EDITOR'S CHOICE

Same-day antimicrobial susceptibility test using acoustic-enhanced flow cytometry visualized with supervised machine learning

Timothy J. J. Inglis[1,2,3,*], Teagan F. Paton[3], Malgorzata K. Kopczyk[3], Kieran T. Mulroney[1,2,4] and Christine F. Carson[2]

Abstract

Purpose. Antimicrobial susceptibility is slow to determine, taking several days to fully impact treatment. This proof-of-concept study assessed the feasibility of using machine-learning techniques for analysis of data produced by the flow cytometer-assisted antimicrobial susceptibility test (FAST) method we developed.

Methods. We used machine learning to assess the effect of antimicrobial agents on bacteria, comparing FAST results with broth microdilution (BMD) antimicrobial susceptibility tests (ASTs). We used *Escherichia coli* (1), *Klebsiella pneumoniae* (1) and *Staphylococcus aureus* (2) strains to develop the machine-learning algorithm, an expanded panel including these plus *E. coli* (2), *K. pneumoniae* (3), *Proteus mirabilis* (1), *Pseudomonas aeruginosa* (1), *S. aureus* (2) and *Enterococcus faecalis* (1), tested against FAST and BMD (Sensititre, Oxoid), then two representative isolates directly from blood cultures.

Results. Our data machines defined an antibiotic-unexposed population (AUP) of bacteria, classified the FAST result by antimicrobial concentration range, and determined a concentration-dependent antimicrobial effect (CDE) to establish a predicted inhibitory concentration (PIC). Reference strains of *E. coli, K. pneumoniae* and *S. aureus* tested with different antimicrobial agents demonstrated concordance between BMD results and machine-learning analysis (CA, categoric agreement of 91 %; EA, essential agreement of 100 %). CA was achieved in 35 (83 %) and EA in 28 (67 %) by machine learning on first pass in a challenge panel of 27 Gram-negative and 15 Gram-positive ASTs. Same-day AST results were obtained from clinical *E. coli* (1) and *S. aureus* (1) isolates.

Conclusions. The combination of machine learning with the FAST method generated same-day AST results and has the potential to aid early antimicrobial treatment decisions, stewardship and detection of resistance.

INTRODUCTION

Antimicrobial susceptibility tests (ASTs) are the common point of entry to optimized antimicrobial therapy, antimicrobial resistance surveillance and new antimicrobial agent discovery. ASTs classify the interaction between bacteria and antimicrobial agents, most often expressed in a binary form as sensitive or resistant. But in some instances, antimicrobial

susceptibility needs to be expressed as a quantitative measurement known as the MIC. The current international AST reference method is the broth microdilution (BMD) version of the MIC [1], which takes 18–24 h after primary isolation of the causal bacteria from the initial specimen culture. The time this takes prevents AST from informing the earliest antimicrobial therapy decisions in serious life-threatening

infections [2]. Concerns about increasing antimicrobial resistance (AMR) have prompted a search for faster methods of accurate AST [3].

We previously reported a flow-cytometer method of antimicrobial susceptibility testing (FAST) that generated sensitive/resistant and MIC results with a combination of high speed and accuracy [4]. The data sets generated by the FAST method are too complex for efficient analysis by current flow-cytometer interpretive software originally designed to meet the needs of eukaryotic cell biologists. Previous analytical methods held back expansion of our FAST repertoire to a wider range of antimicrobial/bacterial combinations [5]. While flow cytometry had been used previously to study antimicrobial susceptibility [6–8], it has not been widely adopted for AST. Our flow-cytometry-assisted susceptibility test (FAST) method generated an accurate MIC in less than 3 h [4], but only after manual curation and analysis of data output by an experienced cytometrist using commercial software. A more efficient data-processing pipeline became essential after we modified the FAST sample preparation for high throughput analysis of multiple antimicrobial/bacteria combinations. The improved workflow generated larger numbers of opto-electronic events from each AST experiment. Moreover, the original description of the FAST method used a two-laser, eight-channel acoustic-enhanced flow cytometer. The high throughput workflow employs a four-laser, 16-channel acoustic cytometer equipped with a 96-well plate autosampler, highlighting the need for swift, objective data processing by a less experienced operator. We therefore sought a data-processing solution through machine learning. Here we describe the initial development of a supervised machine-learning ensemble for visualization, classification and analysis of antimicrobial susceptibility, and demonstrate its application in accelerated antimicrobial susceptibility testing of bacteria isolated from time-critical clinical specimens and a broader challenge panel of isolates.

METHODS

Software

Data machines were designed, assembled and run using open access data-mining software (Orange v3.20, University of Ljubljana, Slovenia) [9]; under a Creative Commons license. Orange was run under Windows 10 (Microsoft, CA, USA). Statistical analysis was conducted in Prism v8 (GraphPad, San Diego, CA, USA).

Bacterial strains

Escherichia coli ATCC 25922, *E. coli* ATCC 35218, *E. coli* −2841 (clinical), *Klebsiella pneumoniae* ATCC 700603, *K. pneumoniae* ATCC 700603, *K. pneumoniae* ATCC BAA-1705, *K. pneumoniae* ATCC BAA-1706, *K. pneumoniae* ATCC 13883, *Proteus mirabilis* −9545 (clinical), *Pseudomonas aeruginosa* ATCC 27853, *Staphylococcus aureus* ATCC 25923, *S. aureus* ATCC 29213, *S. aureus* ATCC 33592, *S. aureus* −6885 (clinical), *Enterococcus faecalis* ATCC 29212. Clinical isolates: *E. coli* 1A and *S. aureus* 9B. Curated by Western Australian

Culture Collection, Department of Microbiology, PathWest Laboratory Medicine WA. Source of clinical isolates, Department of Microbiology, PathWest Laboratory Medicine WA.

Antimicrobial susceptibility series

We used commercial, custom, pre-dispensed 96-well microtitre plates containing dilution series of antimicrobial agents for standardized broth microdilution MIC for (a) Enterobacterales (SEMPA1, Sensititre, Oxoid, UK) and (b) Gram-positive cocci (SEMSE3, Sensititre, Oxoid, UK). The plates contained the following antibiotic dilution series bracketing the European Committee on Antimicrobial Susceptibility Testing (EUCAST) break points:

> SEMPA1, containing amikacin (0.5–32 µg ml⁻¹), aztreonam (1–64 µg ml⁻¹), ciprofloxacin (0.12–8 µg ml⁻¹), colistin (0.25–32 µg ml⁻¹), cefepime (1–32 µg ml⁻¹), gentamicin (0.25–16 µg ml⁻¹), imipenem (0.25–32 µg ml⁻¹), levofloxacin (0.12–8 µg ml⁻¹), meropenem (0.12–32 µg ml⁻¹), piperacillin-tazobactam (1/4-64/4 µg ml⁻¹), trimethoprim-sulphamethoxazole (0.12/2.38-16/304 µg ml⁻¹), ceftazidime (0.5–32 µg ml⁻¹) and tobramycin (0.25–16 µg ml⁻¹).

> SEMSE3, containing amikacin (2–64 µg ml⁻¹), azithromycin (0.5–8 µg ml⁻¹), ciprofloxacin (0.12–8 µg ml⁻¹), clarithromycin (0.25–8 µg ml⁻¹), clindamycin (0.12–4 µg ml⁻¹), cefoxitin (1–16 µg ml⁻¹), gentamicin (0.12–4 µg ml⁻¹), levofloxacin (0.12–8 µg ml⁻¹), linezolid (0.5–16 µg ml⁻¹), moxifloxacin (0.06–4 µg ml⁻¹), norfloxacin (1–16 µg ml⁻¹), ofloxacin (0.25–4 µg ml⁻¹), penicillin (0.03–0.5 µg ml⁻¹), teicoplanin (0.25–16 µg ml⁻¹), tobramycin (0.12–4 µg ml⁻¹) and vancomycin (0.5–16 µg ml⁻¹).

Bacterial analysis: a modified version of the original FAST method was used [4]. In brief, an acoustic flow cytometer (Attune NxT, ThermoFisher Scientific, Eugene, OR, USA) was coupled to a 96-well autosampler to generate well-by-well analysis of SYTO9 nucleic acid intercalating dye (ThermoFisher Scientific, Eugene, OR, USA) -stained bacterial cells after co-incubation with a series of increasing concentrations of the antimicrobial agents listed above. Challenge bacteria were processed as follows: two to three single bacterial colonies were picked from blood agar (5 % horse blood agar, Excel Laboratory Products, Western Australia) and resuspended in Sensititre de-ionized water to achieve a turbidity of 0.5 McFarland standard, as indicated on the Sensititre nephelometer by its central green bar. Then, 55 µl of this suspension was inoculated into a Sensititre dosage tube containing 11 ml cation-adjusted Mueller–Hinton broth (ThermoFisher Scientific, Lenexa, KS, USA) at 1:200 dilution to achieve a notional suspension density of 5×10^5 cells ml⁻¹. The dosage tube was incubated without shaking for 1 h at 35.5 °C. The dosage tube was mixed thoroughly, loaded into the auto-inoculator (AIM, Oxoid, UK), and the contents used to inoculate a EUCAST pattern Sensititre plate (SEMPA1 Enterobacterales, SEMSE3 Gram-positives) with an inoculum volume of 100 µl. The 96-well plate was sealed and incubated without shaking for 1 h (Enterobacterales), or 3 h (Gram-positives). We then added

15 ml of Hank's balanced salt solution (HBSS, Excel Laboratory Products, WA) and 20 µl of 5 mM SYTO9 to an empty dosage tube. Using the AIM, 150 µl of the stain and HBSS was dispensed into each well of a standard format 96-well plate (Nunclon delta surface). Next, 50 µl of bacterial culture from each well of the Sensititre plate was added to the respective well on the 96-well plate containing stain and HBSS, to a total volume of 200 µl per well. The remaining 50 µl in the plate was re-sealed, and used as a comparator for FAST. The plate with stained contents was incubated at ambient temperature in darkness, with 400 r.p.m. shaking, for 8 min. The 96-well plate was then inserted into the flow cytometer auto-sampler (Attune, ThermoFisher, Eugene, OR, USA) for data acquisition by the flow cytometer.

Data generation for AST

Data were generated according to the Data File Standard for Flow Cytometry, Version 3.1 [10] by a single acoustic flow cytometer (Attune NxT, Life Technologies, ThermoFisher Scientific, Eugene, OR, USA) equipped with four lasers (405, 488, 561 and 637 nm) and 16 analysis channels [forward scatter, FSC; side scatter, SSC; violet laser (VL) channel 1, VL1 – 440/50, VL2 – 512/25, VL3 – 603/48, VL4 – 710/50, blue laser (BL) channel 1, BL1 – 530/30; BL2 – 590/40, BL3 – 694/40; yellow laser (YL) channel 1, YL1 – 585/16, YL2 – 620/15, YL3 695/40, YL4 – 780/60; red laser (RL) channel 1, RL1 – 670/14, RL2 – 720/30, RL3 – 780/60] with minor adjustment to the FSC detector to improve small particle resolution.

Data handling

FAST data sets were obtained as flow-cytometer data files and converted into comma-separated variable (CSV) format, then transferred without editing to a stand-alone computer (Legion, Lenovo) for data analysis with the data machines described below (Fig. 1).
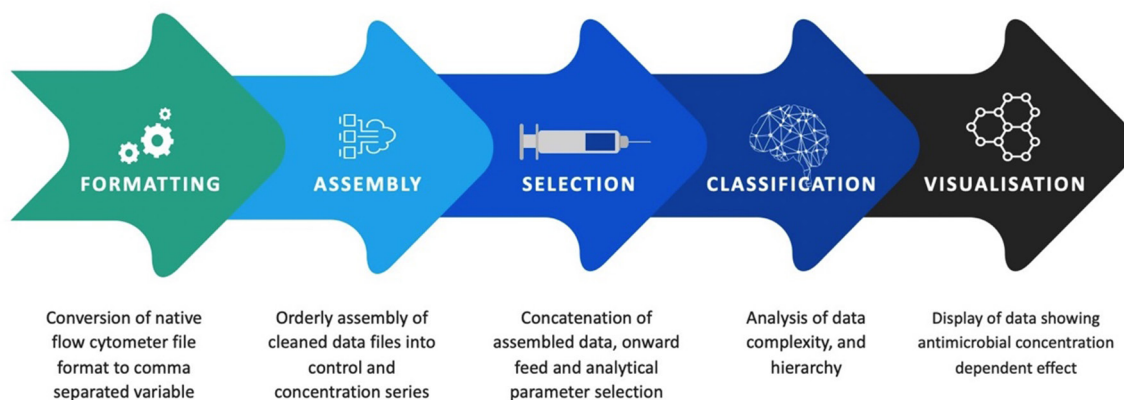
Step one

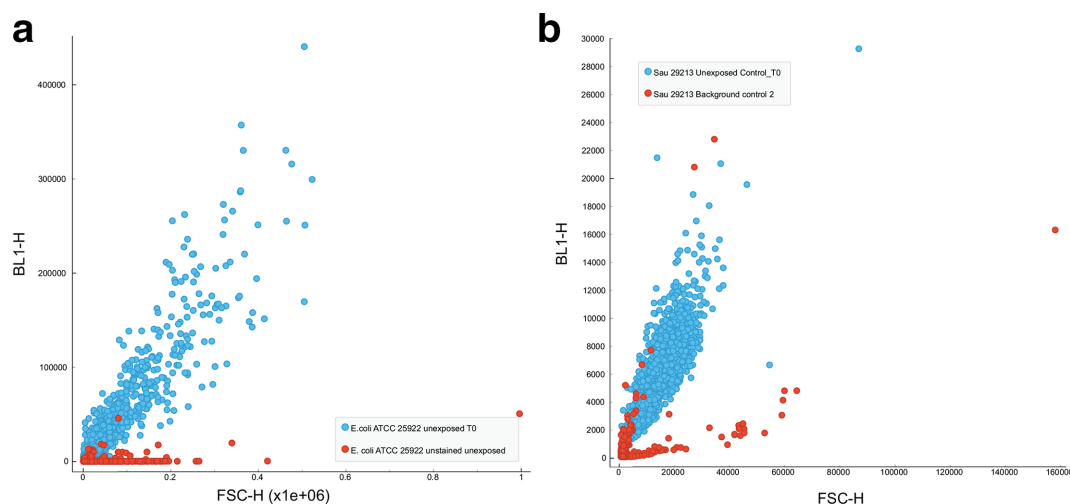### Data machine 1: antimicrobial unexposed population determiner

Bacteria. *E.coli* ATCC 25922; *K. pneumoniae* ATCC 700603, *S. aureus* ATCC 25923, *S. aureus* 29 312.

The FAST analysis method we previously reported relies on definition of an antimicrobial agent-unexposed bacterial population, called the unexposed cellular morphotype (UCM) [4]. These measurements were analysed in a bivariate plot of forward scatter, and the 488 nm fluorescence, which captured the signal from the optical events corresponding to bacterial suspensions stained by adding the intercalating nucleic acid dye, SYTO9.

In order to determine the most suitable channels for analysis, we used a supervised machine-learning approach that combined a series of classification and visualization tools [9, 10] to understand the flow-cytometer channels that best defined an antimicrobial agent-unexposed population (AUP). We used data from replicate blank samples (suspension medium containing SYTO 9) and antimicrobial-unexposed population wells, applied equally weighted data selection to both these sources, and concatenated these data sets for classification and visualization. We then used information rank, principal component analysis and a tree classifier set to a depth of three layers to determine the most informative data projections. We then used a bivariate scatter map, polynomial classification and a scatter plot to determine the best boundaries for the AUP, and displayed this with a frequency-density histogram. The population of interest was selected by manually drawing a rectangular box on the scatter plot to set the boundaries of the AUP zone and exclude background particulate and electronic noise to distinguish the AUP from the contents of the blank well. This blue-coloured event population was used consistently throughout all three data machines (Figs 2–4) and the ensemble to aid visualization.

FORMATTING — Conversion of native flow cytometer file format to comma separated variable

ASSEMBLY — Orderly assembly of cleaned data files into control and concentration series

SELECTION — Concatenation of assembled data, onward feed and analytical parameter selection

CLASSIFICATION — Analysis of data complexity, and hierarchy

VISUALISATION — Display of data showing antimicrobial concentration dependent effect

Fig. 1. Analysis of antimicrobial susceptibility flow cytometer data by supervised machine learning relies on a standardised data handling workflow, comprising conversion of flow cytometer data into .CSV format, assembly of an orderly collection of cleaned data files, which are then linked (concatenated), analytical parameters selected, classified into a hierarchy to optimse information gain, and then displayed to present antimicrobial concentration-dependent effects that can be calibrated against a Minimum Inhibitory Concentration.

Fig. 2. Antimicrobial-unexposed population. Antimicrobial unexposed bacteria (blue) and background particulate noise (red), showing (left) *E. coli* and (right) *S. aureus*. Differences in unexposed bacterial density are due to balancing the numbers of unstained unexposed (blue) and background noise (control, red) events to avoid classification bias.

Subsequent colour generation in concentration series varied due to automated line colour generation.

Data machine 2: ordinal antimicrobial susceptibility classifier

After defining the AUP with data machine 1 (Fig. 2), we concatenated the unexposed bacterial suspension data with the lowest and uppermost antimicrobial-exposed bacterial suspension files. The optimal data-capture parameters for the ordinal susceptibility classification were selected as before [principally by tree classification, after looking at data rank and the dimensionality in principal component analysis]. The channels used for the scatter plot were those used to determine the AUP in data machine 1, so that the same zone could be selected, and carried across to the frequency distribution histogram. These same parameters were used to set the axes of the scatter map, and polynomial classification as a check on the consistency of the AUP. The distribution histogram was used to visualize any a concentration-dependent antimicrobial effect within the tested antimicrobial concentration range, but displayed with the optimal combination of parameters from the tree classification above. The second data machine therefore ordered antimicrobial susceptibility to indicate whether the MIC lay below, within or above the range of tested antimicrobial concentrations but did not define a precise value. To enable delineation of the AUP in the scatter plot, the zoom function and low jitter were used to clarify the position of the AUP, consistently displayed as the light blue population.

Data machine 3: antimicrobial susceptibility classifier for predicted inhibitory concentration

After classifying antimicrobial susceptibility into one of three broad categories in data machine 2, we built a third data machine that included equally weighted data samples at all concentrations in the antimicrobial series also used for broth

microdilution MICs (Fig. 5). Having established that an antimicrobial concentration-dependent effect was in the tested concentration range, the same data-processing sequence was used to perform a quantitative version of data machine 2 with the same analytical processes to demonstrate and measure concentration-dependent effects on data from the optimized channel combination determined in data machines 1 and 2. Principal component analysis showed that a minimum of five data parameters were needed to capture at least 90 % variation in these complex bacterial data sets (Fig. 3at). The tree classification function (Fig. 3b) often identified supplementary flow-cytometer channels for determination of a predicted inhibitory concentration (PIC; approximation to MIC) in addition to the channels used in the original FAST method report [4]. These additional channels were used according to the priority indicated by the tree classifier to analyse the data obtained from the entire antimicrobial concentration series and compare it against the AUP. A PIC was determined from the antimicrobial–bacterial combination series when a >50 % reduction in frequency density from the AUP density curve was observed (Fig. 3f). An important exception to the analytical process were aminoglycoside class agents where this density reduction was only evident when a proportionate data sample (e.g. 10%, rather than a fixed event number) was used for data input at each antimicrobial concentration due to the rapid action of aminoglycosides seen in sensitive bacteria. The PIC was then interpreted against the 2019 EUCAST susceptibility test standards.

Step two

Expanded challenge panel

Bacteria. Bacteria: *E. coli* ATCC 25922, *E. coli* ATCC 35218, *E. coli* −2841 (clinical), *K. pneumoniae* ATCC 700603, *K. pneumoniae* ATCC 700603, *K. pneumoniae* ATCC BAA-1705, *K. pneumoniae* ATCC BAA-1706, *K. pneumoniae* ATCC 13883,
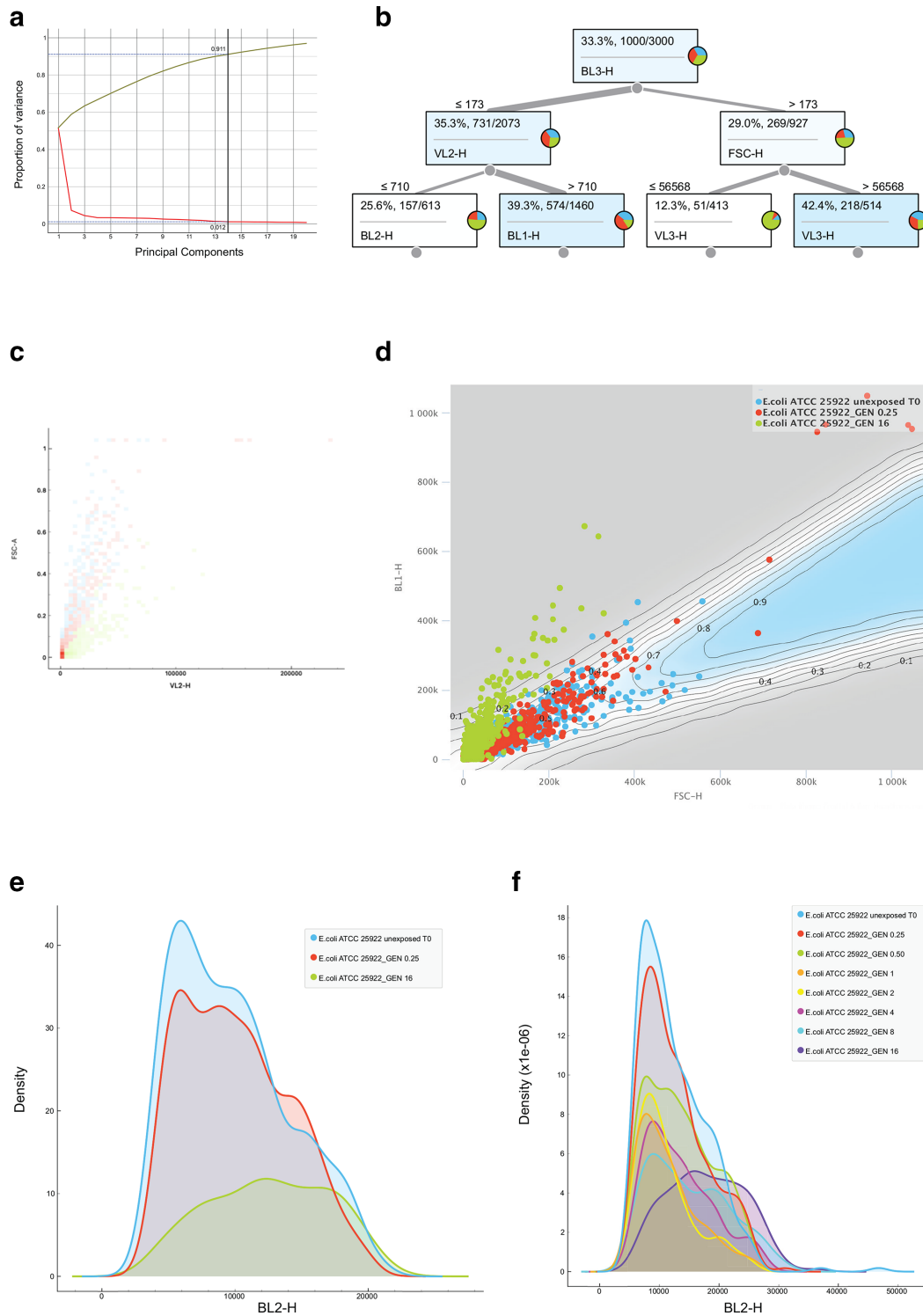
Fig. 3. Visualising antimicrobial susceptibility: Gentamicin-exposed *E. coli.* Top panel: Data Machine 2 (a) Principal Component Analysis, Scree diagram showing dimensionality of data, (b) tree classification with reference to the AUP (Blue). right) tree classification with reference to the AUP (Blue). Middle panel: (c) scatter map and (d) polynomial classification both with toggle on/off for specific concentrations to enable detection of concentration-dependent effect. Bottom panel, falling bacterial population density in AUP zone. (e) Data Machine 2. Density histogram of antimicrobial-unexposed population (AUP, blue), and lowest (red) and highest (green) Gentamicin-concentration-exposed *E. coli* ATCC 25922. The loss of events between low and high Gentamicin concentration indicates a likely concentration-dependent effect within the tested range of concentrations. (f) Data Machine 3. The corresponding frequency distribution histogram featuring all tested concentrations and shows progressive loss of event density in the AUP zone. Predicted inhibitory concentration (PIC) = 2 µg ml[−1].
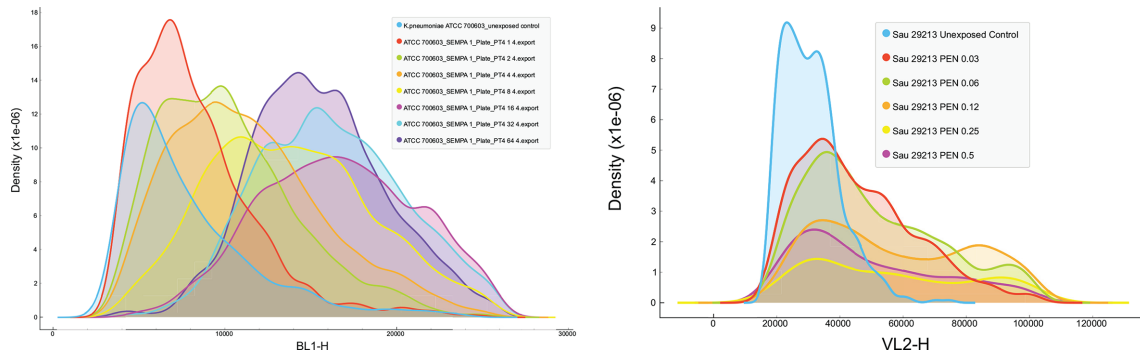
Fig. 4. Antimicrobial resistant profiles. Left, *K.pneumoniae* ATCC 700603 exposed to a series of Piperacillin/Tazobactam concentrations with a high predicted effective concentration (PIC = ≥64 µg ml⁻¹), indicating resistance. Right, *S.aureus* ATCC 29213 and penicillin concentration series, with an MIC at the breakpoint (breakpoint = 0.12–0.25; PIC > 0.12 µg ml⁻¹). NB the BMD result of0.5 µg ml⁻¹ was inaccurate.
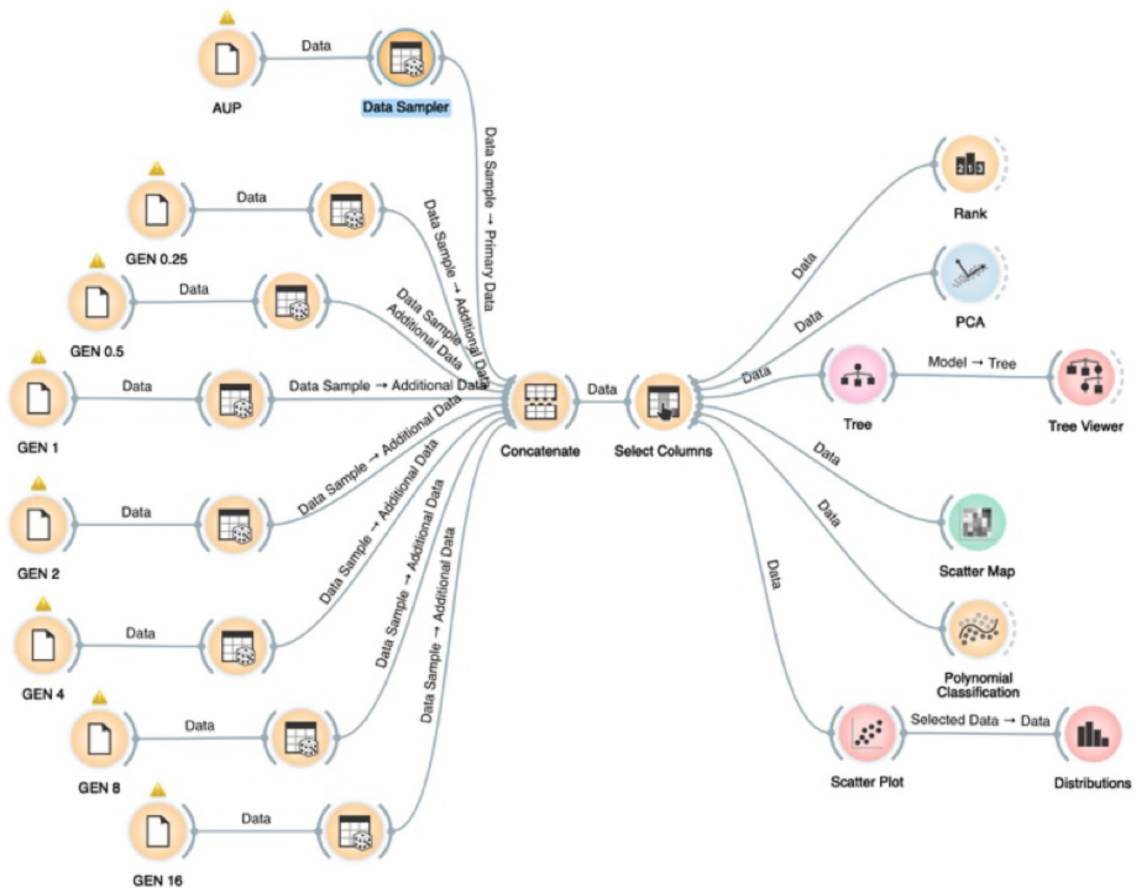
Fig. 5. Data machine 3. Data mining workflow used to assemble flow cytometer data files from *K. pneumoniae* ATCC 700603 to determine Gentamicin PIC, showing data structure, sampling, selection, classification and visualisation (Orange v3.20.1).

P. mirabilis −9545 (clinical), *P. aeruginosa* ATCC 27853, *S. aureus* ATCC 25923, *S. aureus* ATCC 29213, *S. aureus* ATCC 33592, *S. aureus* −6885 (clinical), *E. faecalis* ATCC 29212.

To explore the breadth of application of this data-analysis method, we used data from an expanded panel of nine Gram-negative, and five Gram-positive bacterial species including the four reference strains previously tested. AST was performed in parallel by FAST and Sensititre broth microdilution (BMD) methods against three different classes of antimicrobial agent. Machine-learning analysis was conducted before sighting BMD results in a single pass using the two flow-cytometer channels most commonly used in the standardized FAST method identified with data machine two (above), i.e. FSC-H and BL1-H. No attempt was made to correct apparent discrepancies with BMD results by iterative machine-learning techniques such as unsupervised machine learning.

Step three

Clinical application

The final version of these data machines was then used to process FAST data from new blood culture isolates available after primary overnight sub-culture from the previous day's autoanalyser positive blood cultures. The first two bacterial isolates from different blood culture sets at the start of the morning shift were obtained on blood agar plates to give us representative Gram-negative and Gram-positive bacteria. These were identified to species level by MALDI-TOF mass spectrometry (score >2.0; MALDI Biotyper Reference Laboratory, Bruker Daltonics, Bremen, Germany) respectively as *E. coli* and *S. aureus* (designated 1A and 9B) and processed by FAST method consecutively. FAST results were generated in real time and the BMD results were read the following day, ensuring objectivity of analysis by the machine-learning algorithm. As the SEMPA1 and SEMSE3 96-well plates contain a much larger range of antimicrobials than regularly used in our centre, three antimicrobial agents with the highest immediate clinical utility were selected for each isolate: *E. coli* – gentamicin, piperacillin/tazobactam, meropenem, *S. aureus* – penicillin, cefoxitin, vancomycin.

RESULTS

Flow-cytometer data were analysed from nine different Gram-negative and five different Gram-positive reference strain bacteria, and two clinical isolates; one Gram-negative and one Gram-positive. Initial definition of the AUP was achieved by analysing each unexposed bacterial population using data machine 1 (Fig. 2). This process was subsequently incorporated into data machines 2 and 3. From the example AUP data sets used here, the most informative cytometer channels for beta-lactam antibiotic combinations were usually channels BL1-H and FSC-H, shown on the *x*- and *y*-axes of the AUP scatter plots (Fig. 2).

The density distribution of events inside the border of the high-lighted AUP zone for the lowest and highest tested antibiotic

concentration were plotted against the most informative channel obtained from a simple tree classifier in data machine 2 (Fig. 3e; *E. coli* ATCC 25922 v gentamicin, lowest concentration 0.25 µg ml⁻¹, highest concentration 16 µg ml⁻¹). This data set shows a concentration-dependent fall in bacterial cell density within the tested antimicrobial concentration range. Though the MIC cannot be determined from this visualization method, the predicted inhibitory concentration appears to lie between the minimum and maximum tested concentration and is therefore in range, justifying additional analysis by data machine 3 to obtain a predicted inhibitory concentration or PIC (Fig. 3f; *E. coli* ATCC 25922 v gentamicin), which in this case was 2 µg ml⁻¹ gentamicin. Using a 50 % fall in peak event density on the most highly ranked channels, the approximate categoric agreement between machine-learning analysis of all SEMPA1 and SEMSE3 MICs was 91.07 % when compared with proprietary flow-cytometry software (FlowJo) and 89.29 % when compared with BMD (Table 1). The approximate essential agreement between machine learning and proprietary software was 96.43 % and with BMD was 100 %. Application of the finalized machine-learning algorithm to the composite Gram-negative and Gram-positive challenge panel, in which each isolate was tested against three different classes of antimicrobial agent, produced an overall categoric AST agreement with BMD MIC results in 35/42 (83 %) and essential agreement in 28/42 (67 %) (Table 2). Gram-negative and Gram-positive categoric and essential agreement was 23/27 (85 %) CA and 16/27 (59 %) EA and 12/15 (80 %) CA and 12/15 (80 %) EA, respectively.

Application of the ensemble machine-learning pipeline to FAST data generated PIC results on the same working day (Table 3). The isolates were received at 08:30 h and both analyses were complete by 14:59 h. Total handling time was approximately 10 m for the Gram-negative isolate, and a further 15 m for the Gram-positive. PIC values matched the corresponding MIC for five of six antimicrobials. The one discordant result (*E. coli*, piperacillin/tazobactam) was resolved by correcting for stain uptake variation during the 96-well plate analysis by using the AUP control closest to the antimicrobial series, without AUP controls from other parts of the plate. The lower stain uptake was evident to an experienced user during plate processing, before CSV files were ready for machine learning. Definitive BMD results were available after completion of pipeline analysis, 24 h after inoculation. Manual data transfer, analysis and recording took an average of around 90 min per Sensititre plate.

DISCUSSION

Though the rapid generation of AST data by FAST is valuable in time-critical bacterial infections such as septicaemia and bacterial pneumonia, it needs to be complemented by the speed and consistency offered by an automated analytical pipeline. Once validated with a sufficiently broad range of bacterial species and antimicrobial susceptibilities, the prototype supervised machine-learning pipeline we describe here could be automated for unsupervised machine learning, so

Table 1. Data-machine development and calibration series, antimicrobial susceptibility test results

Species, strain	Antimicrobial agent	BMD[a]	FAST		BP[d]	S-R[e]	corrn[f]
			ps[b]	sml[c]			
E. coli	Amikacin	2	1	2	8	S	
ATCC 25922	Aztreonam	1	1	1	1	S	
	Ciprofloxacin	≤0.12	≤0.12	≤0.12	0.25	S	
	Colistin	1	≤0.25	≤0.25	2	S	
	Cefepime	≤1	≤1	≤1	1	S	
	Gentamicin	2	0.5	0.5	2	S	2/S
	Imipenem	≤0.25	≤0.25	≤0.25	2	S	
	Levofloxacin	≤0.12	≤0.12	≤0.12	0.5	S	
	Meropenem	≤0.12	≤0.12	≤0.25	2	S	
	Piperacillin/tazobactam	4	4	2	8	S	
	Co-trimoxazole	≤0.12	≤0.12	0.25	2	S	
	Ceftazidime	≤0.5	≤0.5	≤0.5	1	S	
	Tobramycin	2	0.5	≤0.25	2	S	1/S
K. pneumoniae	Amikacin	1	1	1	8	S	
ATCC 700603	Aztreonam	>64	32	>64	1	R	
	Ciprofloxacin	1	0.5	0.5	0.25	R	
	Colistin	1	≤0.25	≤0.25	2	S	
	Cefepime	8	≤1	32	1	R	
	Gentamicin	8	4	4	2	R	
	Imipenem	0.5	0.5	2	2	S	
	Levofloxacin	2	0.5	2	0.5	S	
	Meropenem	≤0.12	≤0.12	≤0.12	2	S	
	Piperacillin/tazobactam	32	8	>64	8	R	
	Co-trimoxazole	2	2	4	2	I	
	Ceftazidime	>32	32	>32	1	R	
	Tobramycin	8	2	4	2	R	
S. aureus	Amikacin	2	2	2	16	S	
ATCC 25923	Azithromycin	1	≤0.5	1	2	S	
	Ciprofloxacin	0.5	0.25	0.25	1	S	
	Clarithromycin	0.25	0.25	0.25	2	S	
	Clindamycin	≤0.12	≤0.12	0.25	0.5	S	
	Cefoxitin	4	2	2	4	S	
	Gentamicin	0.5	0.25	0.5	1	S	
	Levofloxacin	0.25	≤0.12	≤0.12	1	S	
	Linezolid	1	1	1	4	S	
	Moxifloxacin	≤0.06	≤0.06	≤0.06	0.25	S	
	Norfloxacin	≤1	≤1	≤1	NA		

Table 1. Continued

Species, strain	Antimicrobial agent	BMDᵃ	FAST		BPᵈ	S-Rᵉ	corrnᶠ
			psᵇ	smlᶜ			
	Ofloxacin	0.5	0.25	0.25	1	S	
	Penicillin	≤0.03	≤0.03	≤0.03	0.125	S	
	Teicoplanin	0.5	≤0.25	0.5	2	S	
	Tobramycin	0.25	0.5	0.25	1	S	
	Vancomycin	2	1	1	2	S	
S. aureus	Amikacin	4	2	4	16	S	
ATCC 29213	Azithromycin	2	1	1	2	S	
	Ciprofloxacin	0.5	≤0.12	≤0.12	1	S	
	Clarithromycin	0.5	≤0.25	≤0.25	2	S	
	Clindamycin	≤0.12	≤0.12	≤0.12	0.5	S	
	Cefoxitin	4	4	4	4	S	
	Gentamicin	1	0.25	0.25	1		
	Levofloxacin	0.25	≤0.12	0.25	1	S	
	Linezolid	4	2	2	4	R	
	Moxifloxacin	≤0.06	≤0.06	≤0.06	0.25	S	
	Norfloxacin	2	≤1	≤1	NA		
	Ofloxacin	≤0.25	≤0.25	≤0.25	1	S	
	Penicillin	>0.5	0.25	0.5	0.12	R	
	Teicoplanin	1	≤0.25	0.5	2	S	
	Tobramycin	1	0.5	0.25	1	S	
	Vancomycin	≤1	≤1	≤1	2	S	

a, BMD, broth microdilution.
b, ps, proprietary software.
c, sml, supervised machine learning.
d, BP, EUCAST susceptible breakpoint (μg ml^{-1}).
e, S-R, sensitive/resistant categorization.
f, corrn, corrected by re-training pipeline.

that every successive analysis is incorporated into an iterative learning process (glossary of machine-learning terms, Table 4). In this proof-of-concept study, errors obtained when we applied FAST to critical clinical isolates could be rectified in the future by additional recursive analysis. Supervised machine learning allows recognition of a signal outlier detected during prototype development as here, but might in future be handled by an automated pipeline. The clinical potential of the FAST method depends in part on its ability to deliver valid AST results on the same day that bacterial growth is first detected in blood and other critical cultures. The prototype machine-learning algorithm we describe here shows how data mining could be used to achieve this outcome and place AST results closer in time to rapid bacterial identification by MALDI-TOF in the clinical laboratory workflow.

The classification of bacteria into antimicrobial sensitive or antimicrobial resistant is one of the most clinically useful determinations made by the hospital microbiology laboratory, since it informs treatment choice, infection control interventions and antimicrobial resistance surveillance. Antimicrobial susceptibility is also the entry point to screening new candidate antimicrobial drugs. Here we report our use of machine-learning methods to classify and visualize antimicrobial susceptibility using multi-parameter flow-cytometer analysis of bacterial populations to determine AST with improved speed and accuracy. Currently used AST methods are approximate indicators of therapeutic efficacy [11]. These methods lack precision, but persist in their current form because they are a widely accessible solution to the high demands

Table 2. Expanded bacterial challenge set, single-pass antimicrobial susceptibility test results

Species	Strain	Antimicrobial	PIC(cat.)[a]	MIC(cat.)[b]	CA[c]	EA[d]
K. pneumoniae	ATCC 1705	Meropenem	4 (I)	8 (I)	Y	Y
K. pneumoniae	ATCC 1706	Meropenem	≤0.12 (S)	2 (S)	Y	N
K. pneumoniae	ATCC 13883	Meropenem	≤0.12	≤0.12	Y	Y
K. pneumoniae	ATCC 700603	Meropenem	≤0.12 (S)	≤1 (S)	Y	N
E. coli	ATCC 25922	Meropenem	≤0.12 (S)	≤0.12 (S)	Y	Y
E. coli	ATCC 35218	Meropenem	0.5 (S)	0.12 (S)	Y	N
E. coli	−2841	Meropenem	≤0.12 (S)	≤0.12 (S)	Y	Y
P. mirabilis	−9545	Meropenem	≤0.12 (S)	≤0.12 (S)	Y	Y
P. aeruginosa	ATCC 27853	Meropenem	≤0.12 (S)	0.5 (S)	Y	N
K. pneumoniae	ATCC 1705	Ceftazidime	4 (I)	8 (I)	Y	Y
K. pneumoniae	ATCC 1706	Ceftazidime	≤0.5 (S)	32 (R)	N	N
K. pneumoniae	ATCC 13883	Ceftazidime	≤0.5 (S)	≤0.5 (S)	Y	Y
K. pneumoniae	ATCC 700603	Ceftazidime	8 (R)	32 (R)	Y	N
E. coli	ATCC 25922	Ceftazidime	≤0.5 (S)	≤0.5 (S)	Y	Y
E. coli	ATCC 35218	Ceftazidime	16 (R)	≤0.5 (S)	N	N
E. coli	−2841	Ceftazidime	0.5 (S)	0.5 (S)	Y	Y
P. mirabilis	−9545	Ceftazidime	≤0.5 (S)	≤0.5 (S)	Y	Y
P. aeruginosa	ATCC 27853	Ceftazidime	1 (S)	4 (I)	N	N
K. pneumoniae	ATCC 1705	Gentamicin	0.5 (S)	≤2 (S)	Y	N
K. pneumoniae	ATCC 1706	Gentamicin	0.25 (S)	1 (S)	Y	N
K. pneumoniae	ATCC 13883	Gentamicin	1 (S)	0.5 (S)	Y	Y
K. pneumoniae	ATCC 700603	Gentamicin	4 (I)	8 (I)	Y	Y
E. coli	ATCC 25922	Gentamicin	1 (S)	2 (S)	Y	Y
E. coli	ATCC 35218	Gentamicin	16	1	N	N
E. coli	−2841	Gentamicin	0.5 (S)	0.5 (S)	Y	Y
P. mirabilis	−9545	Gentamicin	2 (S)	4 (S)	Y	Y
P. aeruginosa	ATCC 27853	Gentamicin	≤0.25 (S)	0.5 (S)	Y	Y
Gram negative					23/27 (85%)	16/27 (59%)
S. aureus	ATCC 25923	Penicillin	≤0.03(S)	≤0.03(S)	Y	Y
S. aureus	ATCC 29213	Penicillin	0.12 (S)	>0.5(R)	N	N
S. aureus	ATCC 33592	Penicillin	≥0.5 (R)	>0.5	Y	Y
S. aureus	−6885	Penicillin	≥0.5 (R)	>0.5	Y	Y
E. faecalis	ATCC 29212	Penicillin	≥0.5 (R)	>0.5	Y	Y
S. aureus	ATCC 25923	Cefoxitin	≤1	1	Y	Y
S. aureus	ATCC 29213	Cefoxitin	2	4	Y	Y
S. aureus	ATCC 33592	Cefoxitin	2	16(R)	N	N
S. aureus	−6885	Cefoxitin	16 (R)	>16	Y	Y
E. faecalis	ATCC 29212	Cefoxitin	≥16 (R)	>16	Y	Y

Continued

Table 2. Continued

Species	Strain	Antimicrobial	PIC(cat.)[a]	MIC(cat.)[b]	CA[c]	EA[d]
S. aureus	ATCC 25923	Vancomycin	≤0.5(S)	2(S)	Y	Y
S. aureus	ATCC 29213	Vancomycin	1 (S)	1(S)	Y	Y
S. aureus	ATCC 33592	Vancomycin	1 (S)	2(S)	Y	Y
S. aureus	−6885	Vancomycin	1 (S)	4(R)	N	N
E. faecalis	ATCC 29212	Vancomycin	2(S)	4(S)	Y	Y
Gram positive					12/15 (80%)	12/15 (80%)
Total					35/42 (83%)	28/42 (67%)

a, PIC (cat.); predicted inhibitory concentration (categoric result [S,I,R]).

b, MIC (cat.); broth MIC (cat.); broth microdilution MIC (categoric result [S,I,R]).

c, CA; categoric agreement [yes/no].

d, EA; essential agreement [yes/no].

on clinical laboratories despite growing pressure for earlier selection of effective antimicrobial therapy to reduce mortality from severe bacterial infections [12, 13]. Though the FAST method addresses this need by combining accuracy and speed, the high-end analytic flow-cytometry skills it relies on may not be available in many clinical microbiology laboratories. The available commercial flow data-analysis pipelines lack the capacity to analyse the numerous bacterial datasets generated by flow cytometers to keep up with clinical demands. The proprietary software we use is able to generate PIC results in a shorter time frame when operated by an expert user, but lacks the capability of machine learning unless coded for immediate processing of native flow cytometer files, or the adaptive potential of an automated machine-learning algorithm. Noting these constraints, use of the FAST method for high AST throughput in a busy clinical laboratory generates a heavy bioinformatic

processing burden, and when performed manually might cause problems with laboratory accreditation standards. The conversion of our manual, tube-based FAST method [4] to a high-throughput, semi-automated 96-well plate format dramatically increased the volume of data for analysis, and prompted us to explore alternative approaches to data handling and analysis. In the present study, we demonstrate that supervised machine learning provides a data-processing pipeline that, once calibrated against current reference susceptibility test methods, is capable of the classification and visualization necessary for accurate quantitative AST result prediction. In view of the critical role AST plays in selection of antimicrobial therapy, our machine-learning ensemble presents a prototyping method to meet a current bioinformatics shortfall in clinical microbiology. Our reliance on calibration against the biological endpoints in reference AST methods highlights the need for agreed bioinformatic

Table 3. Clinical isolates, single-pass antimicrobial susceptibility test results

Species, isolate	Antimicrobial agent	BMD[a]	FAST ps[b]	FAST sml[c]	BP[d]	S-R[e]	corrn[f]
E. coli	Piperacillin/tazobactam	4	2	>64	8	R	4/S
1A	Gentamicin	0.5	0.5	0.5	2	S	
	Meropenem	≤0.12	≤0.12	≤0.12	2	S	
S. aureus	Penicillin	>0.5	>0.5	>0.5	0.12	R	
5B	Cefoxitin	>16	16	16	4	R	
	Vancomycin	1	0.5	1	2	S	

a, BMD, broth microdilution.

b, ps, proprietary software.

c, sml, supervised machine learning.

d, BP, EUCAST susceptible breakpoint (µg ml^{-1}).

e, S-R, sensitive/resistant categorization.

f, corrn, corrected by re-training pipeline.

Table 4. Machine-learning terms used

Phrase	Explanation
Machine learning	The use of computers to automate learning from input data
Supervised (SML)	Use of computers to analyse large data sets, starting with a set of labelled examples (training set) to predict specific outcomes
Unsupervised (UML)	Fully automated machine learning that does not require the expertise of a user to supervise training
Data mining	The examination of large data sets to generate new information
Classification	The prediction of class labels or categories using a series of learning steps
Tree classification	A commonly used form of predictive modelling in which recursive binary partitioning is applied to categoric variables.
Visualization	The presentation of data in picture form to assist decision-making
Concatenation	Connection of a series of discrete sets of similarly formatted data to preserve their source as a feature
Jitter	Displacement of data points to assist their individual visibility by reducing two-dimensional overlap
Principal component analysis (PCA)	The transformation of potential correlates into sets of non-correlating data known as principal components in which the first component has greatest variance.

standards for bacterial flow-cytometry data-analysis procedures beyond the standard output file format [10].

FAST assays are based on detection of changes in dye-staining properties of one or more bacterial sub-populations on exposure to an antibacterial agent before, and independent of, cell death, rupture and dispersal. The changes in bacteria we can measure with a flow cytometer cover a range of parameters including size, shape, internal complexity and colour within the detectable spectrum [14]. The use of supervised machine learning on these data sets provides a new approach to gaining insight into antimicrobial resistance mechanisms. For example, our approach could be applied to investigate hetero-resistance in multidrug-resistant strains of bacteria [15, 16]. The supervised machine-learning pipeline demonstrated here could be also used to assess the utility of alternative dyes for rapid AST, or the efficacy of new candidate antimicrobial agents. We emphasize that the present study is a proof of concept for application of machine-learning techniques to rapid AST. Application of these techniques to a bacterial challenge panel exposed AST discrepancies when compared with the broth microdilution method. These cannot be fully explained by the low accuracy of some BMD endpoints. There are machine-learning techniques that could be used to further improve the accuracy of FAST data interpretation, such as additional recursive

analysis by automated (unsupervised) machine learning on data sets generated from much larger bacterial strain collections. Other discrepant results can be analysed in additional flow-cytometer channels in a supervised machine-learning calibration process.

One of the notable outcomes of our data-machine ensemble is the single-cell population-based evidence for antimicrobial susceptibility and resistance. This supervised machine-learning ensemble and its major data machine components present a method of visualizing antimicrobial susceptibility as a series of antimicrobial concentration-dependent effects in one or more measurable characteristics of the antimicrobial-exposed bacterial population. Antimicrobial resistance corresponds to a loss of these concentration-dependent changes in flow-cytometer-based bacterial population analysis at single-cell resolution. Exactly what form these changes take and their optimal measurement is likely to depend on the mechanisms and dynamics of antimicrobial action, and thus, the combination of antimicrobial agent and the target microorganism. For example, beta lactam antibiotics inhibit cell-wall formation, interfere with bacterial cell division and rely on time of exposure above the MIC [17]. Aminoglycosides, by contrast, rely on inhibition of protein synthesis and depend on the peak antimicrobial concentration and other indicators [18]. These differences in the biology of antimicrobial agent action produce different concentration-dependent effects visualized in the scatter plots and corresponding frequency distribution histogram analyses. Additional systems biology tools are needed to explain exactly what these measurable parameters signify at a molecular level.

In conclusion, supervised machine learning enabled us to determine AST classifications without the high-end analytic skills of an expert flow-cytometer user or dedicated flow-cytometry analytic software. Further steps are now needed to incorporate our prototype machine-learning algorithm into flow-cytometer operating software to enable prescriptive unsupervised machine learning during single-cell bacterial population analysis. There is also a need to benchmark these tests and their supporting software to fulfil laboratory quality system and accreditation requirements. Fully automated quantitative bacterial cell analysis has far-ranging implications for the clinical microbiology laboratory.

References

1. **International Standards Organisation**. 2006. ISO 20776-1:2006. Clinical laboratory testing and *in vitro* diagnostic test systems – Susceptibility testing of infectious agents and evaluation of performance of antimicrobial susceptibility test devices – Part 1: Reference method for testing the *in vitro* activity of antimicrobial agents against rapidly growing aerobic bacteria involved in infectious diseases. https://www.iso.org/standard/41630.html

2. **Inglis TJJ, Urosevic N**. Where sepsis and antimicrobial resistance countermeasures converge. *Front Public Health* 2017;5:6.

3. **HM Government**. *Rapid Diagnostics: Stopping Unnecessary Use of Antibiotics*. Westminster, UK: The Review on Antimicrobial Resistance; 2015.

4. **Mulroney KT, Hall JM, Huang X, Turnbull E, Bzdyl NM** *et al.* Rapid susceptibility profiling of carbapenem-resistant *Klebsiella pneumoniae*. *Sci Rep* 2017;7:1903.

5. **Inglis TJJ**. Editorial conference report 5th international one health congress, saskatoon, Canada, 22–25 June 2018. *J Med Microbiol* 2018;67:1665–1666.

6. **Steen HB, Boye E, Skarstad K, Bloom B, Godal T** *et al.* Applications of flow cytometry on bacteria: cell cycle kinetics, drug effects, and quantitation of antibody binding. *Cytometry* 1982;2:249–257.

7. **Boye E, Steen HB, Skarstad K**. Flow cytometry of bacteria: a promising tool in experimental and clinical microbiology. *Microbiology* 1983;129:973–980.

8. **Gant VA, Warnes G, Phillips I, Savidge GF**. The application of flow cytometry to the study of bacterial responses to antibiotics. *J Med Microbiol* 1993;39:147–154.

9. **Demšar J, Curk T, Erjavek A, Gorup C, Hocevar T** *et al.* Orange: data mining toolbox in python. *J Machine Learning Res* 2013;14:2349–2353.

10. **Spidlen J, Moore W, Parks D, Goldberg M, Bray C** *et al.* Data file standard for flow cytometry, version FCS 3.1. *Cytometry A* 2010;77:97–100.

11. **Andrews JM, Howe RA, BSAC Working Party on Susceptibility Testing**. BSAC standardized disc susceptibility testing method (version 10). *J Antimicrob Chemother* 2011;66:2726–2757.

12. **Kumar A, Roberts D, Wood KE, Light B, Parrillo JE** *et al.* Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med* 2006;34:1589–1596.

13. **Pruinelli L, Westra BL, Yadav P, Hoff A, Steinbach M** *et al.* Delay within the 3-hour surviving sepsis campaign guideline on mortality for patients with severe sepsis and septic shock. *Crit Care Med* 2018;46:500–505.

14. **Shapiro HM**. Multiparameter flow cytometry of bacteria: implications for diagnostics and therapeutics. *Cytometry* 2001;43:223–226.

15. **Nodari CS, Ribeiro VB, Barth AL**. Imipenem heteroresistance: high prevalence among Enterobacteriaceae *Klebsiella pneumoniae* carbapenemase producers. *J Med Microbiol* 2015;64:124–126.

16. **Bardet L, Baron S, Leangapichart T, Okdah L, Diene SM** *et al.* Deciphering heteroresistance to colistin in a *Klebsiella pneumoniae* isolate from Marseille, France. *Antimicrob Agents Chemother* 2017;61:pii: e00356-17.

17. **Osthoff M, Siegemund M, Balestra G, Abdul-Aziz MH, Roberts JA**. Prolonged administration of β-lactam antibiotics - a comprehensive review and critical appraisal. *Swiss Med Wkly* 2016;146:w14368.

18. **Pagkalis S, Mantadakis E, Mavros MN, Ammari C, Falagas ME**. Pharmacological considerations for the proper clinical use of aminoglycosides. *Drugs* 2011;71:2277–2294.