Electronic Supplementary Information for

Using Machine Learning to Identify and Quantify Endotoxins from Different Bacterial Species using Liquid Crystal Droplets

Authors

Shengli Jiang[1,2], JungHyun Noh[1,3], Chulsoon Park[1,2], Alexander D. Smith[2],

Nicholas L. Abbott[3*], Victor M. Zavala[2*]

[2]Department of Chemical and Biological Engineering, University of Wisconsin- Madison,
1415 Engineering Dr, Madison, WI 53706, USA.

[3]Smith School of Chemical and Biomolecular Engineering, Cornell University,
113 Ho Plaza, Ithaca, NY 14853, USA.

[1]Equally contributing authors

[*]**Equal corresponding Authors:** victor.zavala@wisc.edu and nla34@cornell.edu

Supplementary Experimental Methods

Detailed description of serial dilutions of endotoxin solutions. Prior to dilution, each stock solution of endotoxin (1 mg/mL) was stirred by vortexing at 3000 rpm for 5 seconds. Our experiments were conducted using two different procedures, which we detail below:

(i) With microcentrifuge tubes of type A:

- First, the stock solution of endotoxin (1 mg/mL) was diluted to 10 μg/mL using PBS buffer (no added SDS) in the type A microcentrifuge tube. Specifically, 9 μL of the stock solution of endotoxin was added to 891 μL of PBS buffer to make a solution with a total volume of 900 μL. The solution was mixed by vortexing at 3000 rpm for 20 seconds.
- The 10 μg/mL solution was then diluted 10-fold to 1 μg/mL using PBS buffer in a microcentrifuge

tube of type A by adding 90 μL of 10 μg/mL endotoxin solution to 810 μL of the PBS buffer. Subsequently, the solution was vortexed at 3000 rpm for 5 seconds. Further serial dilutions were performed to reach a final concentration of 0.1 pg/mL.

- The procedure above was repeated to achieve concentrations of endotoxin ranging from 1 μg/mL to 0.01 pg/mL. After dilution, each solution was vortexed for 5 seconds. 700 μL of each diluted solution was transferred to a polystyrene tube.
- Finally, 35 μL of LC emulsion was added to 700 μL of each endotoxin solution, and the endotoxin-LC emulsions were incubated for 30 minutes, prior to flow cytometry measurement.

(ii) With cleaned type B microcentrifuge tubes:

- The stock solution was diluted to 10 μg/mL using PBS containing 27 μM SDS in the rinsed microcentrifuge tube of Type B. Specifically, 9 μL of the stock solution was added to 891 μL of the PBS buffer, thus a total volume of the solution becomes 900 μL. The diluted solution was mixed by vortexing for 20 seconds.
- The 10 μg/mL solution was diluted 10-fold to 1 μg/mL in a microcentrifuge tube of Type B using 27 μM SDS in PBS. Specifically, 90 μL of 10 μg/mL endotoxin solution was added to 810 μL of the PBS buffer to achieve a total solution volume of 900 μL. Afterward, the solution was mixed by vortexing at 3000 rpm for 5 seconds. Further serial dilutions were performed to reach a final concentration of 0.01 pg/mL. Each solution was vortexed for 50 seconds. 700 μL of each diluted solution was transferred to a polystyrene tube.
- Finally, 35 μL of LC emulsion was added to 700 μL of each endotoxin solution and the endotoxin-LC emulsions were incubated for 30 minutes, prior to flow cytometry measurement.

Preparation of 10 mM PBS buffer. Powdered phosphate-buffered saline was dissolved in 1 L of deionized water to prepare 10 mM phosphate-buffered saline at 137 mM NaCl, 0.27 mM KCl, pH 7.4 at 25℃. For this process, a 1 L volumetric flask was used.

Preparation of 27 μM or 5 μM SDS solution. First, 10 mM SDS solution was prepared by adding 0.568 g of SDS to 20 mL of 10 mM PBS buffer in a 20 mL glass vial. The solution was mixed by vortexer (Fisher Scientific brand, catalog number 0215370) at 3000 rpm for 60 seconds. The 10 mM SDS solution was then diluted 10-fold to reach a concentration of 1 mM SDS in a 20 mL glass vial. After mixing by vortexing, 540 μL of 1 mM SDS solution was added to 19.46 mL of 10 mM PBS buffer. The final solution containing 27 μM SDS was mixed by vortexing at 3000 rpm for 30 seconds. A similar procedure was used to prepare 5 μM SDS in PBS.

Supplementary Figures

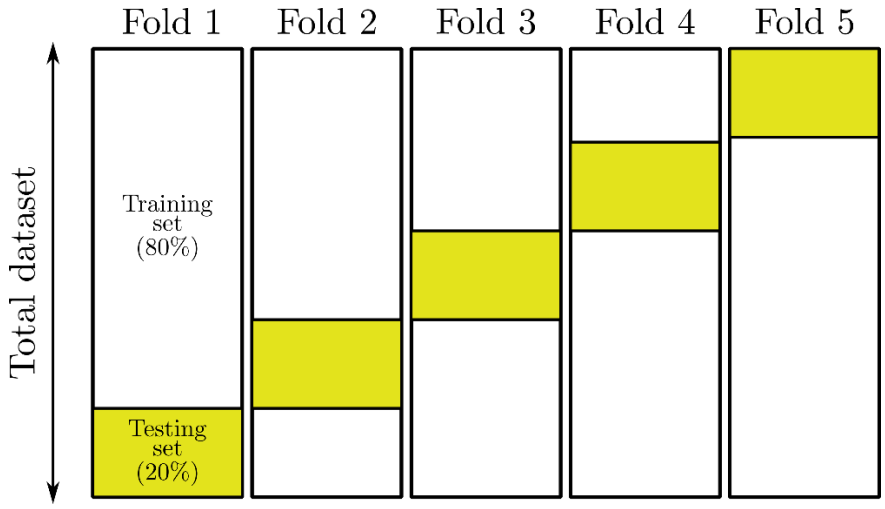

Figure S1. Data training and testing (or prediction) procedure based on a 5-fold approach. In the first fold, 80% of data was randomly selected for training and 20% for prediction. After training the first fold, 80% of the total dataset was selected again randomly as the second fold training and 20% of the data was selected for prediction. After five trainings, all test data can form a complete dataset, thus all the data in the dataset was tested.

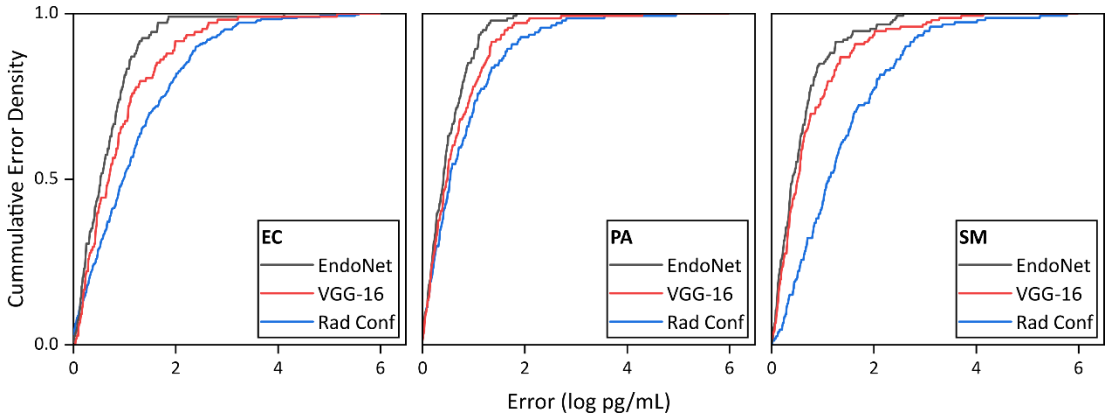

Figure S2. Cumulative frequencies for prediction errors. Cumulative frequency for prediction error using different methods for EC (left), PA (middle), and SM (right). For all bacterial species, the EndoNet curve (black) dominates that of VGG-16 and radial configuration curves (red and blue). This comparison demonstrates that EndoNet consistently has a higher probability of having the smallest prediction error.

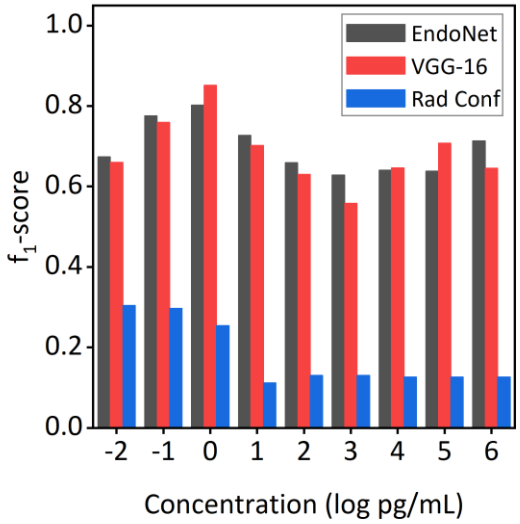

Figure S3. f_1-score for LPS species classification. f_1-score for LPS classification (SM, EC or PA) at various concentrations of endotoxin. EndoNet has an overall higher f_1-score for multi-class classification. When the f_1-score is closer to one, the classification has a better accuracy.

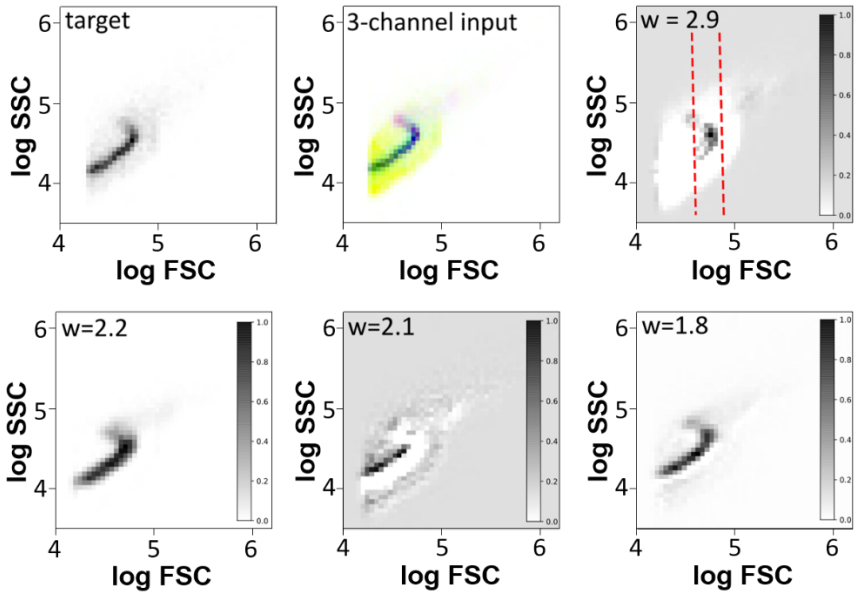

Figure S4. Convolutional feature map analysis of EndoNet. EndoNet feature maps (activated images) that have the highest weights of classification for SM at 1 pg/mL. The target image is in $\mathbb{R}^{50\times50}$, and the input to the CNN (reference images) has three channels ($\mathbb{R}^{50\times50\times3}$). Between the two red dashed lines is the characteristic "S-region". The feature map with $w = 2.9$ is looking at the "S-region" similar to the RC method. The feature maps with $w = 2.2, 2.1$ and 1.8 are focusing on the diagonal patterns of the original input to the CNN. Color bars next to the feature maps indicate the normalized pixel intensity (a large value indicates a

high filter activation).