MALAWI UNIVERSITY OF SCIENCE AND TECHNOLOGY
ACADEMY OF MEDICAL SCIENCES

DEPARTMENT OF BIOLOGICAL SCIENCES

NAME: ROPAFADZO ZOE KWINDANI

TOPIC: THE APPLICATIONS OF MACHINE LEARNING IN HEALTHCARE(MEDICAL MICROBIOLOGY)

COUNTRY: MALAWI

INSTITUITION: THE MALAWI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ABSTRACT

In the last decade, machine learning((ML), gained significant momentum in health care, especially for advanced diagnostics and improved patient care. Machine learning is an application of artificial intelligence (AI) that provides computer based systems with the ability to automatically learn and improve from experience without being explicitly programmed. The basic premise of machine learning is to build models that can receive input data and use statistical analysis to predict an output while updating outputs as new data becomes available. The models are coded in different programming languages including ; python, r and java just to mention a few. However, for most African countries, resource scarcity is a challenge although the technology has proved to be of immense potential and also crucial for addressing the unique challenges faced by the continent.

INTRODUCTION

Machine learning can be categorized into supervised and unsupervised algorithms. Supervised algorithms are used when the data used to train is classified and labeled while unsupervised algorithms are used in unlabeled data. Nowadays, machine learning is used in a wide range of applications such as social media services, identifying disease and diagnosis, drug discovery and clinical researches. This position paper highlights the applications of machine learning in health care and how machine learning and AI techniques can be used as innovative tools of solving some of the problems faced by the medical field in the African continent.

1. **PYBACT**

Pybact is a software that was written in python for bacterial identification. The code simulates the predefined behaviour of bacterial species by generating a simulated data set based on the frequency table of biochemical tests from microbiology textbook. Results indicated that the classifiers could accurately predict its respective bacterial class with 98% accuracy. This algorithm has assisted microbiologists in rapidly identifying pathogens and understanding the prevalence of certain strains in different regions.

The figures below show the algorithms code input and output in one of python's extensive libraries.

```
import modules: random, os, sys, and psyco
read data from input file
construct an array and import input data as array members
for each bacteria (from a list of bacterial species):
    generate class label for N number of strains
    for each biochemical test (from a list of biochemical tests):
        populate the value as 0 (representing negative results);
        if the biochemical test has a predefined positive percentage of 0
        then populate all value as 0
        else
            for each biochemical test (from a list of biochemical tests):
                assign a value of 1 (representing positive results) to a random position
                until the cumulative count of 1 equals the predefined positive percentage
            append biochemical test to the list
write results to output file in the form of a tab delimited data matrix
```

Table 1

PyBact's data generation algorithm

The input data matrix is depicted as follows (showing only the first 5 biochemical test and the first 2 bacterial species):

Indole	MR	VP	Citrate	H_2S	
99	99	75	97	0	*V.cholerae*
98	99	9	99	0	*V.mimicus*

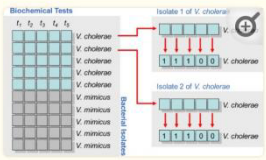


Figure 1

Schematic representation of biochemical test data generation by PyBact

After the algorithm's procedures have been performed, the generated output yields the following data matrix (this example shows only the first 5 biochemical test and 5 strains of the first 2 bacterial species):

Indole	MR	VP	Citrate	H_2S	
1	1	0	1	0	*V.cholerae*
0	1	0	1	0	*V.cholerae*
1	0	1	1	0	*V.cholerae*
1	1	1	0	0	*V.cholerae*
1	1	1	1	0	*V.cholerae*
1	1	0	1	0	*V.mimicus*
1	1	0	0	0	*V.mimicus*
1	1	0	1	0	*V.mimicus*
0	0	0	1	0	*V.mimicus*
1	1	0	1	0	*V.mimicus*

2. ANTIMICROBIAL RESISTANCE PREDICTION FROM CLINICAL MALDI-TOF MASS SPECTRA(ARPCM)

Antimicrobial resistance has become a serious global concern. After having been exposed to antibiotics, bacteria can quickly develop resistance due to their short growth cycle and various adaptation mechanisms. Both supervised and unsupervised machine learning tools have been successfully used to predict early antibiotic resistance, and thus support clinicians in selecting appropriate therapy. The ARPCM model was written in python and r programming to predict antibiotic resistance patterns of bacterial isolates based on their genomic data. The data set combines more than 300,000 mass spectra with more than 750000 antimicrobial resistance phenotypes from different medical institutions. Validation on a panel of clinically important pathogens, including Staphylococcus aureus, Escherichia coli and Klebsiella pneumoniae, resulting in areas under the receiver operating characteristic curve of 0.80, 0.74 and 0.74, respectively, demonstrated the potential of using machine learning to substantially accelerate antimicrobial resistance determination and change of clinical management. Furthermore, a retrospective clinical case study of 63 patients found that implementing this approach would have changed the clinical treatment in nine cases, which would have been beneficial in eight cases (89%). MALDI-TOF mass spectra-based machine learning may thus be an important new tool for treatment optimization and antibiotic stewardship.

README.md

resistance prediction using logistic regression), it is sufficient to issue the following commands:

```
poetry shell # Not necessary if you are already in the virtual environment
python plot_fig4_curves_per_species_and_antibiotic_2panels.py
```

Afterwards, the output file `fig4.png` will be created, which reproduces the E. coli panel of Figure 3 in the paper:

The figures below showcase a visual representation of the model, input and output code in one of pythons extensive libraries.

README.md

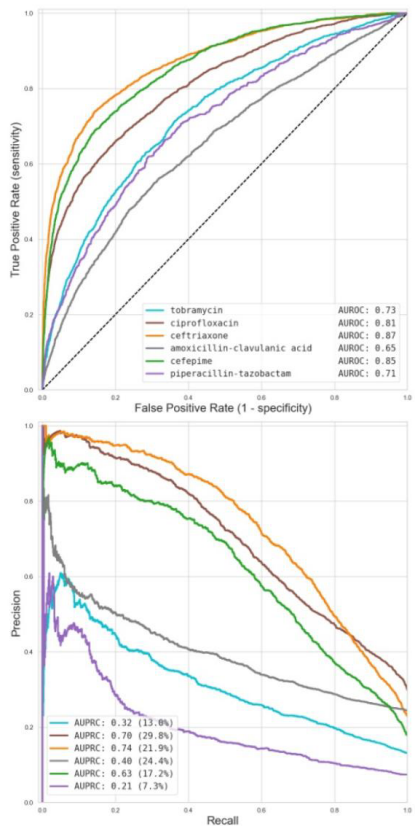
antibiotic	model						
Amoxicillin-Clavulanic acid	lightgbm	77.06	0.82	43.83	1.83	67.02	1.41
	lr	75.93	0.76	40.96	2.86	65.81	1.41
	rf	75.71	0.15	41.13	1.29	66.27	1.76
	svm-linear	54.09	10.54	30.84	1.63	56.93	2.08
	svm-rbf	62.48	13.59	39.91	1.84	64.23	1.63
Cefepime	lightgbm	88.99	0.63	69.85	2.90	88.17	1.47
	lr	87.54	0.82	63.18	3.07	85.59	1.22
	rf	84.91	0.41	66.99	2.65	86.92	1.75
	svm-linear	71.46	19.17	47.35	4.90	76.04	3.49
	svm-rbf	58.30	35.40	64.24	1.94	85.24	1.51
Ceftriaxone	lightgbm	88.42	0.84	79.41	2.13	89.55	1.36

README.md

resistance prediction using logistic regression), it is sufficient to issue the following commands:

```
poetry shell # Not necessary if you are already in the virtual environment
python plot_fig4_curves_per_species_and_antibiotic_2panels.py
```

Afterwards, the output file `fig4.png` will be created, which reproduces the E. coli panel of Figure 3 in the paper:



CONCLUSION

We can conclude that machine learning is an ideal technique and therefore essential in health care. However there still exists possible factors that may potentially hinder accurate and definitive bacterial identification and antimicrobial resistance prediction. That is why it is important for African researchers who are interested in artificial intelligence to come together and share ideas on how best they can utilize the available resources at their disposal and build a solid legacy for African AI.

REFERENCE

1.Bassi S. A primer on python for life science researchers. PLoS Comput Biol. 2007;3:e199. [PMC free article] [PubMed] [Google Scholar]

2. Biopython. Available from: http://www.biopython.org.

3. Frederiksen W, Tønning B. Possible misidentification of Haemophilus aphrophilus as Pasteurella gallinarum. Clin Infect Dis. 2001;32:987–989. [PubMed] [Google Scholar]

4. Giacomini M, Ruggieroa C, Calegari L, Bertone S. Artificial neural network based identification of environmental bacteria by gas-chromatographic and electrophoretic data. J Microbiol Meth. 2000;43:45–54. [PubMed] [Google Scholar]

5. Janda JM, Abbott SL. Bacterial identification for publication: when is enough enough? J Clin Microbiol. 2002;40:1887–1891. [PMC free article] [PubMed] [Google Scholar]

6. Lapage SP, Bascomb S, Willcox WR, Curtis MA. Identification of bacteria by computer: general aspects and perspectives. J Gen Microbiol. 1973;77:273–290. [PubMed] [Google Scholar]

7. Logan NA. Bacterial systematics. Oxford: Blackwell Scientific Publ; 1994. [Google Scholar]

8. Murray PR, Baron EJ, Jorgensen JH, Landry ML, Pfaller MA. Manual of clinical microbiology. Washington DC: ASM Press; 2007. [Google Scholar]

9.Rawson T.M., Ming D., Ahmad R., Moore L.S.P., Holmes A.H. Antimicrobial use, drug-resistant infections and COVID-19. Nat. Rev. Microbiol. 2020;18:409–410. doi:

10.1038/s41579-020-0395-y. [PMC free article] [PubMed] [CrossRef] [Google Scholar]
10. Beam A.L., Kohane I.S. Big Data and Machine Learning in Health Care. JAMA. 2018;319:1317–1318. doi: 10.1001/jama.2017.18391. [PubMed] [CrossRef] [Google Scholar]

11. Macesic N., Polubriaginof F., Tatonetti N.P. Machine learning: Novel bioinformatics approaches for combating antimicrobial resistance. Curr. Opin. Infect. Dis. 2017;30:511–517. doi: 10.1097/QCO.0000000000000406. [PubMed] [CrossRef] [Google Scholar]

12. Rajpurkar P., Chen E., Banerjee O., Topol E.J. AI in health and medicine. Nat. Med. 2022;28:31–38. doi: 10.1038/s41591-021-01614-0. [PubMed] [CrossRef] [Google

13., A. P. et al. Diagnostic performance of the Xpert Carba-RTM assay directly from rectal swabs for active surveillance of carbapenemase-producing organisms in the largest Brazilian University Hospital. J. Microbiol. Methods 171, 105884 (2020).

14. van Belkum, A., Welker, M., Pincus, D., Charrier, J. P. & Girard, V. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry in clinical microbiology: what are the current issues? Ann. Lab. Med. 37, 475–483 (2017).

15. Croxatto, A., Prod'hom, G. & Greub, G. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. FEMS Microbiol. Rev. 36, 380–407 (2012).

16. Dierig, A., Frei, R. & Egli, A. The fast route to microbe identification: matrix assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS). Pediatr. Infect. Dis. J. 34, 97–99 (2015).

17. Hou, T.-Y., Chiang-Ni, C. & Teng, S.-H. Current status of MALDI-TOF mass spectrometry in clinical microbiology. J. Food Drug Anal. 27, 404–414 (2019).

18. Kim, J.-M. et al. Rapid discrimination of methicillin-resistant Staphylococcus aureus by MALDI-TOF MS. Pathogens 8, 214 (2019