PNAS RESEARCH ARTICLE | MICROBIOLOGY

# Phenotype-Based Threat Assessment

Jing Yang[a,1], Mohammed Eslami[b,1], Yi-Pei Chen[b,1], Mayukh Das[a,1], Dongmei Zhang[a,1], Shaorong Chen[a], Alexandria-Jade Roberts[a], Mark Weston[b], Angelina Volkova[b], Kasra Faghihi[b], Robbie K. Moore[a], Robert C. Alaniz[a], Alice R. Wattam[c], Allan Dickerman[c], Clark Cucinell[c], Jarred Kendziorski[a], Sean Coburn[a], Holly Paterson[a], Osahon Obanor[a], Jason Maples[a], Stephanie Servetas[d], Jennifer Dootz[d], Qing-Ming Qin[a], James E. Samuel[a,2], Arum Han[e,g,2], Erin J. van Schaik[a,2], and Paul de Figueiredo[a,f,2]

Edited by Ralph Isberg, Tufts University School of Medicine, Boston, MA; received July 22, 2021; accepted January 27, 2022

**Bacterial pathogen identification, which is critical for human health, has historically relied on culturing organisms from clinical specimens. More recently, the application of machine learning (ML) to whole-genome sequences (WGSs) has facilitated pathogen identification. However, relying solely on genetic information to identify emerging or new pathogens is fundamentally constrained, especially if novel virulence factors exist. In addition, even WGSs with ML pipelines are unable to discern phenotypes associated with cryptic genetic loci linked to virulence. Here, we set out to determine if ML using phenotypic hallmarks of pathogenesis could assess potential pathogenic threat without using any sequence-based analysis. This approach successfully classified potential pathogenetic threat associated with previously machine-observed and unobserved bacteria with 99% and 85% accuracy, respectively. This work establishes a phenotype-based pipeline for potential pathogenic threat assessment, which we term PathEngine, and offers strategies for the identification of bacterial pathogens.**

bacterial pathogen | machine learning | threat assessment | adherence | toxicity

## Significance

Assessing the threat posed by bacterial samples is fundamentally important to safeguarding human health. Whole-genome sequence analysis of bacteria provides a route to achieving this goal. However, this approach is fundamentally constrained by the scope, the diversity, and our understanding of the bacterial genome sequences that are available for devising threat assessment schemes. For example, genome-based strategies offer limited utility for assessing the threat associated with pathogens that exploit novel virulence mechanisms or are recently emergent. To address these limitations, we developed PathEngine, a machine learning strategy that features the use of phenotypic hallmarks of pathogenesis to assess pathogenic threat. PathEngine successfully classified potential pathogenic threats with high accuracy and thereby establishes a phenotype-based, sequence-independent pipeline for threat assessment.

How is bacterial pathogenic potential predicted? Classical characterization of pathogenic potential used Koch's postulates (1). Recently, however, machine learning (ML) analysis of whole-genome sequences (WGSs) from bacteria has been used to predict pathogenic potential. For example, PaPrBaG is an ML approach for detecting novel pathogens from next-generation sequencing (NGS) data (2). This approach predicts pathogenicity by training ML models on a large number of simulated reads from pathogens and nonpathogens (2). Other approaches using ML models derived from sequencing data have also been described (e.g., Institute for Machine Learning and Analytics, DeePaC, and PathoFact) (3–5). A key feature of these recent ML methods is that pathogenic potential is predicted based solely on WGS reads without any other biological context. Although these methods are powerful, they have the potential to falsely characterize pathogens that have cryptic genetic variation, which is often observed in different strains of the same species. For example, genetic comparison of virulence genes between hyperinvasive and apathogenic *Neisseria meningitidis* provided no answers for the differences in pathogenicity; however, transcriptomic data suggested that virulence was linked to buffered expression of cryptic genetic loci (6, 7). Another report found a similar pattern where expression patterns rather than genome sequences determined antimicrobial resistance profiles to 12 different drugs in a collection of strains of *Acinetobacter baumannii* (8). ML models trained with genome sequence data alone are also susceptible to reporting false-positives where nonpathogenic features are mislabeled as virulence determinants (9). To overcome these limitations, some have emphasized the pressing need to enhance our knowledge of phenotypic characteristics, such as severity of infection and virulence phenotypes in infection models, at the same rate as that of bacterial genomes that are coming online (9).

With these ideas in mind, we pursued the development of an alternative strategy using an ML pipeline termed PathEngine that assessed data from classic pathogenic phenotypes. PathEngine is a general approach that uses an ensemble ML model across these assays to provide robust and accurate potential pathogenic threat assessment of phylogenetically diverse bacteria. In addition, after demonstrating that PathEngine worked using well-characterized hallmarks of pathogenic potential, we explored whether other assays not classically associated with virulence could also be used. Here, we describe this strategy to assess pathogenic potential of bacteria using four phenotypic assays. We used a diverse set of bacterial pathogens and nonpathogens, totaling a mere 40 bacterial strains belonging to 16 genera and 29 species, for training and testing each model, which was significantly more economical than reports for similar tasks in previous WGS-based approaches (Fig. 1*A*) (2, 3).

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1] J.Y., M.E., Y.-P.C., M.D., and D.Z. contributed equally to this work.

[2] To whom correspondence may be addressed. Email: jsamuel@tamu.edu, arum.han@ece.tamu.edu, vanschaik@tamu.edu, or pjdefigueiredo@tamu.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2112886119/-/DCSupplemental.

Published April 1, 2022.

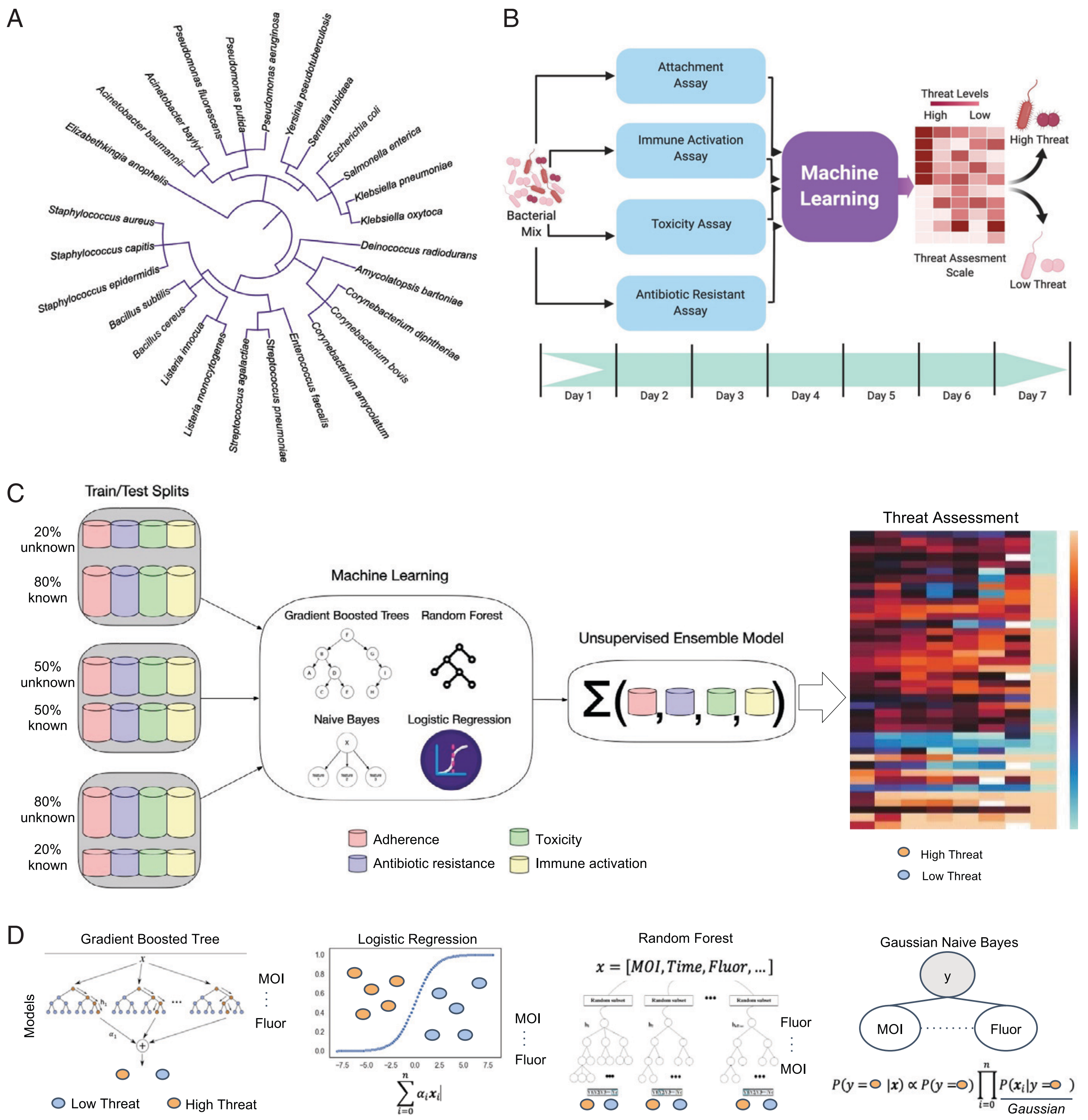

**Fig. 1.** Framework for generation of an ML platform that enables bacterial threat assessment. (*A*) Bacterial strains used in this work are phylogenetically divergent. (*B*) An overall framework in a time frame for threat assessment. (*C*) Overview of architecture of ML workflow includes data requirements and processing, model selection, and threat assessment. Unknown and known bacterial pathogens used in the threat assessment by the indicated ML models. (*D*) Overview of computational architecture of the four different ML models used in this work. MOI, multiplicity of infection. Fluor, fluorescence.

To make this strategy work with such a limited number of strains, we made each assay multimodal. This provided a large training set for improving the effectiveness of ML models in assessing the potential pathogenic threat. Therefore, we developed assays based on hallmarks of pathogenesis that generated graded readouts of bacterial pathogenicity with many data points (Fig. 1*B*). These phenotypes included classic virulence mechanisms such as bacterial adherence to host cells and toxicity to host cells. In addition, we used antibiotic resistance (AR), which is not a classic virulence mechanism, to show the power of our ML pipeline using phenotypic data. We also demonstrated that an assay that features bacteria-induced expression of host innate immune genes could be used for potential pathogenic threat assessment using PathEngine. For multimodal data-driven approaches to provide unbiased findings, the analysis of large datasets (e.g., >10,000 data points) is required, and in some of our assays more than 1 million data points were analyzed. However, we overcame limitations of the small dataset sizes generated in some assays such as AR (<1,000 data points) by using ensemble data points in all assays instead of plus or minus strain-level analyses. This approach enabled successful training of ML models despite limited bacterial strains. Based on this work, an ensemble method was applied to aggregate the model's predictions across observations and assays to output a pathogenic score per organism (Fig. 1*C*). The ensemble method combined weak predictions from the observations into a strong classifier, a well-studied technique in the ML community (10). In addition, the ensemble approach allowed the assays to complement each other to make accurate predictions. To guarantee robustness to the model in deployed scenarios, we evaluated it with two tests: Test 1 (T1) quantified the model's ability to predict pathogenic potential of untested observations from previously tested bacteria, while Test 2 (T2) quantified the model's ability to predict pathogenic potential of previously untested bacteria.

## Results

**Bacterial Adherence to Host Cells.** Bacterial colonization is an essential first step in the virulence programs for most human pathogens, and the ability to adhere to host cells avoids the mechanical clearance mechanisms of the body (11). Adherence to host cells is therefore one example of a potential pathogenic phenotype. Classic examples of well-characterized adherence phenotypes include the interactions between A549 pneumocytes and *Pseudomonas aeruginosa* or *Staphylococcus aureus,* both lung pathogens of patients with cystic fibrosis (12–16). To obtain multimodal model data from bacterial adherence assays that can be used in ML algorithms, an automated image-based bacterial adherence assay was used. We incubated A549 cells with fluorescently tagged Gram-negative bacterial strains *P. aeruginosa* (PAO1) (as a positive control) and its attachment-deficient mutant PAO1Δ*pilA*, as well as *Escherichia coli* DH5α (negative controls). As expected, image analysis revealed that bacterial adherence to host cells by control strains differed significantly from the dose-dependent adherence of the wild-type organism (Fig. 2 *A* and *B*). Similar results were observed when corresponding Gram-positive bacterial strains of *S. aureus,* together with its mutant Δ*saeR* and *Bacillus subtilis,* were used in these assays (*SI Appendix*, Fig. S1 *A* and *B*). Importantly, findings garnered from experiments performed using fluorescence microscopy–based approaches were validated using colony-forming unit (CFU) assays that enumerated bacterial adherence to host cells (*SI Appendix*, Fig. S1*C*). After establishing the control parameter benchmarks, a diverse set of pathogenic and nonpathogenic bacterial strains from National Institute of Standards and Technology (NIST) collections were also tested and used to train ML models (Fig. 1*A* and *SI Appendix*, Fig. S2).

Based on the results from adherence assays performed with control strains and the NIST collection, we trained four ML models [i.e., Gradient Boosted Trees (GBT) (17), Logistic Regression (LR) (18), Random Forests (RFs) (19), and Gaussian Naïve Bayes (GNB) (20)] to assess threat based on phenotypic measurements from the imaging studies (Fig. 1*D*). For this training exercise, every training point for the assay included five features: 1 to 3) the average/minimum/maximum number of adherent bacteria per host cell, 4) the total number of host cells present in the image, and 5) the size of the host nucleus as a proxy to measure the size of the host cell. The models were then trained to make a prediction of pathogenic potential per image. The results indicated that RF was the best performing T1 model, with an average balanced accuracy of 65% ± 4% (Fig. 2*C*). RF was also the best performing T2 model with an average balanced accuracy of 56% ± 6% (Fig. 2*D*). Additional host cell lines were also investigated to explore if the predictions were biased toward respiratory pathogen by A549 cells. Our hypothesis was that using the diverse set of pathogens in the training model would remove any bias. Human umbilical vein endothelial cells (HUVECs) and a human hepatocyte carcinoma cell line (HepG2) were tested in the same bacterial adherence assays, and similar results were observed, confirming our hypothesis. For T1, the balanced accuracy was 64% for HUVECs and 63% for HepG2 (*SI Appendix*, Fig. S3 *A* and *C*). For T2, the balanced accuracy was 57% for HUVECs and 56% for HepG2 (*SI Appendix*, Fig. S3 *B* and *D*). This demonstrates that although different pathogens target different host cell types, there is no one-size-fits-all cell line that best captures adherence phenotype. Consequently, the models predictions did not change when a diverse group of bacteria was used. However, using only one assay is problematic because although *Salmonella enterica* is a pathogen, it was not resolved as such using adherence alone (21). We were not surprised by the poor performance of ML models based on adherence phenotypes alone given the limited nature of the datasets employed. We expected, however, that combining several assays would yield improved accuracy.

**Bacterial Toxicity to Host Cells.** Cytotoxicity through invasion or production of toxins is another hallmark of bacterial pathogenesis as exemplified by Enterohemorrhagic *Escherichia coli* (EHEC)-producing Shiga-like toxin bacteria, which have been extensively characterized (22, 23). We developed a multimodal toxicity assay using time as a measured dimension. THP-1 cells, a human monocytic cell line that can be differentiated in vitro to macrophages, were used as toxicity with Shiga toxins, and other modes of cytotoxicity exemplified by *Salmonella* spp. are well documented (24, 25). Diverse bacteria from our test panel (*SI Appendix*, Table S1) at different time points were tested to assess the effects of damage to the cell membrane and/or cell death by pathogens. Cell death was measured using automated imaging to monitor the uptake of propidium iodide (PI) by cells over time (26, 27) (Fig. 3 *A* and *B* and *SI Appendix*, Fig. S1*D*). To validate the automated imaging findings, cell viability was measured by flow cytometry (*SI Appendix*, Fig. S1*E*). Bacteria that express Shiga-like toxin, a classic A-B toxin that induces programmed cell death of host cells (23, 28, 29), were used as positive controls in these experiments (Fig. 3 *A* and *B*). Analysis of the data using ML models revealed that the best performing T1 and T2 models were GBT and RF, with balanced accuracies of 77% ± 1% (Fig. 3*C*) and 68% ± 10% (Fig. 3*D*), respectively. These results suggested that host cell toxicity induced by bacteria provided a promising additional assay for potential pathogenic prediction by ML models.

**Bacterial AR.** Although this phenotype is not a classical pathogenic trait, it is correlated with virulence and therefore was included in our assays (30, 31). The ability to query this assay using the diverse set of strains in both the pathogen and nonpathogen categories and our ML pipelines should have been able to resolve its predictive capacity. To test whether the feature of AR could be used in the ML models, we first performed standard Kirby-Bauer Disk Diffusion assays to estimate AR (32). Six different antibiotics, kanamycin, ampicillin, chloramphenicol, tetracycline, polymyxin B, and ceftazidime, were selected to cover diverse mechanisms of action. Consistent with previous reports (33, 34), we found that the strains known to express AR determinants (e.g., beta-lactamases), displayed AR in these assays (Table 1 and *SI Appendix*, Table S2). Encouraged by these findings, we used the assay to measure resistance levels of strains in the test panel and then applied the results to train the four ML models. We found that the GBT model, which sequentially learns the best decision

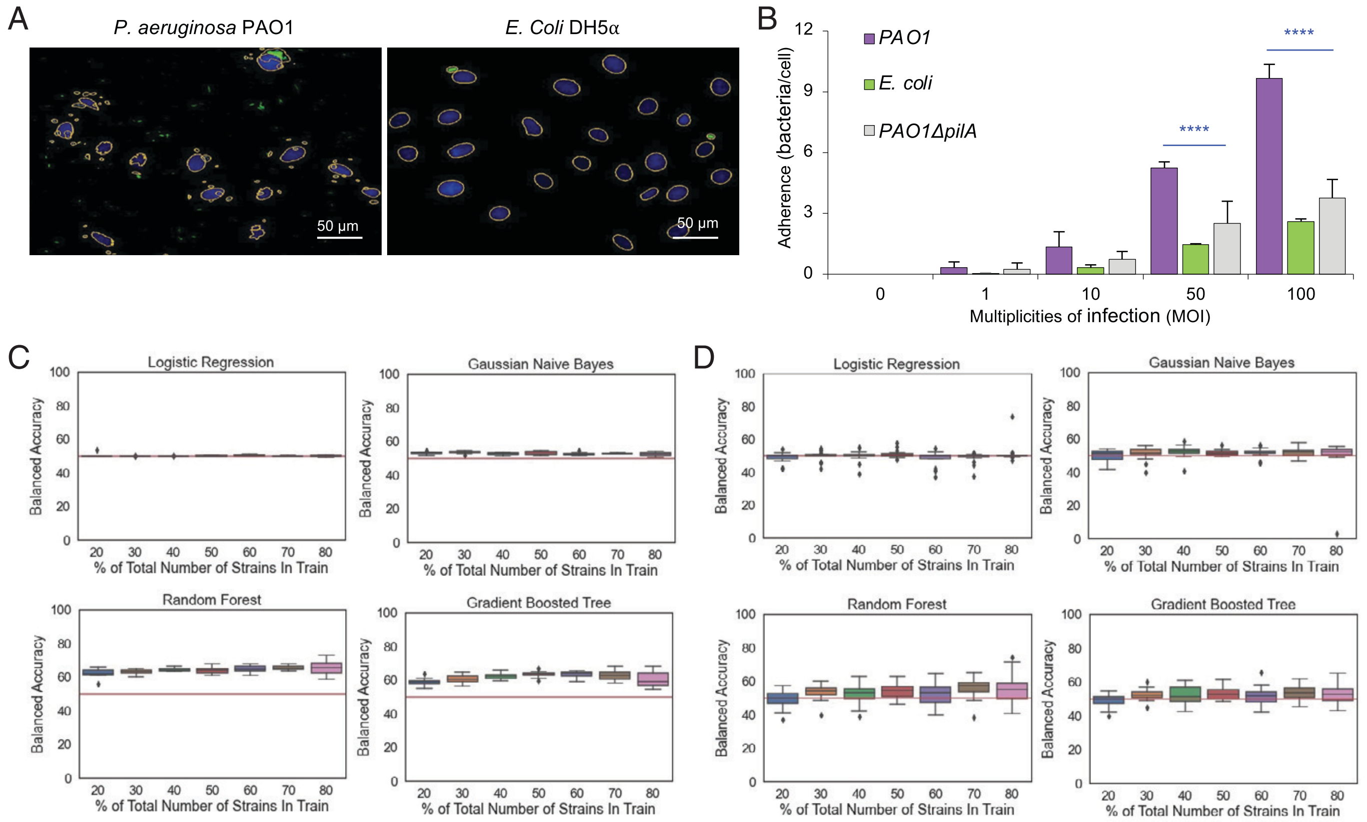

**Fig. 2.** Bacterial adherence performance in evaluating bacterial threat assessment using the ML model. (*A*) Representative images of adherence assays for *P. aeruginosa* and *E. coli* as positive and negative controls. The adherent bacteria and their corresponding target host cells were counted and marked with outlines. Host cells (blue) were stained by DAPI, and bacteria (green) were GFP-tagged. (Scale bar: 50 µm.) (*B*) Average adherent bacterial counts per A549 cell under various MOIs. Data represent the means ± SDs from three independent experiments. At each MOI, $n \geq 15$. Significant difference in adherent bacteria at MOIs of 50 and 100 was observed (****: *P* value < 0.0001). (*C* and *D*) Performance of the four ML models in Test 1 (*C*) or Test 2 (*D*) for adherence assay. All models were characterized to determine the percentage of data required to plateau in performance. Each machine learning algorithm was run 20 times, with the error bars showing the 95% confidence interval from the accuracy scores in each run. The accuracy referred to the percentage of strains assigned correctly by the models.

tree that separates pathogens from nonpathogens, outperformed all other models, achieving T1 performance of 82% ± 3% balanced accuracy (Fig. 4*A*). The best performing T2 model was the RF model, with an average balanced accuracy of 64% ± 6% (Fig. 4*B*). To evaluate whether using more clinically relevant antibiotics would affect the model's predictive power, we tested a separate set of six clinically relevant antibiotics including azithromycin, ciprofloxacin, doxycycline, imipenem, meropenem, and rifampin. We found that the best T1 performance was 85% ± 1% (*SI Appendix*, Fig. S4*A*) and the best T2 performance was 65% ± 16% (*SI Appendix*, Fig. S4*B*), The results from two antibiotic sets demonstrated that AR provided an important bullet in differentiating pathogens and nonpathogens and that the performance of more clinically relevant antibiotics did not significantly differ from that of others (Fig. 4 *A* and *B* and *SI Appendix*, Fig. S4 *A* and *B*).

**Bacteria-Induced Innate Immune Activation.** Molecular pathways that detect pathogens based on pathogen-associated molecular patterns are commonly found in immune cells such as macrophages. These innate immune cells recognize bacteria based on conserved cellular structures found on both pathogen and nonpathogen. We decided to incorporate a cell type for this assay using the commercially available nuclear factor κB (NF-κB)/Jurkat/green fluorescent protein (GFP) transcriptional reporter cell line (Systems Biosciences). Our original hypothesis was that these cells would not be useful for determining pathogenic potential because the NF-κB pathway would be universally activated by bacteria. However, we were curious to see if determining fluorescent intensity of the reporter in individual cells and using time as a factor to create multimodal data might allow our ML pipeline to make accurate predictions. We found that the reporter cells showed differential GFP expression when incubated with pathogenic and nonpathogenic bacteria, such as *S. enterica* and *E. coli* DH5α (Fig. 4*C*). Similar results were also obtained using fluorescence microscopy imaging (*SI Appendix*, Fig. S1*F*). To assess the utility of monitoring innate immune signaling for potential pathogenic threat assessment, bacterial strains from NIST collections were tested (*SI Appendix*, Fig. S5) and evaluated using ML models. We found that RF and LR were the best performing T1 and T2 models, with maximal average balanced accuracy of 64% ± 0.4% and 63% ± 7% (Fig. 4 *D* and *E*), respectively. Collectively, these data suggested that host cell immune activation provided an appropriate feature for evaluating a potential threat predicted by ML.

To interrogate the hypothesis that integrating individual ML phenotype models into an ensembled ML model would enhance threat assessment, we therefore developed an ensemble ML model, PathEngine, that combined predictions from weakly supervised models (35) to generate predictions at both an assay level, where we computed the pathogenic potential of a microbe per assay, and across assays, from which a prediction of pathogenic potential of each microbe could be determined. After aggregating the observations across each strain, we obtained an improved accuracy. For T1, the AR, host cell toxicity, adherence assays, and host immune activation achieved accuracies of 95%, 91%, 77%, and 70%, respectively (Fig. 5*A*

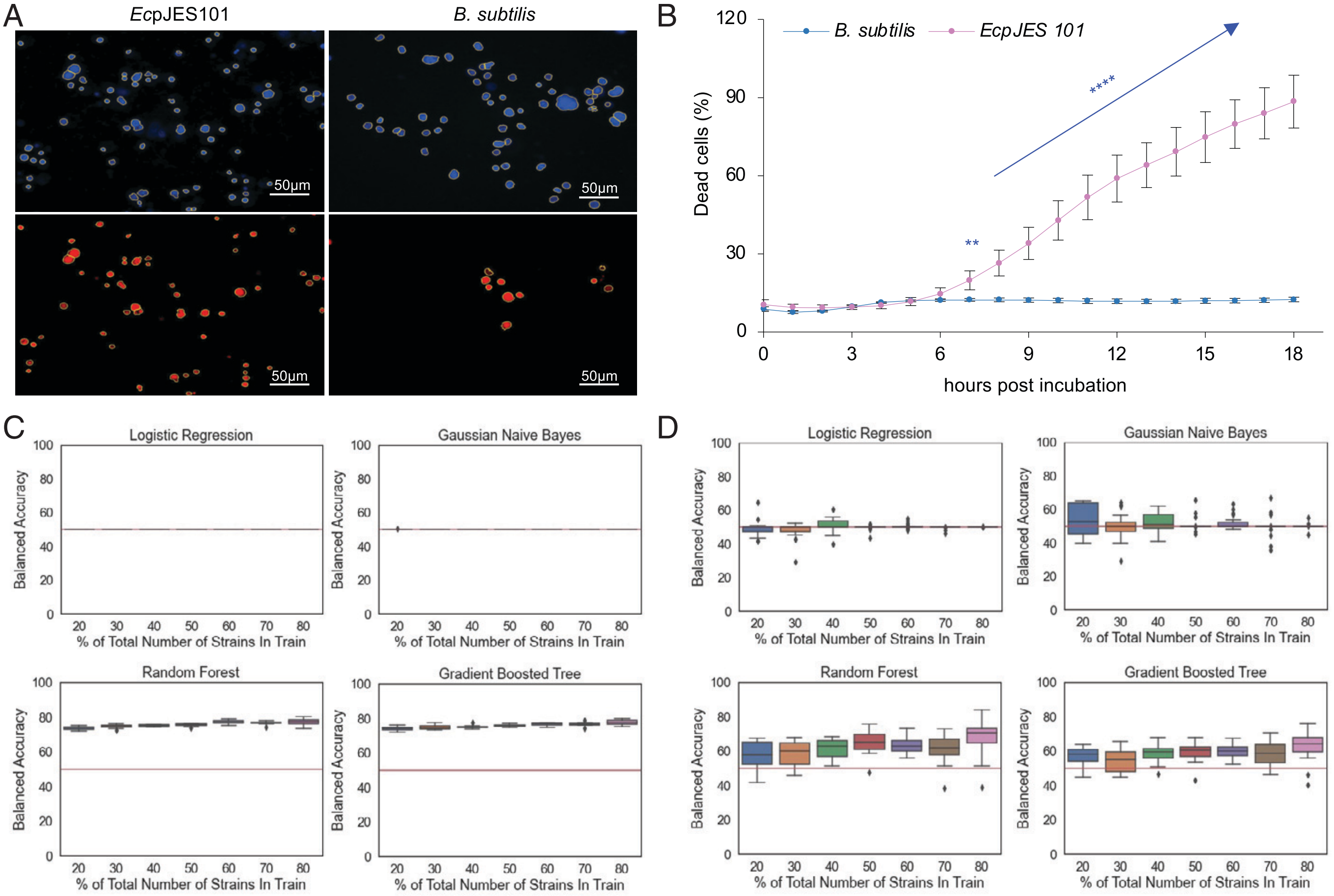

**Fig. 3.** Bacteria-induced host cell toxicity performance in threat assessment by ML models. (*A*) Representative images of toxicity assay for THP1 cells induced by bacteria or Shiga toxin at 18 h post infection/incubation (h.p.i.). Total cells (blue) were counted by Hoechst staining and dead cells (red) by PI staining. Cells were automatically counted and marked with outlines. (Scale bar: 50 μm.) (*B*) Time course of THP1 cell death coincubated with *B. subtilis* or Shiga toxin producing *E. coli (EcpJES 101)* at an MOI of 1 for 18 h.p.i. Data represent the mean ± SD from three independent experiments, each experimental data point $n \geq 9$. Significant difference in adherent bacteria at MOIs of 50 and 100 was observed (** *P* value < 0.001, **** *P* value < 0.0001). (*C* and *D*) Performance of the four indicated ML models in Test 1 (*C*) or Test 2 (*D*) for toxicity assay. All models were characterized to determine the percentage of data required to plateau in performance. Each machine learning algorithm was run 20 times, with the error bars showing the 95% confidence interval from the accuracy scores in each run. The accuracy referred to the percentage of strains assigned correctly by the models.

and *SI Appendix*, Figs. S3*E* and S4*C*). The final prediction by integrating all four assays in T1 achieved an accuracy of 99% (Table 2). For T2, AR, host cell toxicity, adherence assays, and immune activation achieved accuracies of 77%, 76%, 60%, and 66%, respectively (Fig. 5*B* and *SI Appendix*, Figs. S3*F* and S4*C*). The final prediction by integrating all four assays in T2 achieved an accuracy of 85% (Table 2). We also tested different combinations of the four assays and obtained accuracies of 91% ∼98% for T1 and 76% ∼85% for T2 (Table 2). Comparative analysis showed much higher efficacy of threat assessment using integrated assays in PathEngine compared to that of individual assays (Fig. 5 *C* and *D*).

## Discussion

The results from this study show that the development of a multiplexed bacterial phenotyping system using standard laboratory equipment enables assessment of potential bacterial pathogenic threat levels using ML algorithms termed PathEngine. Notably, PathEngine identified the pathogenic potential of *Corynebacterium aurimucosum* (NIST0013), a microbe that classically had been considered a contaminant of clinical microbiological samples but now is appreciated as a possible source of clinically significant bacteremia (36). This demonstrates the power of using PathEngine to assess pathogenic potential and validates our original hypothesis that ML using phenotypic data could assess threat.

PathEngine advances the capability of phenotype-based pathogenicity assessment in multiple aspects. First, the ensemble ML framework quantified the importance of each assay shown in Fig. 5 *C* and *D* and Table 2. In this work, bacterial toxicity and AR were the two most powerful assays with the most balanced accuracy in predictions of potential bacterial pathogenicity. The ability of PathEngine to resolve these nuances in assays suggests that any set of assays could be used, moreover that PathEngine would predict the one that delivers the highest accuracy. Therefore, developing or testing other assays will eventually yield the best sources of data. Second, the weight assignment of each assay is a difficult parameter to tune in ML models. PathEngine automatically weighted weak assays less and strong assays more through the probable outputs from ML models (as shown in Eqs. **1** and **2**). Third, the ML models can be trained beforehand and then deployed to predict new pathogens. The process can be completely automated, whereas traditional phenotypic assessment is analyzed manually.

Despite these advantages, we note that PathEngine does not provide detailed insights on the virulent mechanism of pathogenic infections, as the current main aim is the detection of pathogenic potential of uncharacterized bacteria. Nonetheless, it is possible to enhance this framework by increasing assay diversity. For example,

**Table 1. Kirby-Bauer Disk Diffusion susceptibility testing of a collection of pathogenic and nonpathogenic bacterial strains**

| Zone diameter (mm) | Kanamycin | Ampicillin | Tetracycline | Chloramphenicol | Polymyxin B | Ceftazidime |
|---|---|---|---|---|---|---|
| *PAO1ΔpilA* | 28.3 ± 0.5<br>S | 6.0 ± 0<br>R | 6.0 ± 0<br>R | 6.0 ± 0<br>R | 18.0 ± 0<br>S | 27.7 ± 0.5<br>S |
| *Bm16MΔvjbR* | 56.7 ± 1.7<br>S | 31.0 ± 0.8<br>S | 53.0 ± 0.8<br>S | 39.7 ± 0.8<br>S | 14.0 ± 0.8<br>S | 22.3 ± 0.9<br>S |
| *Pseudomonas fluorescens* | 34.7 ± 0.5<br>S | 6.0 ± 0<br>R | 23.0 ± 0<br>S | 15.0 ± 0<br>I | 6.0 ± 0<br>R | 26.0 ± 0.8<br>S |
| *EHEC-T* | 34.0 ± 0.5<br>S | 22.3 ± 0.5<br>S | 25.0 ± 0<br>S | 27.0 ± 0<br>S | 18.3 ± 0.5<br>S | 31.3 ± 0.9<br>S |
| *EHEC-NT* | 34.0 ± 0.5<br>S | 23.0 ± 0<br>S | 27.0 ± 0<br>S | 29.3 ± 0<br>S | 18.0 ± 0.8<br>S | 31.7 ± 0.5<br>S |
| *SaJE2* | 31.7 ± 0.5<br>S | 24.7 ± 0.5<br>R | 28.3 ± 0.5<br>S | 23.7 ± 0.5<br>S | 12.0 ± 0<br>S | 15.0 ± 0<br>S |
| *SaJE2ΔsaeR* | 32.7 ± 0.5<br>S | 26.0 ± 0.8<br>R | 28.3 ± 0.5<br>S | 21.3 ± 0.5<br>S | 11.0 ± 0<br>I | 14.3 ± 0.5<br>S |
| *PA-14* | 30.3 ± 0.5<br>S | 6.0 ± 0<br>R | 10.7 ± 0.5<br>R | 8.7 ± 0.9<br>R | 18.7 ± 0.5<br>S | 32.3 ± 0.5<br>S |
| *B. subtilis* | 42.3 ± 0.5<br>S | 31.0 ± 0.8<br>S | 20.3 ± 0.5<br>S | 26.7 ± 0.9<br>S | 15.7 ± 0.5<br>S | 20.3 ± 1.2<br>S |
| *S. enterica* | 34.0 ± 0.5<br>S | 26.0 ± 0<br>S | 25.7 ± 0.5<br>S | 28.7 ± 0.5<br>S | 17.7 ± 0.5<br>S | 29.0 ± 0.8<br>S |
| *EcpJES101* | 41.7 ± 0<br>S | 8.0 ± 0<br>R | 29.0 ± 0<br>S | 30.0 ± 0.8<br>S | 19.0 ± 0<br>S | 35.0 ± 0<br>S |
| *EcDH5α* | 42.7 ± 0.8<br>S | 29.0 ± 0.8<br>S | 6.0 ± 0<br>R | 6.0 ± 0<br>R | 20.3 ± 0<br>S | 39.7 ± 0.5<br>S |
| *PAO1* | 27.0 ± 0<br>S | 6.0 ± 0<br>R | 10.7 ± 0.5<br>R | 6.0 ± 0<br>R | 18.0 ± 0<br>S | 30.7 ± 0.5<br>S |

Notes: S, susceptible; R, resistant; I, intermediate refer to the Zone Diameter Interpretive Chart; BD BBL, Sensi-Disk Antimicrobial Susceptibility Test Disk. Data represent the mean ± SD from three independent experiments, each experimental data point $n \geq 3$.

additional reporters from different immune signaling pathways would provide insights into the mechanism of host immune response activation to diverse pathogens. Furthermore, PathEngine's ensemble model can readily handle the additional information collected by the assays.

Previous studies of sequencing-based ML methods to predict bacterial pathogenicity have been reported; for example, PaPrBaG reported an accuracy of 88%, which is similar to the accuracy reported here, and DeePac reported an accuracy of 98% (2, 3). However, PaPrBaG and DeePac, both sequence-based methods for pathogenicity prediction, require data from a corpus of 2,836 and 2,878 bacterial strains (2, 3). In addition, both of these sequence-based methods simulate sequence reads to train their ML models. In contrast, PathEngine, which accounts for biological context, achieved comparable pathogenic predictions using data on only four assays from 40 strains, an order of magnitude fewer strains than sequence-based approaches (*SI Appendix*, Table S3). We therefore expect that the described technology will significantly improve as the size of the training corpus increases. Moreover, the proposed strategy could be very helpful for the early prediction of future epidemics caused by unknown or novel pathogens, especially when combined with NGS-based pathogenicity prediction algorithms. The method established in this study can be applied for other types of biology discovery in which small phenotypic data are available and sequencing data are not required.

## Materials and Methods

**Bacterial Culture Conditions.** All the bacterial strains used in this work are listed in *SI Appendix*, Table S4. Each bacterial strain was grown per appropriate conditions as stated. All of the strains were selected based on the aims to provide a diverse set of BSL-2 bacterial pathogens with respect to Gram stain, source of the strains, route of infection, host tissue tropism, and virulence factors. Virulence factors of particular interest included toxins, adhesins, and antimicrobial resistance. To enhance the ML predictive power, primary pathogens and opportunistic pathogens, as well as nonpathogens, are included.

**Cell Cultures.** A549 cell line (American Type Culture Collection, CCL 185) was maintained in F-12K medium supplemented with 10% fetal bovine serum (FBS) and incubated in 37 °C, 5% $CO_2$. The cells were fed every 3 to 4 d and passaged at 85 to 95% confluency using 0.25% Trypsin-0.53 mM ethylenediaminetetraacetic acid. The cells were then seeded in the vessels using a subcultivation ratio of 1:3 to 1:8. THP1 cells were grown and maintained in RPMI-1640 with glutamine, supplemented with 10% FBS, 1x Hepes, pH 7.0, and 1 mM sodium pyruvate and incubated in 37 °C, 5% $CO_2$. The THP1 cells were split every 2 to 3 d when concentration was proximate to $1 \times 10^6$/mL NF-κB/Jurkat/GFP reporter cell line (System Biosciences, Cat No. TR850A-1) and were grown and maintained in RPMI-1640 medium, 10% FBS, and 2 mM L-glutamine and incubated at 37 °C in 5% $CO_2$. The cells were split every 2 to 3 d to maintain the density 0.5 to $1 \times 10^6$/mL.

**Bacterial Adherence Assay.** A549 cells ($1 \times 10^4$ cell/well) were seeded onto 96-well plates (Corning, 3882) a day prior to the adherence assays. To determine the kinetics of bacterial association, bacteria were cultured to optical density $(OD)_{600}$ of 0.4, then harvested by centrifugation and washed three times using 1x phosphate-buffered saline (PBS; pH 7.4). All bacteria were either GFP tagged or stained by BactoView Red (Biotium 40101). Each bacterial CFU was calculated in advance for estimating the concentration. A549 cells were overlaid with bacterial suspensions with a multiplicity of infection (MOI) of 0, 1, 10, 50, and 100. Each condition had been performed in triple wells in three biological experimental replicates. The bacteria were spun down at 1,000 rpm for 10 min and incubated at 37 °C, 5% $CO_2$ for 1 h. Cells were washed five times using warmed 1x PBS, then fixed for 15 min with 4% formaldehyde at room temperature. The

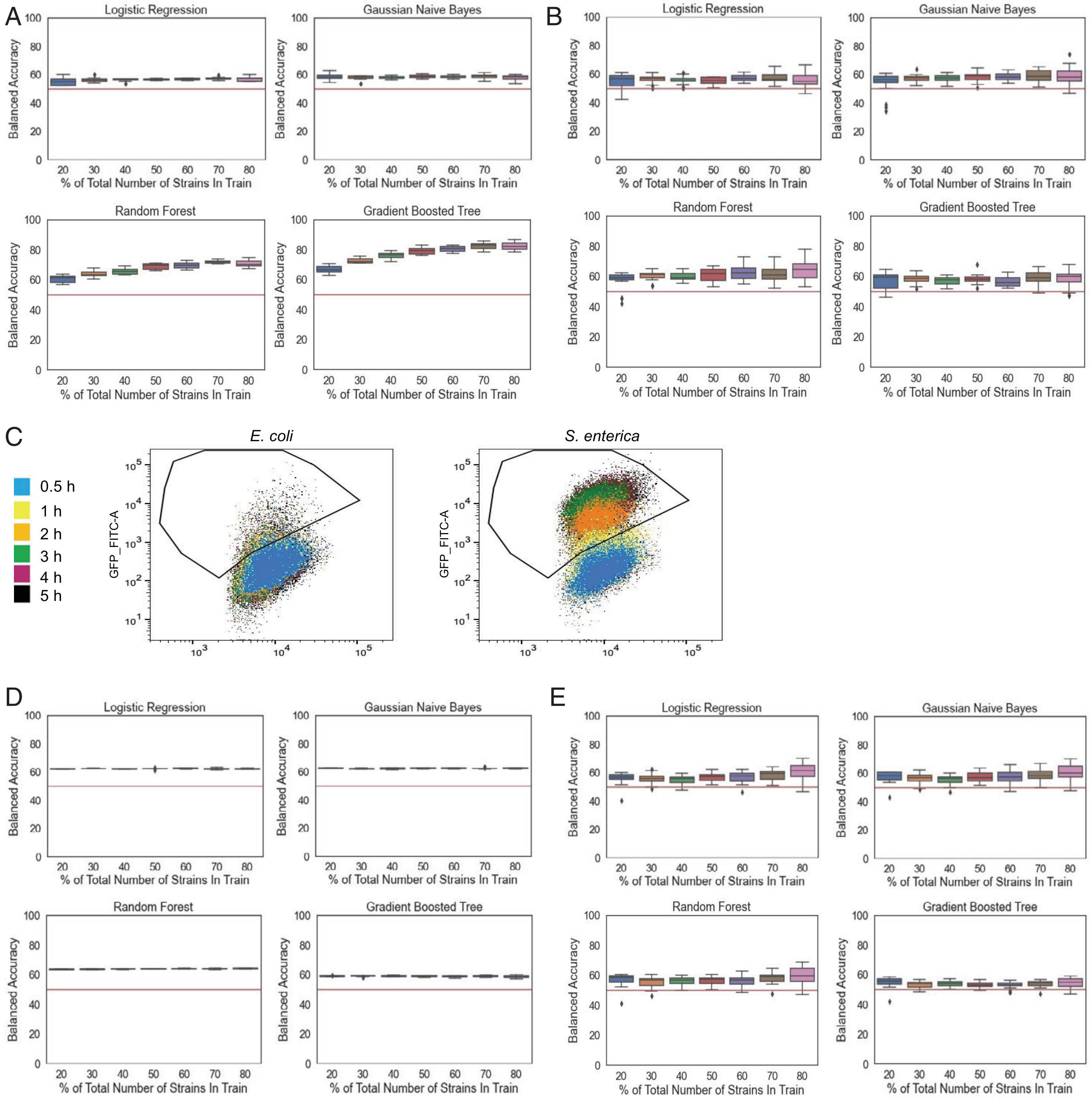

**Fig. 4.** ARs detection and immune activation in the ML models for bacterial threat assessment. (*A* and *B*) Performance of the four ML models in Test 1 (*A*) or Test 2 (*B*) for bacterial AR assays. (*C*) Representative flow cytometry plots of GFP reporter activation induced by *S. enterica* and *E. coli* and the quantification of activated NF-κB/Jurkat/GFP T lymphocyte reporter cells at various hours post infection (h.p.i) at an MOI of 1. GFP signal was measured using BD Fortessa X-20 (FITC: 488-nm laser with bandwidth filter 525/50) at various h.p.i. (*D* and *E*) Performance of the four indicated ML models in Test 1 (*B*) or Test 2 (*C*) for immune activation assay. All models were characterized to determine the percentage of data required to plateau in performance. Each machine learning algorithm was run 20 times, with the error bars showing the 95% confidence interval from the accuracy scores in each run. The accuracy referred to the percentage of strains assigned correctly by the models.

host cells were stained with 50 ng/mL DAPI in the last step. The images were captured and analyzed using BioTek Cytation 5.

**Host Cell Toxicity Assay.** Bacteria were grown overnight and harvested by centrifugation. The bacterial pellet was resuspended in RPMI media without phenol red (Agilent Technologies, Cat. No. 103336–100). THP1 cells ($5 \times 10^5$ cells/well) were seeded onto 96-well plates. Cells were first stained by PI (0.5 μg/mL) (Thermo Fisher Scientific, Cat. No. P3566) and Hoechst (10 μg/mL) (Invitrogen, Cat. No. H3569). Shiga toxin (25 ng/mL) and each bacterium (MOI of 1) were then coincubated with host cells at 37 °C. Images were automatically captured every hour in a total 18-h period. Flow cytometry (BD LSR Fortessa X-20 with 405-, 488-, 561-, and 649-nm lasers) was used to measure the dead cells at PE (phycoerythrin)-Texas Red channel (561-nm laser with bandwidth 610/20 filter). Three technical replicates of each condition in each experiment were performed. The counts of dead cells and total cells were analyzed using BioTek Cytation 5.

**Immune Activation Assay.** All bacterial cultures were centrifuged at 10,000 rpm for 2 min, and the bacterial pellet was resuspended in 1 mL RPMI media. Bacterial concentration was estimated by measuring $OD_{600}$. One million Jurkat reporter cells were incubated with bacteria at an MOI of 1 or 10 for 0.5, 1, 2, 3,

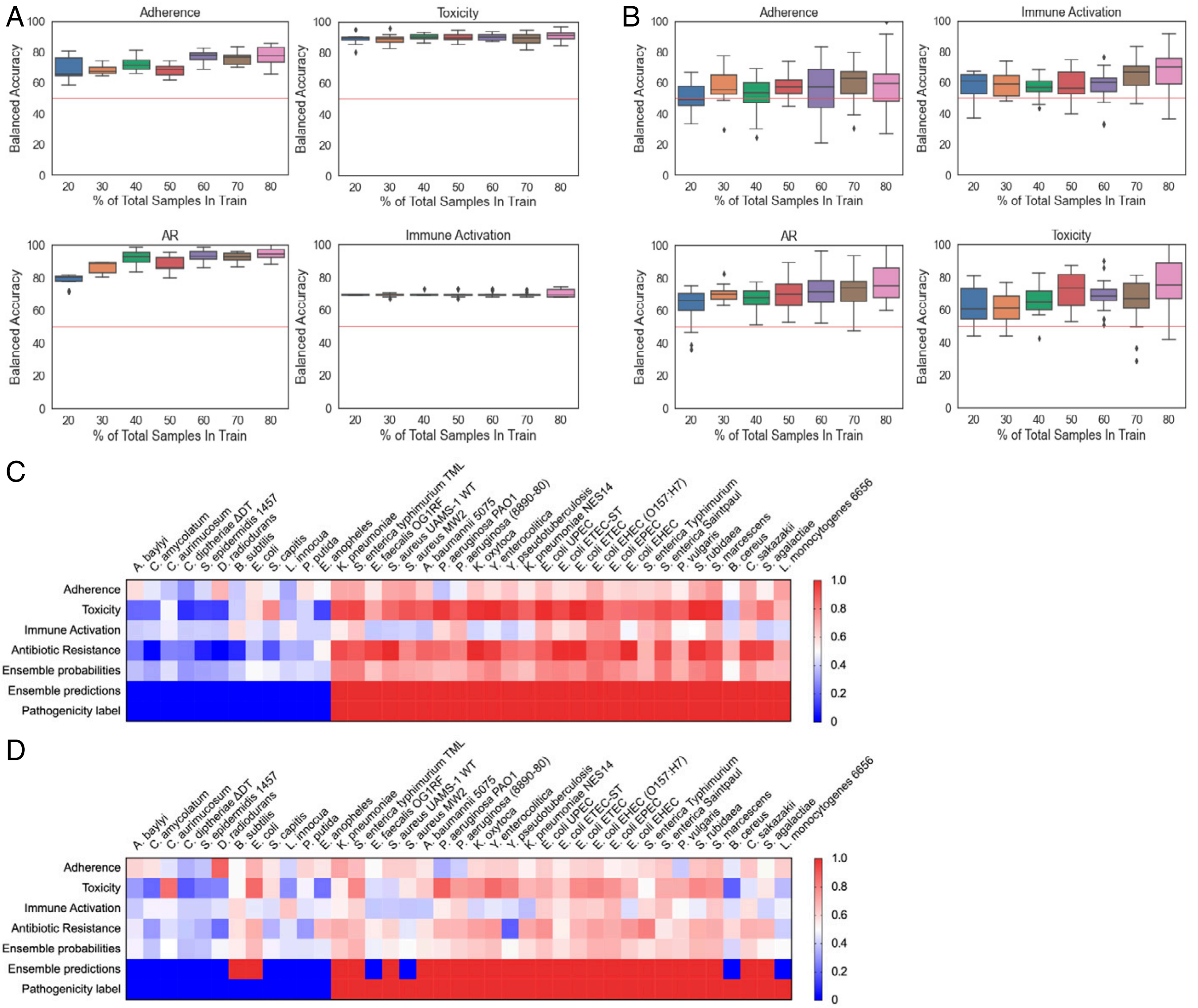

**Fig. 5.** The ensemble ML model of PathEngine improves the accuracy of threat assessment. (*A* and *B*) Aggregated performance for all four phenotypic assays, bacterial adherence, host immune activation, AR, and bacterial toxicity in PathEngine prediction for Test 1 (*A*) and in Test 2 (*B*). Each machine learning algorithm was run 20 times, with the error bars showing the 95% confidence interval from the accuracy scores in each run. (*C* and *D*) The observations from each strain and each phenotypic assay were aggregated to make one prediction per strain in the PathEngine ensemble model. The accuracy was estimated by comparing the actual threat status and the predicted threat status for each strain. Bacterial pathogenic potential was quantified by individual assays and the ensemble assay in Test 1 (*C*) and Test 2 (*D*). Pathogenic scores obtained from ML predictions for each assay between 0 (blue) and 1 (red). 0 represents a strong nonpathogen, and 1 represents a strong pathogen. The ensemble probabilities show when the pathogenic scores from all four assays are ensembled together. Ensemble predictions convert the ensemble probabilities to 0 or 1 with a cutoff at 0.5 for comparing to the pathogenicity label in the last column.

4, and 5 h. Cells were then stained using LIVE/DEAD Fixable Dead Cell Stain Kit (Invitrogen, Cat. No. L34960) for flow cytometry detection. The cells were eventually fixed using 4% formaldehyde and washed by PBS supplemented with 1% bovine serum albumin (BSA) to preserve samples for flow cytometry measurement. NF-κB/GFP reporter–activated cells were detected using fluorescein isothiocyanate (FITC) (488-nm laser with bandwidth filter 525/50), while live/dead cells were detected using PE-Texas Red channels (561-nm laser with bandwidth 610/20 filter).

**AR Assay.** Six antibiotics were tested, including kanamycin 10 μg (BD, No. 316424), ampicillin 10 μg (BD, No. 231264), tetracycline 30 μg (BD, No. 231344), chloramphenicol 30 μg (BD, No. 231274), polymyxin B 300 units (BD, No. 231324), and ceftazidime 30 μg (BD, No. 231633) using Kirby-Bauer Disk Diffusion susceptibility testing methods accepted by CLSI (Clinical and Laboratory Standards Institute) in the Texas A&M Veterinary Medical Laboratory. All bacteria were initially grown on Brain Heart Infusion (BHI) agar plates (Hardy Diagnostics, A20). Single colonies were inoculated to BHI broth and grown to achieve turbidity of 0.5 McFarland standard. Organisms that did not require blood for growth were tested using plain Mueller-Hinton agar, and fastidious organisms that required blood were tested using Mueller-Hinton agar with 5% blood. Sterile cotton swabs were used to inoculate bacteria onto agar plates. Appropriate agar plates were used for different organisms. The entire plate surface was covered by the bacterial inocula. Plates were air-dried for 15 min at room temperature, and then antibiotic discs were stamped onto the agar surface. All plates were incubated at 37 °C overnight, and the diameters of the zones of inhibition were measured after overnight incubation using calipers and interpreted as susceptible, intermediate, or resistant to each antimicrobial drug according to CLSI recommendations (34, 37). All experiments had three independent biological replicates.

**Generation of ML Models and Performance of the ML.** Four ML models, including the GBT, LR, RF, and GNB, were selected to generate ML. An LR is a generalized linear model whose parameters learn a hyperplane that separates the classes (pathogen/nonpathogen) based on the sum of the inputs and parameters. The GNB learns latent variables, while RFs and GBT learn decision trees that best inform the separation between pathogen and nonpathogens and are

**Table 2. Aggregated accuracy, precision, recall, and F1 scores for each strain by aggregating across assays for Test 1 and Test 2**

| | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| **Test 1** | | | | |
| All four assays | 99.0 | 99.0 | 99.0 | 99.0 |
| Adherence + Toxicity + AR | 98.0 | 98.0 | 98.0 | 98.0 |
| Adherence + AR | 97.0 | 96.0 | 100.0 | 98.0 |
| Adherence + Toxicity | 91.0 | 91.0 | 96.0 | 94.0 |
| Toxicity + AR | 98.0 | 98.0 | 100.0 | 99.0 |
| **Test 2** | | | | |
| All four assays | 85.0 | 91.0 | 87.0 | 89.0 |
| Adherence + Toxicity + AR | 85.0 | 88.0 | 91.0 | 89.0 |
| Adherence + AR | 80.0 | 84.0 | 89.0 | 86.0 |
| Adherence + Toxicity | 76.0 | 79.0 | 90.0 | 84.0 |
| Toxicity + AR | 83.0 | 90.0 | 86.0 | 88.0 |

Notes: The four assays include bacterial adherence, bacterial toxicity, AR, and host cell immune activation. accuracy, the ratio of correct predictions to all predictions; precision, the ratio of correctly predicted pathogens to the total predicted pathogens; recall, the ratio of correctly predicted pathogens to the total pathogens present; F1, the weighted average of precision and recall.

thus more complex than the LR model. We noted that the metric of evaluation used in these studies was balanced accuracy, which ensured that predictions of the majority class were penalized. The framework was evaluated with two types of tests:

1. Test 1: The ability of the ML models to predict pathogenic potential from microbes in its corpus.
2. Test 2: The ability of the ML models to predict pathogenic potential for microbes not in its corpus.

In T1, independent runs of each assay for an organism were held out to form the test corpus with no overlap of the data points in the train and test sets, while in T2, all assays from a set of organisms were held out to form the test corpus with no overlap of strains in the train and test sets. Both of these tests were necessary, as each test measures an aspect of deployment of this framework in a real-world setting. Both tests swept the size of a training corpus by using 20 to 80% of the observations (Test 1) or strains (Test 2) for training and the rest for testing (Fig. 1*C*) to provide insights into how many individual observations or bacterial strains needed to be evaluated in phenotypic tests before each model's performance statistically plateaued. The predictions per observation were then statistically aggregated to make a prediction per microbe per phenotypic assay, and finally, a prediction per microbe across all phenotypic assays.

The predictions per assay were evaluated for their ability to predict pathogenic potential of each observation. An example of an observation for the host cell immune activation assay was each measured cellular event that passed gating. We called these predictions observation-level predictions. These observation-level predictions needed to be aggregated to provide a score for the pathogenic potential for each microbe. This framework could be thought of as combining well-known ML frameworks: 1) self-supervised learning and 2) weakly supervised learning. In self-supervised learning, a pretext task was used to train a model on a related but independent objective, and that model was then used to make predictions for the true objective. In this case, the true objective was the prediction of pathogenic potential of each microbe, while the related independent objective was a prediction for each observation. In weakly supervised learning, the data's features and associated labels were *inexact*; namely, each observation was given a coarse-grained label. As an example, for the toxicity assay, a pathogen may not have exhibited pathogenic properties in the earlier hours of a time course of analysis, yet our framework still labeled those time points as associated with a pathogenic phenotype. Concepts from these two learning frameworks were combined to generate predictions at both an assay level, where we would compute the pathogenic potential of a microbe per assay, or across assays, from which a prediction of pathogenic potential of each microbe could be reached. In both cases, we used a weighted average of the predictions of each event (the pretext task), where the weights were the confidence in the prediction of the ML model per observation. Mathematically, this was described per assay Eq. **1** and across assays Eq. **2**:

$$p_{m,a} = \frac{1}{N}\sum_{i=0}^{N} w_{m,a,i} * P_{m,a,i} \quad [1]$$

$$p_m = \frac{1}{N}\sum_{a=0}^{A} w_{m,a} * P_{m,a} \quad [2]$$

where $p_m$ ($p_{m,a}$) is the pathogenic potential of microbe $m$ (in assay $a$); $w_{m,a}$ ($w_{m,a,i}$) is the model's confidence in its prediction of microbe $m$ in assay $a$ ($i^{th}$ observation); and $P_{m,a}$ ($P_{m,a,i}$) is a 1 if the model's prediction of microbe $m$ in assay $a$ ($i^{th}$ observation) is a pathogen and 0 if not; $N$ is the total number of observations; and $A$ is the total number of assays. This formula provided an extensible framework to both allow each assay to make its own unique set of observations, as well as to account for additional assays as they became available. A threshold could be set on the threat assessment based on the calculated pathogenic potential . In this work, we selected the best performing model from each assay to use in these aggregations. Eq. **1** provided the additional opportunity to determine the ability of any particular assay to predict pathogenic potential.

All ML models were developed with Python 3.7 using the Pandas and Scikit-learn libraries, with all plots visualized using seaborn. The phenotype assays were parsed and integrated using proprietary software called the Active Discovery Engine. The software platform uses a set of user-defined rules to automatically extract metadata from the data sources to integrate with the experimental data to generate a wrangled data frame for ML. Finally, models were compared using an open source Python-based test harness that evaluates the performance of ML models (38).

**Statistical Analysis.** The mean and SD were calculated from triplicates of each data point of all four assays, adherence, toxicity, immune activation, and AR. Two-way analyses of variance (ANOVAs) were performed to test for significant variation between data points across treatments for three independent experiments. Tukey's multiple comparisons test was used for pairwise comparisons of the significance of each data point between treatments. Two-way ANOVAs and Tukey's multiple comparisons test were performed using Prism 8 version 8.4.2.

**Data Availability.** Raw data, ML algorithms, and analysis of biological assays data have been deposited in GitHub (https://github.com/netrias/PathEngine/tree/master/data_files and https://github.com/netrias/PathEngine).

**ACKNOWLEDGMENTS.** The authors are grateful to the Uniformed Services University, US Food and Drug Administration, US Department of Agriculture, and Michigan State University Shiga-toxin producing *E. coli* (STEC) Center for their contributions to the NIST strains collection. We also thank our other lab members Elizabeth Bustamante, Kelsey Wells, and Ana Cabello for their support. Also, we thank Kalika M. Landua, Steve Fullwood, and Jonathan Ekman of Nikon Instruments Inc. for their support. This work was supported by the Department of Defense under Contract No. W911NF1920013 to A.R.W., J.E.S., A.H., and P.d.F. This ML material is based upon work supported by the Defense Advanced Research Projects Agency and the Department of Interior under Contract No. 140D6319C0029 to M.E., M.W., J.E.S., and P.d.F. The content of the information does not necessarily reflect the position or the policy of the government, and no official endorsement should be inferred. Certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Author affiliations: [a]Department of Microbial Pathogenesis and Immunology, Texas A&M Health Science Center, Bryan, TX 77807; [b]Netrias, LLC, Cambridge, MA 02142; [c]Biocomplexity Institute and Initiative, University of Virginia, Charlottesville, VA 22904; [d]Complex Microbial Systems Group, Biomaterials and Biosystems Division, Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899; [e]Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843; [f]Department of Veterinary Pathobiology, Texas A&M University, College Station, TX 77843; and [g]Department of Biomedical Engineering, Texas A&M University, College Station,TX 77843

Author contributions: J.Y., M.E., M.D., D.Z., J.E.S., E.J.v.S., and P.d.F. designed research; J.Y., M.E., Y.-P.C., M.D., D.Z., S. C, A.-J.R., J.K., S.C, H.P., O.O., and J.M. performed research; M.E., A.V., K.F., R.K.M., R.C.A., S. S, and J.D. contributed new reagents/analytic tools; J.Y., Y.-P.C., M.W., A.R.W., A.D., C.C., and A.H. analyzed data; J.Y. and Q.-M.Q. wrote the paper; and R.K.M. flow cytometry technician.

1. H. Hosainzadegan, R. Khalilov, P. Gholizadeh, The necessity to revise Koch's postulates and its application to infectious and non-infectious diseases: A mini-review. *Eur. J. Clin. Microbiol. Infect. Dis.* **39**, 215–218 (2020).
2. C. Deneke, R. Rentzsch, B. Y. Renard, PaPrBaG: A machine learning approach for the detection of novel pathogens from NGS data. *Sci. Rep.* **7**, 39194 (2017).
3. J. M. Bartoszewicz, A. Seidel, R. Rentzsch, B. Y. Renard, DeePaC: Predicting pathogenic potential of novel DNA with reverse-complement neural networks. *Bioinformatics* **36**, 81–89 (2020).
4. X. Zhao, N. Wang, An interpretable machine learning method for detecting novel pathogens. *Res. Sq.* [Preprint] (2020). https://www.researchsquare.com/article/rs-11084/v2. Accessed 16 March 2020.
5. L. de Nies *et al.*, PathoFact: A pipeline for the prediction of virulence factors and antimicrobial resistance genes in metagenomic data. *Microbiome* **9**, 49 (2021).
6. B. J. Ampattu *et al.*, Transcriptomic buffering of cryptic genetic variation contributes to meningococcal virulence. *BMC Genomics* **18**, 282 (2017).
7. F. Balloux *et al.*, From theory to practice: Translating whole-genome sequencing (WGS) into the clinic. *Trends Microbiol.* **26**, 1035–1048 (2018).
8. C. Roe *et al.*, Bacterial genome wide association studies (bGWAS) and transcriptomics identifies cryptic antimicrobial resistance mechanisms in *Acinetobacter baumannii. Front. Public Health* **8**, 451 (2020).
9. J. P. Allen, E. Snitkin, N. B. Pincus, A. R. Hauser, Forest and trees: Exploring bacterial virulence with genome-wide association studies and machine learning. *Trends Microbiol.* **29**, 621–633 (2021).
10. M. Re, G. Valentini, "Ensemble methods: A review" in *Advances in Machine Learning and Data Mining for Astronomy*, M. J. Way, J. D. Scargle, K. M. Ali, A. N. Srivastava, Eds. (Routledge, 2012), pp. 563–594.
11. J. Pizarro-Cerdá, P. Cossart, Bacterial adhesion and entry into host cells. *Cell* **124**, 715–727 (2006).
12. B. Sadowska *et al.*, Characteristics of *Staphylococcus aureus*, isolated from airways of cystic fibrosis patients, and their small colony variants. *FEMS Immunol. Med. Microbiol.* **32**, 191–197 (2002).
13. A. J. Carterson *et al.*, A549 lung epithelial cells grown as three-dimensional aggregates: Alternative tissue culture model for *Pseudomonas aeruginosa* pathogenesis. *Infect. Immun.* **73**, 1129–1140 (2005).
14. E. Chi, T. Mehl, D. Nunn, S. Lory, Interaction of *Pseudomonas aeruginosa* with A549 pneumocyte cells. *Infect. Immun.* **59**, 822–828 (1991).
15. X. Liang *et al.*, Inactivation of a two-component signal transduction system, SaeRS, eliminates adherence and attenuates virulence of *Staphylococcus aureus*. *Infect. Immun.* **74**, 4655–4665 (2006).
16. J. Yang, Q. M. Qin, P. de Figueiredo, Automated, high-throughput detection of bacterial adherence to host cells. *J. Vis. Exp.* **175**, 10.3791/62764 (2021).
17. T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system. *arXiv* [Preprint] (2016). https://arxiv.org/abs/1603.02754. Accessed 10 June 2016.
18. J. Nelder, R. Wedderburn, Generalized linear models. *J. R. Stat. Soc. [Ser A]* **135**, 370–384 (1972).
19. L. Breiman, Random forests. *Mach. Learn.* **45**, 5–32 (2001).
20. H. Zhang, "The optimality of Naive Bayes" in *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference*, V. Barr, Z. Markov, Eds. (AAAI Press, Menlo Park, CA, 2004), pp. 562–567.
21. A. Fàbrega, J. Vila, *Salmonella enterica* serovar Typhimurium skills to succeed in the host: Virulence and regulation. *Clin. Microbiol. Rev.* **26**, 308–341 (2013).
22. M.-S. Lee *et al.*, Shiga toxins induce autophagy leading to differential signalling pathways in toxin-sensitive and toxin-resistant human cells. *Cell. Microbiol.* **13**, 1479–1496 (2011).
23. M.-S. Lee, S. Koo, D. G. Jeong, V. L. Tesh, Shiga toxins as multi-functional proteins: Induction of host cellular stress responses, role in pathogenesis and therapeutic applications. *Toxins (Basel)* **8**, 77 (2016).
24. M. Genin, F. Clement, A. Fattaccioli, M. Raes, C. Michiels, M1 and M2 macrophages derived from THP-1 cells differentially modulate the response of cancer cells to etoposide. *BMC Cancer* **15**, 577 (2015).
25. E. Valle, D. G. Guiney, Characterization of *Salmonella*-induced cell death in human macrophage-like THP-1 cells. *Infect. Immun.* **73**, 2835–2840 (2005).
26. N. Pick, S. Cameron, D. Arad, Y. Av-Gay, Screening of compounds toxicity against human monocytic cell line-THP-1 by flow cytometry. *Biol. Proced. Online* **6**, 220–225 (2004).
27. C. E. Krämer, W. Wiechert, D. Kohlheyer, Time-resolved, single-cell analysis of induced and programmed cell death via non-invasive propidium iodide and counterstain perfusion. *Sci. Rep.* **6**, 32104 (2016).
28. V. L. Tesh, Induction of apoptosis by Shiga toxins. *Future Microbiol.* **5**, 431–453 (2010).
29. R. Parboosing, G. Mzobe, L. Chonco, I. Moodley, Cell-based assays for assessing toxicity: A basic guide. *Med. Chem.* **13**, 13–21 (2016).
30. A. Beceiro, M. Tomás, G. Bou, Antimicrobial resistance and virulence: A successful or deleterious association in the bacterial world? *Clin. Microbiol. Rev.* **26**, 185–230 (2013).
31. M. Schroeder, B. D. Brooks, A. E. Brooks, The complex relationship between virulence and antibiotic resistance. *Genes (Basel)* **8**, 39 (2017).
32. W. L. Drew, A. L. Barry, R. O'Toole, J. C. Sherris, Reliability of the Kirby-Bauer disc diffusion method for detecting methicillin-resistant strains of *Staphylococcus aureus*. *Appl. Microbiol.* **24**, 240–247 (1972).
33. Z. A. Khan, M. F. Siddiqui, S. Park, Current and emerging methods of antibiotic susceptibility testing. *Diagnostics (Basel)* **9**, 49 (2019).
34. X. Qin, S. J. Weissman, M. F. Chesnut, B. Zhang, L. Shen, Kirby-Bauer disc approximation to detect inducible third-generation cephalosporin resistance in Enterobacteriaceae. *Ann. Clin. Microbiol. Antimicrob.* **3**, 13 (2004).
35. Z. Zhou, A brief introduction to weakly supervised learning. *Natl. Sci. Rev.* **5**, 44–53 (2018).
36. A. M. Z. Evangelia *et al.*, Clinically-significant *Corynecterium aurimucosum* bacteremia: Is it time for a change of perspective? A first documented case and review of the literature. *J. Case Rep. Images Infect. Dis.* **1**, 100003Z16EZ2018 (2018).
37. J. Hudzicki, *Kirby-Bauer Disk Diffusion Susceptibility Test Protocol* (ASM, 2009).
38. H. Eramian, M. Eslami, Test harness to compare machine learning models. https://github.com/SD2E/test-harness. Accessed 4 February 2022.