

Machine Learning and Gas Sensors for Antimicrobial Resistance Identification

Albert Belenguer-Llorens^{1,2}, Emma Slack^{3,4,5}, Adrian Egli^{6,7}, Thomas M. Kessler⁸, Andreas T. Güntner⁹, Catherine R. Jutzeler^{1,2}

¹ *Department of Health Sciences and Technology (D-HEST), ETH Zurich, Switzerland;*

² *SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland*

³ *Laboratory for Mucosal Immunology, Institute for Food, Nutrition and Health, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland*

⁴ *Basel Institute for Child Health, Basel, Switzerland*

⁵ *Sir William Dunn School of Pathology, University of Oxford, Oxford, UK*

⁶ *Institute of Medical Microbiology, University of Zurich, Zurich, Switzerland*

⁷ *Swiss National Reference Center for Emerging Antibiotic Resistance (NARA) Network, University of Fribourg, Fribourg, Switzerland*

⁸ *Department of Neuro-Urology, Balgrist University Hospital, University of Zurich, Zurich, Switzerland*

⁹ *Department of Mechanical and Process Engineering, ETH Zurich, Zurich, Switzerland.
abelenguer@ethz.ch*

Summary:

The global rise of antimicrobial resistance (AMR) requires a paradigm shift away from slow, culture-based diagnostics toward more rapid, data-driven approaches. We propose a novel pipeline that integrates high-resolution volatilomics extracted from gas sensors with machine learning (ML) techniques to bypass traditional time-consuming workflows, enabling rapid identification of resistance profiles directly from microbial emissions. This approach aims to provide a scalable foundation for point-of-care diagnostics, with the potential to reduce antibiotic overuse and support timely, precision-guided therapy.

Keywords: Machine Learning, Volatilomics, Antimicrobial Resistance, Microbiology, Gas Sensors

Background, Motivation and Objective

The global rise of antimicrobial resistance (AMR) is one of the current most concerning and rising public health problems, estimated to cause more than 8 million deaths by 2050 [1]. The AMR crisis is driven and exacerbated by the misuse (e.g. premature discontinuation of antibiotic treatment) and their overuse in both healthcare and the food industry. A major contributing factor is that current clinical diagnostics cannot adequately counter this trend in real time. Conventional microbial identification and susceptibility testing are predominantly culture-based, combining techniques such as matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) and antibiotic susceptibility testing (AST), which usually requires from 24 hours to several weeks. Thus, this inherent delay forces clinicians to rely on empirical broad-spectrum antibiotic therapy, a practice that can compromise patient outcomes and facilitate the emergence of resistant strains.

To address these challenges, we aim to propose a new diagnostic pipeline to bypass the

limitations of traditional microbiology using volatilomics data captured by gas sensors [2]. Microbial metabolism produces specific volatile organic compounds (VOCs) that can serve as rapid biomarkers for both species identification and resistance profiling. Although volatilomics is not yet as widely established in clinical routines as other omics modalities (such as genomics or proteomics), there are recent growing evidence of its utility in bacterial identification [3]. The complexity and high dimensionality of volatilomic data require robust computational frameworks to be clinically actionable. In this context, Machine Learning (ML) algorithms have already demonstrated strong performance in related modalities, such as in genomics and mass spectrometry, where they are used to identify biological signatures.

To this end, we propose a novel ML pipeline designed to identify AMR profiles directly from the headspace of clinical samples. By combining high-resolution discovery using mass spectrometry with the development of tailored

chemoresistive sensor arrays, we aim to establish a rapid, data-driven diagnostic workflow.

Volatilomics in microbiology

Microbial volatilomics involves the systematic characterization of low-molecular-weight metabolites (primarily VOCs) secreted during bacterial proliferation and metabolic shifts under specific conditions such as antibiotic stress. These data represent a high-dimensional space, where target analytes (e.g., indoles, ketones, or short-chain fatty acids) often coexist with complex background matrices at trace-level concentrations (parts-per-billion or trillion).

The technical challenge lies in the dynamic nature of these emissions: VOC profiles are not static but exhibit distinct temporal kinetics (Figure X). Thus, requiring from complex computational methods to analyze the different metabolites interactions/tendencies across time.

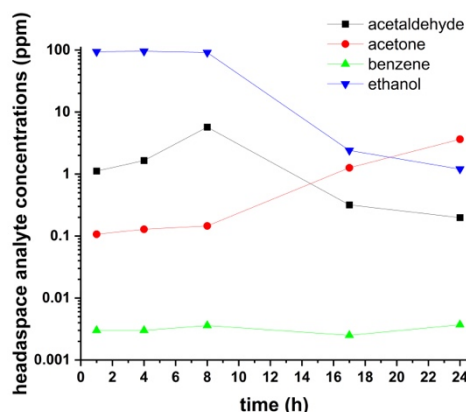


Fig. 1. Example of variation in concentration (ppm) of VOCs for a specific bacteria across time.

Description of the New Method or System

To extract meaningful diagnostic patterns from the VOCs dynamics, we propose a locally deployable ML pipeline designed for integration into resource-constrained diagnostic devices. To ensure compliance with the EU AI Act for high-risk Class III medical devices, our architecture prioritizes interpretability over "black-box" deep learning (DL) models, which are also unsuitable for the limited sample size of these datasets, typical of early-stage clinical studies [4].

The pipeline initiates with a Hidden Markov Model (HMM) layer to characterize the VOCs emission dynamics of resistance vs non-resistance bacteria. That is, given the dynamics over time of new sample, calculate their probability to belong to one of both. To improve performance, we will extract other kinetic-based features such as signal entropy or tendency using a rolling window. Then, to capture the potential

non-linear relations without sacrificing transparency, the final variable set (HMM-derived and kinetic features) will be used to train an ensemble of interpretable algorithms, including XGBoost, Random Forests, Support Vector Machines (SVMs), and Generalized Additive Models (GAMs).

A critical stage of the pipeline is a robust feature selection protocol, which the above models can natively perform with regularization techniques. This identifies the minimal subset of high-impact biomarkers required for accurate AMR prediction, directly informing the design of miniaturized sensor arrays tailored for a specific, limited number of target molecules.

Discussion and on-going work

The proposed pipeline aims to effectively integrate high-dimensional volatilomics and ML for microbial diagnosis in a portable gas sensor-based device. A central hypothesis of our work is that AMR identification cannot rely on isolated VOC analysis; instead, it requires the simultaneous assessment of collective analyte dynamics. Interactions between different volatile families, such as ratios among metabolic byproducts, are expected to yield more robust and specific signatures than single biomarkers alone. Disentangling these complex multi-analyte relationships using ML is key to reducing the dimensionality of the problem. By identifying the minimal optimal set of VOCs required for accurate classification, we can reduce system complexity and thereby enable device miniaturization. Furthermore, the transition from centralized, high-cost laboratory equipment to a streamlined, ML-driven sensor array offers a scalable pathway for the democratization of diagnostics.

On-going work is currently focused on the longitudinal data collection using clinical samples to train and validate our ML pipeline.

References

- [1] Naghavi, Mohsen, et al. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *The Lancet* 1199-1226 (2024); doi: 10.1016/S0140-6736(24)01867-1.
- [2] Bilgin, Mehmet Berk, et al. Microbial and antimicrobial resistance diagnostics by gas sensors and machine learning. *Cell Biomaterials* 1.7 (2025); doi: 10.1016/j.celbio.2025.100125.
- [3] Kemmler, Emanuel, et al. mVOC 4.0: a database of microbial volatiles. *Nucleic Acids Research* 53.D1 (2025); 10.1093/nar/gkae961.
- [4] Aboy, Mateo, Timo Minssen, and Effy Vayena. Navigating the EU AI Act: implications for regulated digital medical products. *NPJ Digital Medicine* 7.1 (2024): 237; doi: 10.1038/s41746-024-01232-3.