AI FOR LIFE: MACHINE LEARNING FOR ANTIBIOTIC SCREENING

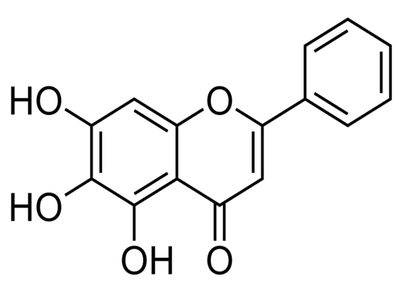
M. McKerlie*, C. Kern, B. Howlin, M. Sacchi and D. Gems

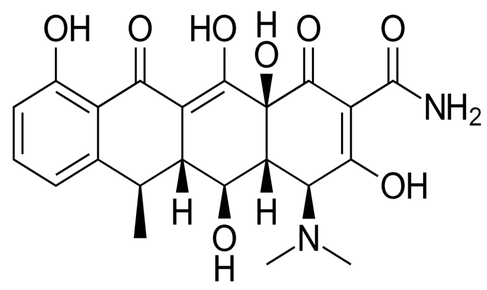
BACKGROUND

Caenorhabditis elegans are microscopic roundworms commonly used to study mechanisms of ageing and extending lifespan interventions. However, the lifespan of *C. elegans (below)* are subject to age related disease and life shortening pharyngeal infection, affecting roughly 40% of the population when under standard culture conditions with *E. coli.*

As a result, there is a possibility a large percentage of similar drug outputs of computational approaches like this are flawed due to effects on bacterial pathogenicity.

Previous studies that have predicted lifespan extension in *C. elegans* do not consider the effect of bacterial pathogenicity on the lifespan of the roundworm. The DrugAge database below lists both Baicalein *(left)* and Doxycycline *(right)* as active lifespan extending compounds but ignore Doxycycline's antibacterial activity against *E. coli* which would fight against the pharyngeal infection causing premature death.



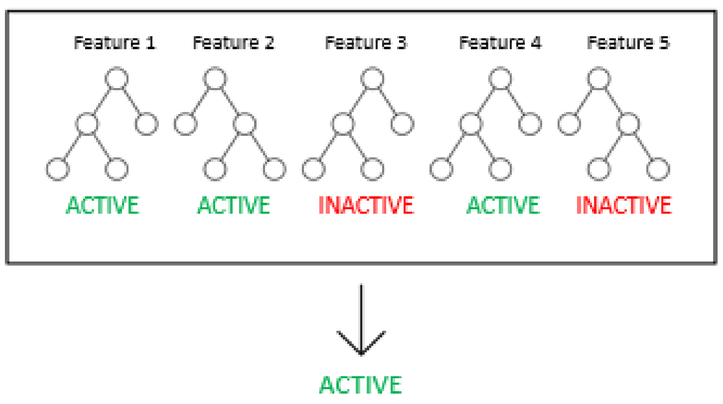


PURPOSE OF THE STUDY:

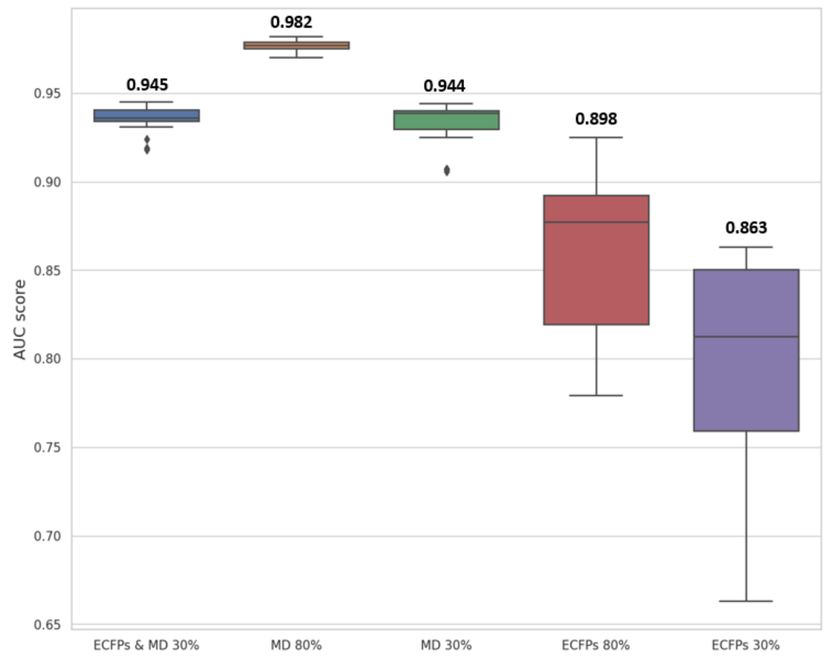
The possibility that drug properties associated with suppression of infection interferes with the detection of potential 'anti-ageing' interventions due to lifespan extension via infection suppression.

METHOD

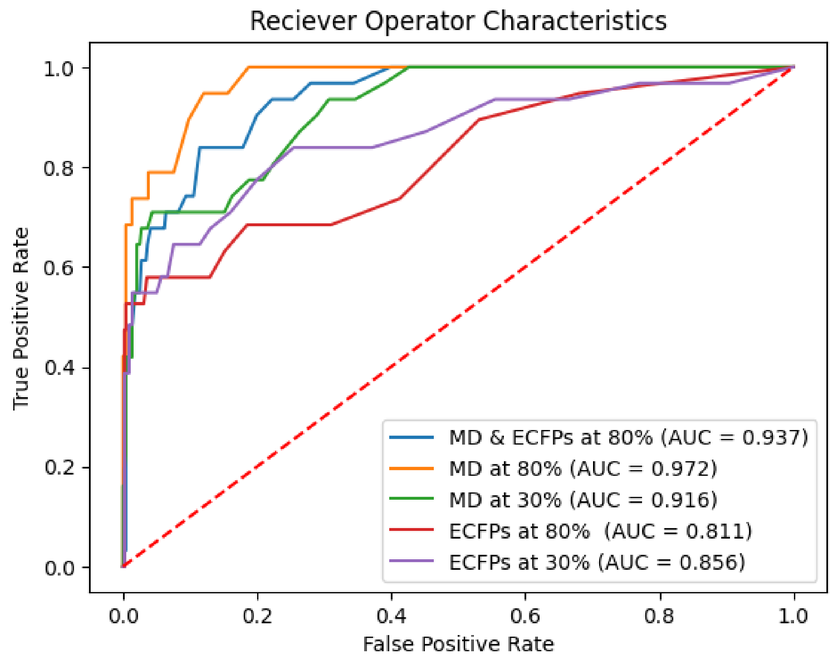
Five random forest classifier models were built in python using the features: molecular descriptors, 1024-bit extended connectivity fingerprints (ECFPs) and a mixture of both at two different activity thresholds. The model was trained on a dataset of 2335 molecules.



RESULTS



AUC scores for the 10 –fold cross validation in the training set.



The ROC curves comparing the performances of the 5 models in the test set.

CONCLUSION & FUTURE RESEARCH

Of the 5 developed, the best performing model was built with molecular descriptors as the features at an activity threshold of 80%. After consideration, it was decided that the model would be more versatile for deployment using a threshold of 30%, so some accuracy was sacrificed for versatility. The final model selected was built using a mixture of 1024-bit ECFPs and molecular descriptors and had a ROC AUC score of **0.937. 27 molecules of the 304** active lifespan extending molecules were highlighted as antibiotically active, and subsequently removed from consideration.

With antibiotic resistance a continuing worry, there is scope for this model to be developed to predict antibiotics against more than just *E. coli.* This could lead to a novel technique for scanning large existing drug databases, potentially leading to drug repurposing of existing molecules with antibiotic properties.

UNIVERSITY OF SURREY

For more information email:

m.mckerlie@surrey.ac.uk

The Institute of HEALTHY AGEING