ROBERT KOCH INSTITUT

Robert Koch Institute

Centre for Artificial Intelligence in Public Health Research (ZKI-PH)

ZKI-PH PhD Programme 2026-2029

Image source: AdobeStock

ZKI-PH PhD Programme 2026-2029

ROBERT KOCH INSTITUT

ZKI-PH_PhD2026_01 ***(ZKI-PH2 & FG13)***
Integrating machine learning and phylogeography to trace *Staphylococcus aureus* outbreaks in healthcare facilities

ZKI-PH_PhD2026_02 ***(ZKI-PH3 & FG13)***
Deep-Learning-based species identification and prediction of antimicrobial resistance (AMR) patterns in bacterial isolates

ZKI-PH_PhD2026_03 ***(ZKI-PH4 & FG37)***
Exploration and integration of AMR drivers in Germany and beyond

ZKI-PH_PhD2026_04 ***(ZKI-PH5 & MF1)***
Representation Learning and Explainable AI for Functional Peptide Discovery

ZKI-PH_PhD2026_05 ***(ZKI-PH2 & ZBS6, ZIG4)***
Deep Learning for uncovering hidden genomic drivers of Vibrio cholerae fitness

ZKI-PH_PhD2026_06 ***(ZKI-PH4 & FG24, FG25)***
Predictive analysis of climate-related medication use and vulnerable population groups by linking survey data, statutory health insurance, and meteorological data in Germany

ZKI-PH_PhD2026_07 ***(ZKI-PH1 & ZKI-PH4)***
Applying AI/ML Methods for the Global Assessment of Antimicrobial Resistance (AMR) and Policy Interventions

ZKI-PH_PhD2026_01 (ZKI-PH2 & FG13)

ROBERT KOCH INSTITUT

Integrating machine learning and phylogeography to trace *Staphylococcus aureus* outbreaks in healthcare facilities

Background

Staphylococcus aureus is considered a major public health problem as an opportunistic pathogen responsible for a wide variety of infections. It is one of the main causes of nosocomial infections (hospital-acquired infections), with multi-drug resistant strains, especially MRSA (Methicillin-Resistant *Staphylococcus aureus*), posing significant treatment challenges, leading to serious conditions like pneumonia, skin infections, and bloodstream infections. Molecular surveillance has improved outbreak detection but integration of spatial, temporal, and genomic data is lacking to resolve transmission pathways. Recent advances in machine-learning–based phylogeographic analysis show promise in capturing relationships between epidemiological and genomic variables. Developing specialized, interpretable machine-learning frameworks for *S. aureus* promises to uncover hidden transmission dynamics, enhance early outbreak detection, and inform targeted, evidence-based infection control across interconnected healthcare settings.

Aim/s

The German National Reference Center for Staphylococci and Enterococci (FG13) characterized *S. aureus* isolates from about 200 laboratories nationwide. Using advanced machine-learning–based phylogeographic models, over 180 single–spa-type isolates collected over 13 years will be analyzed to assess genetic variability, reconstruct transmission networks, and trace strain introductions across hospitals.

Methods

The goal of the project is the development of appropriate machine learning approaches for the phylogeographic analysis of nosocomial *S. aureus.* This analysis will combine classical phylogeographic approaches with state-of-the-art machine learning and deep-learning tools (e.g. Transformer & Graph Neural Networks), with the goal to better understand the transmission dynamics of *S. aureus* within and between hospitals.

Keywords

Computer Vision, Deep Learning, AMR Detection, Bacteria Identification

ZKI-PH_PhD2026_02 (ZKI-PH3 & FG13)

ROBERT KOCH INSTITUT

Deep-Learning-based species identification and prediction of antimicrobial resistance (AMR) patterns in bacterial isolates

Background

Antimicrobial resistance (AMR) poses a major challenge to public health and accounts for more than a million yearly fatality cases worldwide. To quickly identify and characterize potential resistant bacteria, a novel, high-throughput Deep Learning framework will be developed, capable of identifying bacterial species based on their colony morphology and color. Wastewater samples will be separated into individual cells. Each bacterial cell will form a new colony, which will be classified according to their species based on MALDI-TOF Mass Spectrometry. The images of the colonies will serve as input to a deep learning model to classify the bacterial species. Within the project, different growth media will be tested and evaluated to develop a bacterial test system. Furthermore, this setup could be used for AMR diagnostics. To achieve this, antibiotic-specific serial dilution assays could be performed to determine resistance profiles based on a standardized minimum inhibitory concentration (MIC) test.

Aim/s:

This project aims to identify subtle visual differences in bacterial colonies, thereby allowing for bacterial species characterization and potential resistances as a fast and cheap identification/ characterization system.

Methods:

To identify and characterize different bacterial cultures, deep learning models such as convolutional neural networks and vision transformers will be implemented and trained.

Keywords:

Computer Vision, Deep Learning, AMR Detection, Bacteria Identification

ZKI-PH_PhD2026_03 (ZKI-PH4 & FG37)

ROBERT KOCH INSTITUT

Exploration and integration of AMR drivers in Germany and beyond

Background

This PhD project will investigate the spatiotemporal spread of antimicrobial resistance (AMR) in Germany and within the broader European context using data-driven methods. The project builds on recent evidence of substantial regional heterogeneity in AMR incidence. We will explore how differing antimicrobial prescribing patterns, climatic conditions (e.g., temperature, humidity, extreme weather), sociodemographic structures (e.g., age, income, migration, urban-rural gradients), and healthcare utilization jointly shape resistance dynamics. The findings will help to pinpoint regional “hotspots” of resistance, generate hypotheses about underlying drivers, and support more targeted antibiotic stewardship and public-health interventions in Germany and Europe.

Aim/s:

This project will compile and integrate multiple data sources, including routine surveillance data, hospital and outpatient prescribing statistics, environmental indicators, and census-based sociodemographic information. A core aim is to develop a harmonized, high-resolution data pool that enables robust comparison across German regions and as well with selected European countries.

Methods:

The work will combine modern epidemiological analyses with modern data-science workflows, emphasizing transparent data cleaning, linkage, and exploratory visualization. Machine-learning techniques (e.g., tree-based models, regularized regression, and spatiotemporal clustering) will be used to identify key predictors of resistance patterns and potential non-linear interactions between climate and sociodemographic factors. By systematically comparing different model families and validation strategies, the project aims to establish best practices for predictive AMR modelling in heterogeneous settings.

Keywords:

AMR, Climate and Environmental Change, Spatio-Temporal Analysis

ZKI-PH_PhD2026_04 (ZKI-PH5 & MF1)

ROBERT KOCH INSTITUT

Representation Learning and Explainable AI for Functional Peptide Discovery

Background

The global increase in antibiotic resistance requires a change in approach to the discovery of new antimicrobial agents. Although antimicrobial peptides (AMPs) offer a promising solution, their vast sequence space makes traditional experimental screening prohibitively slow. This PhD project aims to integrate advanced protein language models (PLMs) with explainable AI (XAI) frameworks, transforming peptide discovery from a predictive task into an interpretable one. The core innovation lies in using contrastive learning to improve the model's capacity to discern subtle functional differences, coupled with interpretability tools that emphasize the importance of specific amino acids. This dual approach improves the accuracy of bioactivity predictions and provides a transparent rationale for sequence design. The project's ultimate goal is to bridge the gap between in silico generation and experimental validation, thereby fostering the development of targeted, safe and effective therapeutic peptides.

Aim/s

This project aims to develop an interpretable computational framework for peptide discovery. The goal is to identify functional sequence motifs, optimize de novo peptide generation and provide a transparent molecular basis for antimicrobial efficacy that will guide subsequent experimental validation, by combining contrastive representation learning with generative architectures.

Methods

The methodology uses transformer-based architectures and PLMs that have been trained using extensive sequence repositories. A key focus is on implementing novel tokenizer designs and encoding strategies to optimize the contrastive learning signal. The model clusters functional variants in latent space by utilizing dual-encoder frameworks and specific objective functions. Explainability modules are then integrated to extract feature importance, ensuring that the generated candidates are supported by interpretable structural determinants.

Keywords

Contrastive Learning, Representation Learning, Explainable AI (XAI), Generative Transformer Models, Sequence Encoding Strategies

ZKI-PH_PhD2026_05 (ZKI-PH2 & ZBS6 & ZIG4)

ROBERT KOCH INSTITUT

Deep Learning for uncovering hidden genomic drivers of *Vibrio cholerae* fitness

Background

Cholera remains a major public health challenge across Africa, with periodic outbreaks driven by diverse Vibrio cholerae lineages. Despite significant advances in genomic surveillance, uncovering the genetic basis of antimicrobial resistance (AMR) and fitness in *Vibrio cholerae* remains challenging. Traditional comparative genomics and alignment-based methods rely on known resistance genes and reference genomes, likely missing subtle or novel sequence variations that may affect phenotype. Existing machine learning approaches still depend on preselected features and can struggle to generalize across diverse lineages. The AFR15 lineage, which shows enhanced persistence across African settings despite limited canonical AMR markers, exemplifies this blind spot. Addressing this requires more flexible, data-driven models capable of detecting hidden genomic patterns. An alignment-free deep learning framework will be developed that learns sequence representations directly from data, aiming to reveal new determinants of microbial fitness and resistance evolution.

Aim/s

This project seeks to elucidate the genomic drivers underlying the unusual fitness and potential antimicrobial tolerance of the AFR15 *Vibrio cholerae* lineage. Specifically, it aims to develop, train, and interpret a deep learning model to identify sequence motifs linked to phenotypic outcomes, validate these predictions experimentally, and integrate the resulting insights into public health surveillance frameworks.

Methods

This project will design an alignment-free transformer-based deep learning architecture for the analysis of genomic *Vibrio cholerae* sequences. Using self-attention and embedding layers, the model will learn latent representations of sequence segments, enabling phenotype classification without relying on predefined features. Interpretability modules will rank influential genomic regions, guiding targeted laboratory validation.

Keywords

Genomic Epidemiology, *Vibrio cholerae,* Antimicrobial Resistance, Sequence Modeling

ZKI-PH_PhD2026_06 (ZKI-PH4, FG24 & FG25)

ROBERT KOCH INSTITUT

Predictive analysis of climate-related medication use and vulnerable population groups by linking survey data, statutory health insurance, and meteorological data in Germany

Background

In the context of planetary health, the German pharmaceutical supply system needs to ensure the pharmacological safety of an aging society in the face of escalating extreme weather events caused by climate change. Moreover, there are several medications that are necessary to monitor during periods of extreme heat due to potential drug-drug interactions, thermoregulatory effects, or heat-related adverse reactions. Since heat waves alter both, the physiological effects of medications and affect overall morbidity, a data-driven analysis of these dynamics is needed to assess the impact of heat waves and identify those most at risk.

Aim/s

This project aims to assess the impacts of extreme heat events on medicine use and cross-sector prescribing behavior for cardiovascular medications and other drug categories in and across Germany with a focus on quantifying vulnerable population groups. The long-term objective is to create a predictive model that identifies regional supply anomalies early on, establishing the foundation for an AI-supported early warning system. This will provide valuable insights for public health planning, healthcare resource allocation, and targeted prevention strategies aimed at protecting high-risk groups during periods of extreme heat.

Methods

A multimodal database is being created by linking the high-resolution prescription data from the FDZ Gesundheit with historical weather data from the German Weather Service (DWD) and additional geodata. Additionally, the RKI study "Gesundheit 65+" allows for quantifying vulnerable populations Germany potentially affected by periods of extreme heat with regard to medication use, socio-demographic and health determinants. The project will employ machine learning methods, specifically time-series-based regression models and Random Forests, and analyses of geodata to investigate prescribing behavior and medication use across German regions.

Keywords

Climate Impacts, Medication Use, Drug Safety

ZKI-PH_PhD2026_07 (ZKI-PH1 & ZKI-PH4)

ROBERT KOCH INSTITUT

Applying AI/ML Methods for the Global Assessment of Antimicrobial Resistance (AMR) and Policy Interventions

Background

Antimicrobial resistance (AMR) is an escalating global health crisis, contributing to over 700,000 deaths annually and projected to rise to 10 million by 2050 if left unaddressed. The ability of microorganisms to resist antimicrobial drugs threatens the effectiveness of treatments, complicating infection management, and leading to increased healthcare costs. Despite efforts by global organizations such as the World Health Organization (WHO) to curb AMR through various policies and interventions, the impact of these measures remains difficult to assess, especially on a global scale. This gap is due, in part, to the lack of systematic analyses and the integration of diverse data sources across regions. Understanding how AMR trends are shaped by policy changes, antibiotic use, and public health practices is essential for developing targeted and effective future interventions.

Aim/s

This project aims to conduct a comprehensive global assessment of AMR trends, linking them to past and present policy interventions. By leveraging advanced AI techniques and natural language processing, it will evaluate the landscape of AMR policies and interventions, offering evidence-based insights into their impact. By comparing observed regional trends, the project will quantify the impact of past policy decisions on resistance patterns and thus provide valuable guidance for decision-makers on future AMR prevention strategies.

Methods

The project will employ machine learning and natural language processing techniques to analyze vast datasets from various global sources, including national surveillance system and available policy documents. Data will be integrated and used to identify underlying AMR trends across regions and pathogens and model future AMR trends based on current policy interventions. Different text mining strategies will be employed, including topic modelling and tracking, and information processing through large language models and retrieval augmented generation.

Keywords

NLP, Text Mining, AMR Policy Landscape