Electronic Supplementary Information for:

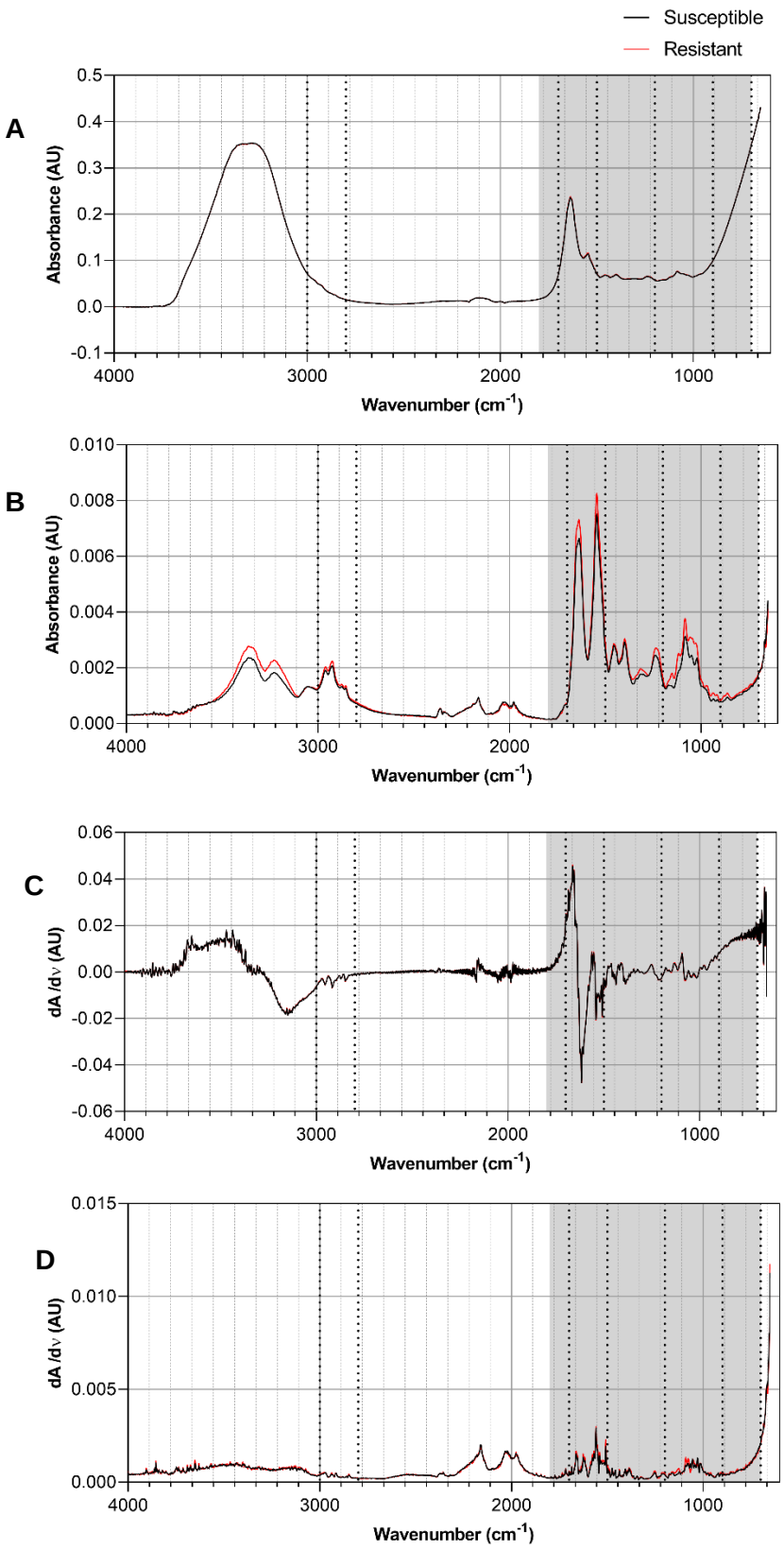
Detecting antimicrobial resistance in *Escherichia coli* using benchtop attenuated total reflectance-Fourier transform infrared spectroscopy and machine learning

Hewa G. S. Wijesinghe,[ab] Dominic J. Hare,[cd] Ahmed Mohamed,[ef] Alok K. Shah,[eg] Patrick N. A. Harris[ahi] and Michelle M. Hill*[ae]

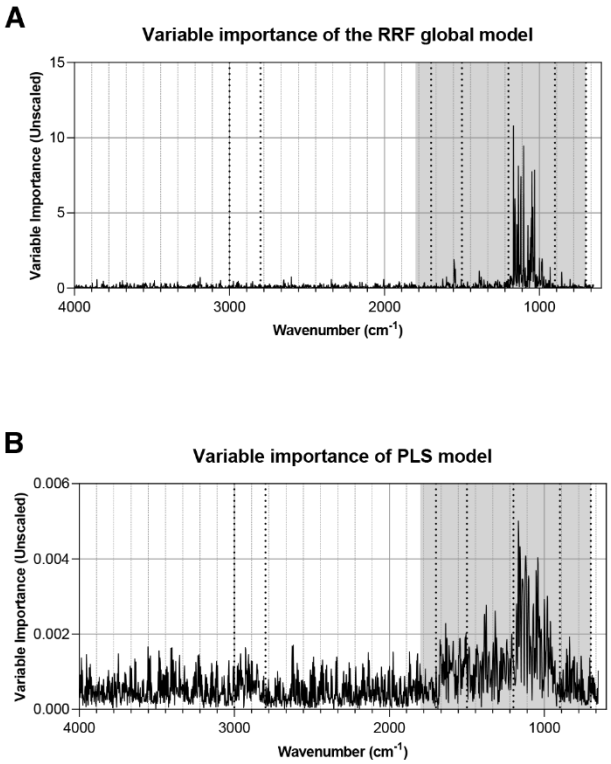
[a] *Centre for Clinical Research, Faculty of Medicine, The University of Queensland, Herston, QLD, 4006, Australia*
[b] *School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, QLD, 4067, Australia.*
[c] *Atomic Medicine Initiative, University of Technology Sydney, Broadway, NSW, 2007, Australia.*
[d] *Monash eResearch Centre, Monash University, Clayton, 3800, Australia (present address).*
[e] *QIMR Berghofer Medical Research Institute, Herston, Brisbane, QLD, 4006, Australia. Email. Michelle.Hill@qimrberghofer.edu.au*
[f] *WEHI, Parkville, VIC, 3052, Australia (present address).*
[g] *CSL Limited, Parkville, VIC, 3052, Australia (present address).*
[h] *Herston Infectious Disease Institute, Royal Brisbane & Women's Hospital, Herston, QLD, 4029, Australia.*
[i] *Central Microbiology, Pathology Queensland, Royal Brisbane & Women's Hospital, Herston, QLD, 4029, Australia.*

CONTENTS:

Data availability statement: All raw data can be accessed via FigShare at: https://figshare.com/s/e2c90fc53cec708bb4b9



ESI Fig 1. Summary spectra for clinical *E. coli* isolates. Average FTIR absorbance (A) and standard deviation (B) plots of raw spectra and the average (C) and standard deviation (D) of Spectra following pre-processing using Savitzky-Golay smoothing (order of 3, window size of 5) and first derivative transformation for 63 ceftriaxone resistant *E. coli* strains (red) and 37 susceptible strains (blue).



ESI Fig 2. Variable importance plot of mean spectral differences using *RFFGlobal* and *PLS* models showing major contributing peak regions.

Table 1. Different machine learning methods used in the study, abbreviated names (used by caret library) and the respective libraries used to implement them in R

Abbreviation	**Method Name**	**R Library**
LogitBoost	Boosted Logistic Regression	caTools
RRF	Regularized Random Forest	randomForest, RRF
RRFglobal	Regularized Random Forest	RRF
dwdLinear	Linear Distance Weighted Discrimination	kerndwd
dwdPoly	Distance Weighted Discrimination with Polynomial Kernel	kerndwd
dwdRadial	Distance Weighted Discrimination with Radial Basis Function Kernel	Kernlab, kern-dwd
gaussprRadial	Gaussian Process with Radial Basis Function Kernel	kernlab
gbm	Gradient Boosting Machines	h2o
gcvEarth	Multivariate Adaptive Regression Splines	earth
hdda	High Dimensional Discriminant Analysis	HDclassif
kernelpls	Partial Least Squares	pls
knn	k-Nearest Neighbors	kknn
mlp	Multi-Layer Perceptron	RSNNS
mlp-WeightDecay	Multi-Layer Perceptron	RSNNS
monmlp	Monotone Multi-Layer Perceptron Neural Network	monmlp
naive_bayes	Naive Bayes	naivebayes
null	Non-Informative Model	NULL
pam	Nearest Shrunken Centroids	pamr
pls	Generalized Partial Least Squares	gpls
ranger	Random Forest	e1071, ranger, dplyr
rbfDDA	Radial Basis Function Network	RSNNS
rf	Quantile Random Forest	quantregForest
sdwd	Sparse Distance Weighted Discrimination	sdwd
simpls	Partial Least Squares	pls
svmLinear	Least Squares Support Vector Machine	kernlab
svmLinear2	Support Vector Machines with Linear Kernel	e1071
svmLinear-Weights	Linear Support Vector Machines with Class Weights	e1071
svmPoly	Least Squares Support Vector Machine with Polynomial Ker-nel	kernlab
svmRadial	Least Squares Support Vector Machine with Radial Basis Function Kernel	kernlab
svmRadialCost	Support Vector Machines with Radial Basis Function Kernel	kernlab
svmRadialSigma	Support Vector Machines with Radial Basis Function Kernel	kernlab
svmRadial-Weights	Support Vector Machines with Class Weights	kernlab
widekernelpls	Partial Least Squares	pls
xgbDART	eXtreme Gradient Boosting	xgboost, plyr
xgbLinear	eXtreme Gradient Boosting	xgboost
xgbTree	eXtreme Gradient Boosting	Xgboost, plyr