ABSTRACT

Multi-drug resistant tuberculosis (MDR-TB) is a disease caused by the *Mycobacterium tuberculosis* that could evade at least two different first-line anti tuberculosis drugs. MDR-TB poses a significant global health challenge, particularly in middle to lower income countries where affordable and rapid diagnostic tools are crucially needed. This situation has brought the idea of leveraging whole genome sequencing and machine learning models for drug resistance predictions. Utilization of *Mycobacterium tuberculosis* genomic data from databases and data pre-processing allows the model to be able to train on the data. The Logistic Regression and Convolutional Neural Network model were trained on the pre-processed genomic data to be able to predict drug resistances. Moreover, both models were evaluated to each other according to the accuracy, sensitivity, specificity, and computational complexity to search for the better model. In accuracy, CNN could outperform LR slightly by outperforming in Rifampicin and Pyrazinamide with a bigger margin than how LR outperforms in Isoniazid and Ethambutol. In sensitivity LR could outperform CNN slightly by outperforming in Rifampicin, Isoniazid, and Pyrazinamide, while CNN could only outperform in Ethambutol. In specificity, CNN could outperform LR slightly by outperforming in Rifampicin and Pyrazinamide with a bigger margin than how LR outperforms in Isoniazid and Ethambutol. Lastly, the computational complexity assessment was invalid due to hardware incompatibility. Overall, each model exhibited its own unique strengths and weaknesses in predicting the first-line anti tuberculosis drug resistances.