the entire strains of the bacteria are available in the database. In this paper, we use the artificial neural network, a technique from machine learning domain to model and automate the identification of A. Baumanni without the database. The peaks from the mass spectra are fed as its features. Since the dataset is huge, we use the principal component analysis to reduce its size. In the following sections of this paper, this research work is elaborated in more details in the literature reviews, methodology, results, discussions and conclusions.

2. LITERATURE REVIEWS

The basis for the needs to identify A. Baumanni has been much deliberated in the previous section. In this section, we explore the techniques of identifying bacteria. There are mainly two approaches, identification at the molecular and phenotypic levels. The former is to go deep and examine the bacteria blueprint i.e., its deoxyribonucleic acid (DNA). Every living thing including bacteria is made up of unique segment of DNA sequences; and thus the accuracy of this approach, for identification, is very high. This method is well known as polymerase chain reaction (PCR) (Falah, Shokoohizadeh & Adabi, 2019; Goudarzi et al., 2019; Lavanya & Uma, 2019) where the bacterial sample undergoes the process of constant heating and cooling. The heating process allows the DNA nucleotides to break free from their bonds and bind to the primer of the targeted bacteria in the cooling process.

Previously, the cost associated with PCR is quite expensive but has become more affordable in recent years. As for the phenotypical method (Anderson & Rather, 2019; Bravo et al., 2019), the agar with the right contents of lactose, soy peptone, casein and sodium chloride are chosen to set up a good environment for the growth of the target bacteria. After that, it must be cultured overnight. Due to its unique biochemical processes that are different from other bacteria, specific biochemical indicators can be observed for A. Baumanni. For example, the level of acidity from the fermentation of lactose and the trace of enzymes to breakdown amino acids such as aspartate and alanine can be used to profile that bacteria.

However, bacteria from the same species but a different strain will be more difficult for this method to differentiate.

Alternatively, the mass spectrometer can be applied to detect the proteins make-up of the target bacteria for greater differentiation between strains from the same species. This technique is known as matrix assisted laser desorption/ionization – time of flight mass spectrometry (MALDI-TOF MS) (Kolk et al., 2019; Mari-Almirall et al., 2019; Neonakis & Spandidos, 2019; Welker et al., 2019). The bacteria sample is mixed with a matrix solution which allows the process of ionization to take place. Once hit by a laser, these analytes vaporize and the positively charged particles are pushed through a magnetic field to hit a panel of detector. The time taken to reach the detector and the intensity of the hit are recorded and compared against the database. These measurements can be used to profile the proteins of the target bacteria. However, reliance on the database is critical. Our approach is to train an artificial neural network model using bacterial mass spectra to automate identification of A. Baumanni that is non-resistant from the one that is resistant to antibiotic.

3. METHODS

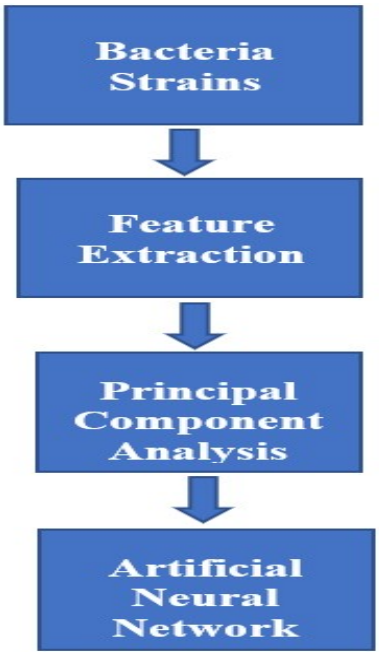

Fig. 1. Overview of the proposed methods for training artificial neural network.

In this section, the entire process involved in training artificial neural network for the identification of non-resistant A. Baumanni is deliberated extensively. Firstly, the bacterial mass spectra generated from cultured A. Baumanni are to be utilized as features for the training. Secondly, due to the overwhelming size of the dataset, the principal component analysis is applied to minimize the number of features. Finally, the dataset is divided into training and testing sets to test the model performance and access whether we have over trained it. Figure 1 summarizes the methods used in this paper.

3.1. Feature Extraction

For detailed explanation, refers to Mahamad Maifiah et al. (2016). The stock of A. Baumanni strains, ATCC 19606 and lpxA mutant 19606R were injected onto separate Mueller Hinton agar plates and incubated at 37 °C for 18 hours to facilitate bacterial growth. The former is a non-resistant strain to polymyxins, a type of antibiotic, that could kill this bacteria by tearing up their cell membrane. However, the latter is resistant to polymyxins. Prior to the incubation, the agar plate was added with 10 mg/L polymyxin B to the lpxA mutant 1906R. This is to ensure that only the resistant population will survive. Following this, for each bacterial strain, a total of three biological replicates and four technical replicates were performed. Due to this, there were 12 samples produced for each strain and a total of 24 samples for both strains.

Next, the biomass of each sample were filled-up with 250 µL extraction solvent and centrifuged at 14000 x *g* for 10 minutes at the temperature of 4°C. This process was performed to release 200 µL intracellular metabolics of the bacteria before it could be injected into the liquid chromatography mass spectrometry (LC-MS) for measurement. The LC-MS used is the Dionex system from Thermo Fisher which could generate positively and negatively-charged ions. As a result, each bacteria strain would produce 24 mass spectra. A total of 48 mass spectra for both strains. These were used as features to train the artificial neural network. However, the system was run for almost 32 minutes and the total ion chromatogram for each sample was massive. In order to reduce the size of this dataset, the principle component analysis was applied.

3.2. Principle Component Analysis

Based on previous section, there are 8000 peaks generated for each mass spectra from the ion chromatogram. Mass spectra is the signal intensity of the ionized intracellular metabolic fragments of the bacteria that were injected into the LC-MS. We are interested to reduce these features but with the minimum loss of information. Principle component analysis (PCA) is one of the best approaches to reduce dimensions. It searches for the best linear combination of these features that can explain most of the information. What we mean by information is the ability of the PCA to explain the most variability between the entire samples. The principle component is defined as follows:

$$Z_i = a_{i,1}(X_1 - \bar{X}_1) + a_{i,2}(X_2 - \bar{X}_2) + \cdots + a_{i,s}(X_s - \bar{X}_s), i = 1, \ldots, s \quad (1)$$

Where $X_1, \ldots, X_s$ are the peaks generated from the mass spectra. $\bar{X}_1, \ldots, \bar{X}_s$ are the mean of these peaks across 24 bacterial samples. Z_i is the set of new features that we want to reduce. In our study, we have reduced from 8000 to 48 features that accounted for 99% of the total variability of the entire samples i.e., the cumulative proportion of variance has reached 0.99. In comparison, this suggests that we can explain 99% of variation in the data with less than 1% of the actual features in the mass spectra. Lastly, the $a_{i,1}, \ldots, a_{i,s}$ are the first until s^{th} weight coefficients of the i^{th} newly derived feature. The partial PCA output for the first five principle components i.e., PCA1 to PCA5 are shown in the appendix.

However, the scale used to measure the mass spectra for positively and negatively charged ions can affect the principal component. For this reason, we have normalized the dataset using the standard z-score defined as follows:

$$v_i' = \frac{v_i - \mu_F}{\sigma_F} \quad (2)$$

Where F is one of the peaks derived from the mass spectra with n observed values of v_i where i = 1, …, n from different samples. μ_F and σ_F are the mean and standard deviation for that feature respectively. The z-score transformations are performed on every value of v_i across the entire 8000 peaks.

3.3. Artificial Neural Network

In the previous section, we have produced PCA1 to PCA48 as the new and reduced features for training the artificial neural network (ANN) (Parisi et al., 2019). ANN can create a model that captures the complex relationship between the predictors and the outcome. This method can satisfy our requirement to model not-so-linear relationship between the new features and identification of non-resistant A. Baumanni strain to polymyxins. It is a data driven method that imitates the way human learns. The architecture is made from three main layers: input, hidden and output. In our study, the input layer comprises neurons that receive values from PCA1 to PCA48. The hidden layer has 10 neurons that employ the resilient backpropagation to update the weights and biases. Finally, the output layer has two neurons that represent the prediction values for the non-resistant and resistant Baumanni strains.

The problem with ANN is that we can over train the model based merely on the existing mass spectra. If there are new mass spectra that come in, the model may not have the same level of accuracy for making a good prediction. In order to evaluate its performance, we need to have a different set of data for testing from the data used for training. For this reason, our dataset is divided into two portions: 60% will be used for training and 40% for testing. Accuracy indicates the overall model performance and is defined as follows:

$$Accuracy = \frac{TP+TN}{P+N} \quad (3)$$

Where TP is the number of times non-resistant and resistant A. Baumanni strains are correctly identified respectively. P and N is the number of non-resistant and resistant A. Baumanni strains respectively. The threshold is set to 0.5. This is used to predict the A. Baumanni strains: (non-resistance = 1) and (resistance = 0). Moreover, sensitivity measures the model ability to correctly identify only the non-resistant A. Baumanni strain and is defined as follows:

$$Sensitivity = \frac{TP}{P} \quad (4)$$

Finally, sensitivity measures the model ability to correctly reject the resistant A. Baumanni strain and is defined as follows:

$$Specificity = \frac{TP}{P} \quad (5)$$

4. RESULTS

Table 1: Model Performance using Hold Out and 5-Folds Cross Validations

Validation Methods	Accuracy	Sensitivity	Specificity
Hold Out	0.75	0.75	0.75

5-Folds Cross Validation	0.73	0.8	0.66

After we have trained the artificial neural network with the training set, the testing set was used to validate the performance of the model. The accuracy, sensitivity and specificity are 0.75 respectively. Due to the biasness in the selection of the sample into 60% of the training set and 40% of the testing set, a different result may be produced instead. In order to ensure the consistency of the results across different sampling approach, we performed a 5-folds cross validation (Rayal et al., 2019). The outcomes from this method indicated that it is just slightly less than the hold out validation method that we have performed earlier. Table 1 summarizes the outcomes from both validation methods. Figure 2 illustrates the boxplots that summarized the overall performance consistency of the model with 5-folds cross validation.

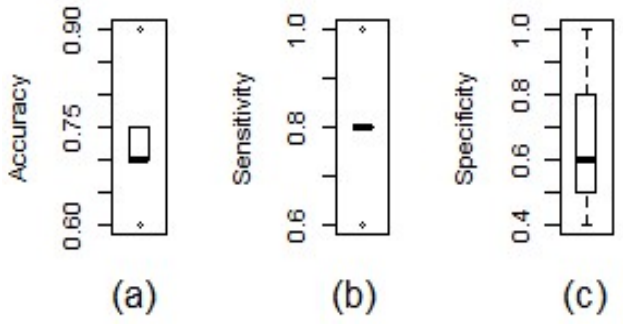

Fig. 2. Boxplot summarized the model overall performance predicting non-resistant A. Baumanni strain. (a) Accuracy. (b) Sensitivity. (c) Specificity.

5. CONCLUSION

Based on the overall performance highlighted from the previous section, the values are considered quite remarkable. This is because the model was trained with only 1% of the original number of 8000 features that were derived from the mass spectra. Although the accuracy did not reach 0.8 target, the average sensitivity from the 5-fold cross validation did reasonably well. This is a good indicator for health science domain since it reduces the type 2 errors. Conversely, the ANN relies heavily on large data to make better prediction. Since our dataset contains only 48 mass spectra used for training, this can be the reason for not reaching this target. In the future, we are considering to replicate more samples from both bacterial strains and thinking of using only the top features with high weight that project to the first principle component. This principle component captures the highest variability of the total variation in the entire dataset.

ACKNOWLEDGEMENT

This work is supported by the International Islamic University Malaysia (IIUM), under the Research Initiative Grant (P-RIGS18-035-0035).

REFERENCES

Anderson, S.E. & Rather, P.N. (2019). Distinguishing Colony Opacity Variants and Measuring Oapcity Variation in Acinetobacter Baumannii. Methods in Molecular Biology. 1946. 151-157.

Bravo, Z., Orruno, M., Navascues, T., Ogayar, E., Ramos-Vivas, J., Kaberdin, V.R. & Arana, I. (2019). Analysis of Acinetobacter Baumannii Survival in Liquid Media and on Solid Matrixes as Well as Effect of Disinfectants . Journal of Hospital Infection. 1-30.

Chamerlain, C.A., Rubio, V.Y., & Garrett, T.J. (2019). Strain-Level Differentiation of Bacteria by Paper Spray Ionization Mass Spectrometry. Analytic Chemistry. 91(8), 4964-4968.

Falah, F., Shokoohizaeh, L. & Adabi, M. (2019). Molecular Identifcation and Genetyping of Acinebacter Baumanni Isolates from Burn Patients by PCR and ERIC-PCR. Scars, Burns and Healing. 5, 1-7.

Goudarzi, H., Mirsamadi, E.S., Ghalavand, Z., Vala, H.M. & Hashemi, A. (2019). Rapid Detection and Molecular Survey of blaVIM, blaIMP and blaNDM Genes among Clinical Isolates of Acinetobacter Baumanni using New Multiplex Real-Time PCR and Melting Curve Analysis. BMC Mirobiology. 122(19), 1-7.

Jain, M., Sharma, A., Sen, M.K., Rani, V., Gaind, R. & Suri, J.C. (2018). Phenotypic and Molecular Characterization of Acinetobacter Baumanni Isolates Causing Lower Respitory Infections among ICU Patients.Microbial Pathogenesis. 128, 75-81.

Kolk, J.H.V.D., Endimiani, A., Graubner, C., Gerber, V. & Perreten, V. (2019). Acinetobacter in Veterinary Medicine, with an Emphasis on Acinetobacter Baimannii. Journal of Global Antimicrobial Resistance. 16, 59-71.

Lavanya, P. & Uma, A.N. (2019). Rapid Detecttion of Gene Encoding OXA Carbapenemases in Acinetobacter using Multiplex PCR. SBV Journal of Basic, Clinical and Applied Health Sciences. 2(1), 25-27.

Mahamad Maifiah, M.H., Cheah, S.E., Johnson, M.D., Han, M.L., Boyce, J.D., Thamlikitkul, V., Forrest, A., Kaye, K.S., Hertzog, P., Purcell, A.W., Song, J., Velkov, T., Creek, D.J., Li, J. (2016). Global Metabolic Analyses Identify Key Differences in Metabolite Levels between Polymyxin-Susceptible and Polymyxin-Resistance Acinetobacter Baumanni. 6, 1-17.

Mari-Almirall, M., Cosgaya, C., Pons, M.J., Nemec, A., Ochoa, T.J., Ruiz, J., Roca, I. & Vila, J. (2019). Pathogenic Acitobacter Species including the Novel Acinetobacter Dijkshoorniae Recovered from Market Meat in Peru. International Journal of Food Microbiology. 305, 1-4.

Neonakis, I.K. & Spandidos, D.A. (2019). Detection of Carbapenemase Producers by Matrix-Assisted Laser Desorption-Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS). European Journal of Clinical Microbiology and Infectious Diseases. 20, 1-7.

Parisi, G.I., Kemker, R., Part, J.L., Kanan, C., Wermter S. (2019). Continual Lifelong Learning with Neural Networks: A Review. 113, 54-71.

Rayal, R., Khanna, D., Sandhu, J.K., Hooda, N., Rana, P.S. (2019). N-sembel: Neural Network based Ensemble Approach. Inernational Journal of Machine Learning and Cybernetics. 10(2), 337-345.

Thorne, A., Luo, T., Durairajan, N.K., Kaye, K.S. & Foxman, B. (2019). Risk Factors for Endemic Acinetobacter Baumanni Colonization. American Journal of Infection Control. 1-4.

Trinh, L.P., Ta, V.S., Luu, T.N.H., Nguyen, M.P., Dinh, T.H., Nguyen, V.H. & Doan T.H., Pham, B.Y. (2019). Studies of Common Antibiotic Resistance-Associated Genes of Acinetobacter Baumanni. Vietnam Journal of Science, Technology and Engineering. 61(1), 58-61.

Welker, M., Belkum, A.V., Girard, V., Charrier, J.P. & Pincus, D. (2019). An Update on the Routine Application of MALDI-TOF MS in Clinical Microbiology. Expert Review of Protoemics. 1-43.

APPENDIX

Table: Partial PCA Output for the First Five Principal Components.

PCA1	PCA2	PCA3	PCA4	PCA5
-3E+09	-1.7E+08	-7.1E+07	-1.7E+08	-2.6E+08
-2.9E+09	-3.6E+08	2.88E+08	-1.1E+08	1.2E+08
-2.4E+09	3.93E+08	1.63E+08	-1.1E+07	-2.5E+07
-3.2E+09	-4.3E+08	1.73E+08	-4.2E+07	2.87E+08
-3.7E+09	-6.2E+08	-3.9E+08	-1.5E+08	1.65E+08
-2.6E+09	-3E+08	2.3E+08	-1.3E+08	1.32E+08
-1.8E+09	-2.2E+08	4.13E+08	2.52E+08	-5.1E+07
-2.6E+09	-3.6E+08	4.23E+08	2.27E+08	-2.4E+08
-2.7E+09	-3.7E+08	2.21E+08	-1.3E+08	-4.2E+07
-2.9E+09	1.7E+08	-3E+08	-1.8E+08	-3.5E+07