Machine learning identification of *Pseudomonas aeruginosa* strains from colony image data

Jennifer Rattray[1,2], Ryan Walden[3], Pedro Marquez Zacarias[1,2,4], Evgeniya Molotkova[1,2], Gabriel Perron[5,6], Claudia Solis-Lemus[7,8], Daniel Pimentel Alarcon[8,9], Sam P. Brown[1,2,*]

[1] School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA
[2] Center for Microbial Dynamics and Infection, Georgia Institute of Technology, Atlanta, GA 30332, USA
[3] Rosenblatt, LLC
[4] Graduate Program in Quantitative Biosciences (QBioS), Georgia Institute of Technology, Atlanta, 30332 GA, USA
[5] Department of Biology, Bard College, Annandale-On-Hudson, NY 12504, USA
[6] Center for Systems Biology and Genomics, New York University, New York, NY, 10003, USA
[7] Department of Plant Pathology, University of Wisconsin-Madison, Madison, WI 53706, USA
[8] Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI 53706, USA
[9] Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI 53706, USA

*corresponding author, email Sam.brown@biology.gatech.edu

Abstract

When grown on agar surfaces, microbes can produce distinct multicellular spatial structures called colonies, which contain characteristic sizes, shapes, edges, textures, and degrees of opacity and color. For over one hundred years, researchers have used these morphology cues to classify bacteria and guide more targeted treatment of pathogens. Advances in genome sequencing technology have revolutionized our ability to classify bacterial isolates and while genomic methods are in the ascendancy, morphological characterization of bacterial species has made a resurgence due to increased computing capacities and widespread application of machine learning tools. In this paper, we revisit the topic of colony morphotype on the within-species scale and apply concepts from image processing, computer vision, and deep learning to a dataset of 69 environmental and clinical *Pseudomonas aeruginosa* strains. We find that colony morphology and complexity under common laboratory conditions is a robust, repeatable phenotype on the level of individual strains, and therefore forms a potential basis for strain classification. We then use a deep

convolutional neural network approach with a combination of data augmentation and transfer learning to overcome the typical data starvation problem in biological applications of deep learning and achieve a validation accuracy averaging above 94% for the classification of individual strains. These results indicate that bacterial strains have characteristic visual fingerprints that can serve as the basis of classification on a sub-species level. Our work illustrates the potential of image-based classification of bacterial pathogens and highlights the potential to use similar approaches to predict medically relevant strain characteristics like antibiotic resistance and virulence from colony data.

Author Summary

Since the birth of microbiology, scientists have looked at the patterns of bacterial growth on agar (colony morphology) as a key tool for identifying bacterial species. We return to this traditional approach with modern tools of computer vision and deep learning and show that we can achieve high levels of classification accuracy (94%) on a within-species scale, despite what is considered a 'data-starved' dataset. Our results show that strains of the environmental generalist and opportunistic pathogen *Pseudomonas aeruginosa* have a characteristic morphological 'fingerprint' that enables accurate strain classification via a custom deep convolutional neural network. Our work points to extensions towards predicting phenotypes of interest (e.g. antibiotic resistance, virulence), and suggests that sample size limitations may be less restrictive than previously thought for deep learning applications in biology, given appropriate use of data augmentation and transfer-learning tools.

Introduction

Since Semmelweis, Lister, and Koch began studying pure cultures of microorganisms, scientists have tried to organize the immense diversity of bacteria into orderly categories to understand their behavior, including pathogenesis. In particular, Koch pioneered the method of determining how specific microorganisms cause distinct diseases and highlighted the importance of laboratory culture for the identification and targeted treatment of pathogenic bacteria (1). By identifying the specific bacterial cause of disease, we can better manage and eradicate the infection. When a sample from an environment or clinical source is plated on solid media at sufficient dilution, individual colonies can be observed, where each colony grows from a single founding bacterial cell (Figure 1A). These clonal groups have characteristic sizes, shapes, edges, textures, and degrees of opacity and color (Figure 1B) (2). These characteristics describe the morphology of a single colony and formed the basis for the earliest attempts at categorization. In modern clinical laboratories, morphological tests play a role but are increasingly replaced by biochemical and genomic methods of classification (3).

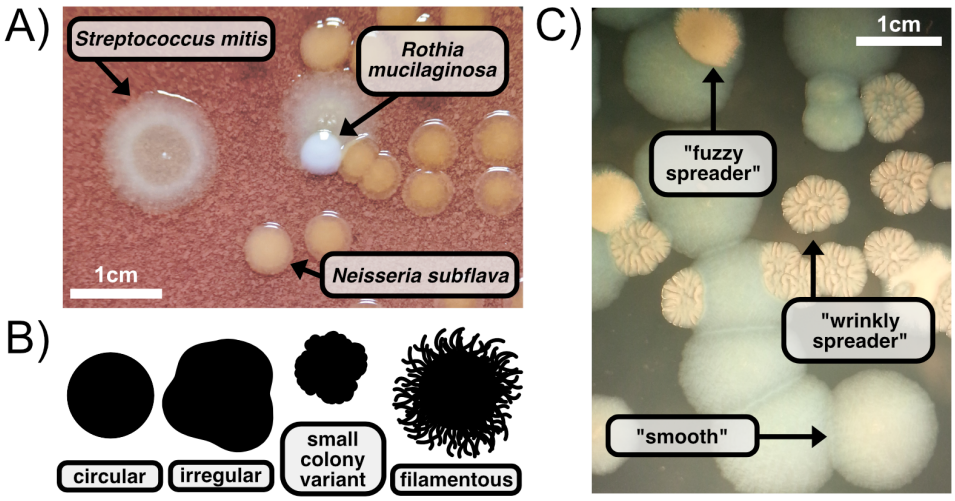


Figure 1. Bacterial colony morphology varies across and within species. A) Morphological identification of bacterial species from a mixed culture plated on Chocolate Agar: *Rothia mucilaginosa* (smallest white circles), *Neisseria subflava* (larger tan round circles), and *Streptococcus mitis* (large bullseye circles). B) A common example of 2D colony morphology features include the appearance of the colony edges. C) Morphological identification of bacterial strains, in this case called "smooth", "wrinkly spreader", and "fuzzy spreader"(4), from a culture containing only *Pseudomonas fluorescens* plated on King's Medium B Agar.

While initial classification was morphology based, advances in genome sequencing technology have revolutionized our ability to classify bacterial isolates. Widespread use of 16S rRNA amplicon sequencing not only allows for robust taxon identification to the genus scale (5,6), but also fosters phylogenetic organization of bacterial taxa (7,8). Marker gene or whole genome sequencing provides finer scale strain identification and allows specific focus on markers of pathogenicity (via identification of virulence factors), drug resistance or other characteristics of interest. Transcriptomic sequencing provides a more direct window into the behavior of pathogens, to assay for example whether virulence factor or drug resistance genes are actually being expressed (9).

While genomic methods are in the ascendancy, morphological characterization of bacterial species has made a resurgence. How a population of bacteria grow is a direct result of the interplay between their genotype, phenotype, and environment (10) and *a priori* contains critical phenotypic information on the behavior of individual microbes. For example, colony morphotype can indicate specific mutations in a bacteria, e.g. small colony variants and cyclic di-GMP regulation (Starkey et al. 2009), or rugosity and phenazine production (12). Morphological characterization and data science have collided to provide insights into image-based prediction of microbial taxa (13–18). The earliest attempts introduced light-scattering for presence-absence detection of microorganisms on surfaces (19,20), which paved the way for image-based identification of *Listeria* species from light scattering patterns of *Listeria* colonies (13). This developed into an automated light scattering sensor called BARDOT (BActerial Rapid Detection using Optical scattering Technology) that produced scatter patterns sensitive enough to distinguish between *Escherichia coli* serotypes (14), virulence mutants (15), and industrial *Staphylococcus* contaminants (16). At the same time, imaged-based prediction diversified to use more accessible standard light microscopes instead of specialized equipment, putting the focus on new analysis methods like colony morphology ontology (18) and deep learning (17,21) .

Deep learning algorithms have gained popularity (17,21,22) as machine learning becomes ubiquitous across fields. Several supervised deep learning models have demonstrated near human accuracy in classification of various image datasets (23) and in the case of bacterial biofilms, deep learning has outperformed human characterization of single and mixed species biofilms (24). Zielinksi et al. investigate several models for colony classification utilizing a deep convolutional

113 neural network (D-CNN) on a rich dataset pre-trained on the ImageNet (25) dataset, achieving a
114 very high validation accuracy of ~ 97%. However, existing work using a D-CNN approach has
115 focused on classification across species, not across strains within a species. Additionally, a
116 ubiquitous challenge in many biology applications (including the one in this paper) is the limited
117 number of samples. For example, the hallmark ImageNet image-based dataset (26–28) contains
118 more than 14 million images across 1000 classes. Outside a colloquial 'rule of thumb' that you
119 need at least 1000 samples per class, there is no absolute way to determine a 'minimum' sample
120 size (29).
121
122 We focus on within species classification of *Pseudomonas aeruginosa*, a gram-negative bacterium
123 capable of causing severe chronic illness and a major cause of acute hospital-acquired infections
124 (30–33). While it is widely viewed as a generalist and can be isolated from a wide number of
125 environments including water and soil, it is frequently found in human/animal-impacted
126 environments (30). *P. aeruginosa* has large variety of mechanisms that allow it to respond to its
127 environment (34–36). There are frequently differences in behavior across strains, like type and
128 production of secreted products (37,38), which can result in observable morphological diversity
129 when cultured on agar surfaces (39–43). Classic examples of morphological variation in P.
130 *aeruginosa* caused by underlying genetics include small colony variants related to cyclic di-GMP
131 regulation (11), rugosity related to phenazine production (12), and mucoidy related to secreted
132 polymer production (44).
133
134 In this paper, we use a collection of 69 clinical and environmental *P. aeruginosa* isolates which
135 present a range of phenotypic features (e.g. exopolysaccharide production, virulence gene

136 expression, and antibiotic resistance profiles) that have been previously described in (45–47).
137 From these 69 isolates we generate a library of 266 *P. aeruginosa* colony images. As expected, we
138 document a high level of variation in the physical appearance of colonies across strains.
139 Additionally, we see that strain morphology in *P. aeruginosa* is consistent across replicates. We
140 then use a D-CNN approach and a combination of data augmentation (48,49) and transfer learning
141 (50) to classify strains of *P. aeruginosa* and achieve an average accuracy of 94%.
142

143 **Results**

144 *Morphological variation across strains and replicates*

145 Individual isolates from our 69-strain collection were spotted onto Luria-Bertani (LB) agar plates
146 supplemented with Congo Red, a dye commonly used in microbiology to stain the biofilm matrix
147 and extracellular polysaccharides (Figure 2). This was done in quadruplicate and generated a
148 library of 266 colonies after quality control (10 colonies were removed due to debris or writing
149 obscuring part of the image). Colonies were grown for 72 hours, which was determined during the
150 pilot experiment to be enough time to reveal major morphological differences while minimizing
151 outward expansion. This resulted in an imaged library of image colonies that ranged in size from
152 12M (approximately 0.74cm x 0.74cm) to 183M pixels (approximately 2.8cm x 2.8cm).
153

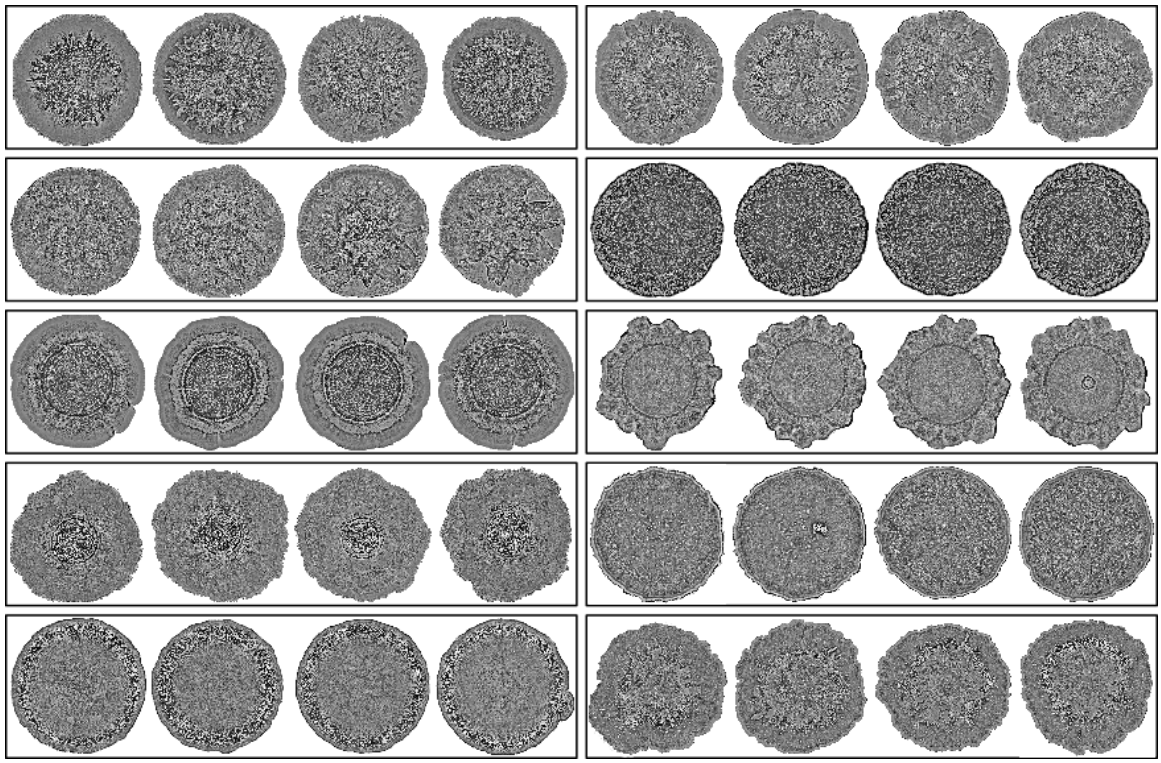


Figure 2. Random sample of the *P. aeruginosa* strain colony dataset. 10 strains (in four-fold replication) were selected randomly without replacement from a complete list of 69 strains via the native sample() command in R. From top to bottom: (left) strain 25, strain, 38, strain 52, strain 113, strain 1; (right) strain 127, strain 316, strain 298, strain 121, strain 174. Colony size has been adjusted for viewing purposes.

In our first approach to characterizing morphological diversity, we use 8 simple descriptive metrics (Figure 3A). Classic morphological descriptive variables include area, radius, perimeter, centroid (the point in which the three medians of the triangle imposed within the image intersect), eccentricity (the ratio of the distance between the foci of the ellipse and its major axis length), circularity (deviation of the perimeter from a superimposed perfect circle), bounding disk center (center of mass of the brightest pixels), and caliper diameter (the caliper dimension obtained by averaging over all orientations).

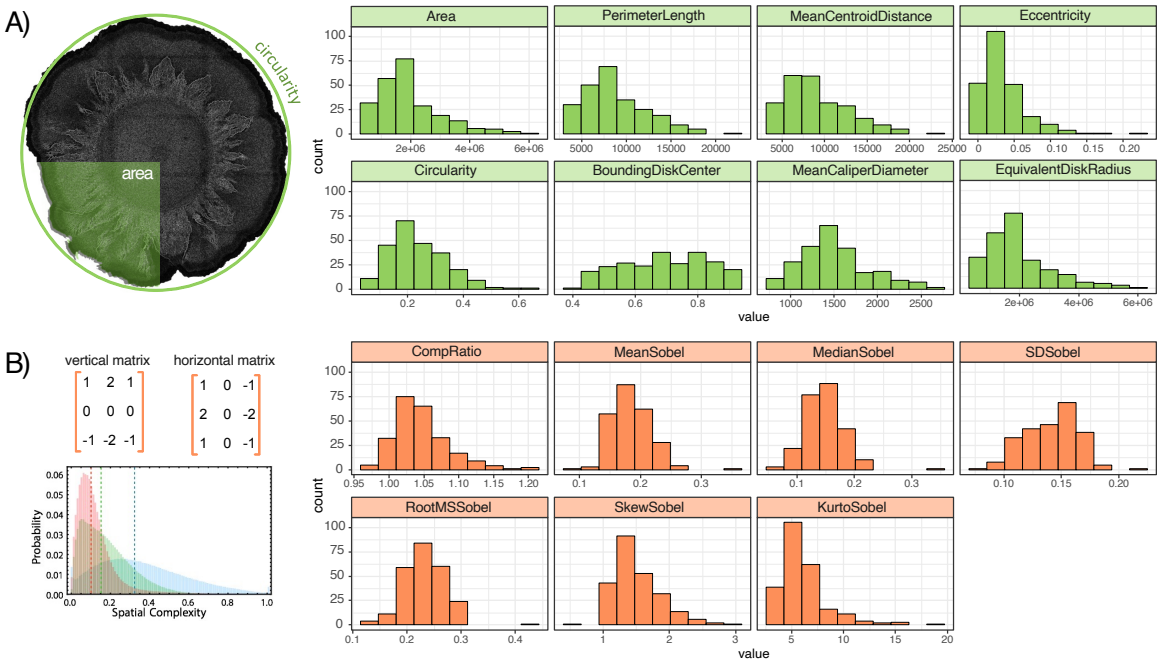
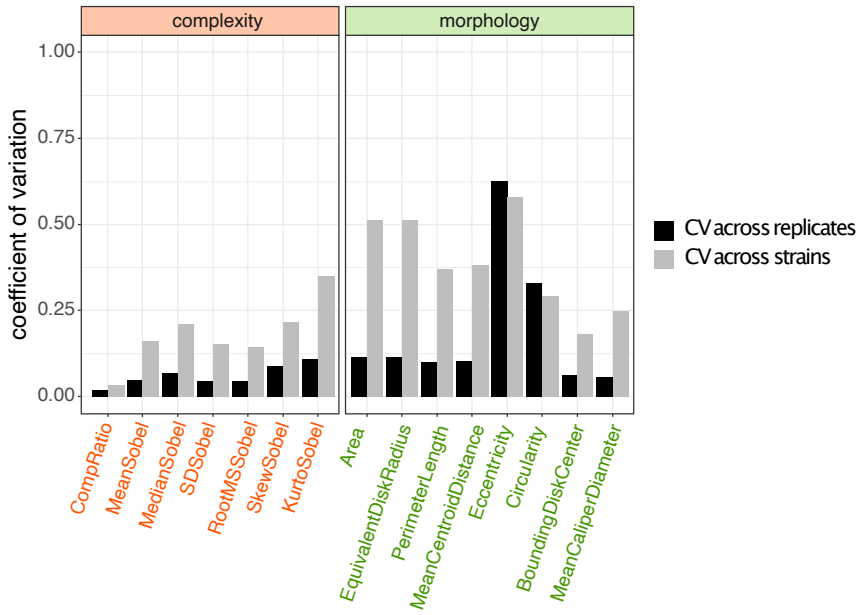


Figure 3. Diversity of *P. aeruginosa* strains in both classic morphological descriptive variables and derived complexity descriptive variables across 69 strains and 266 colonies. Histograms are built from all replicates of all strains. A) classic metrics used to describe colony appearance. B) Derived metrics from image processing and computer vision to describe colony complexity including compression ratio (relative reduction in size of data) and 6 descriptive statistics derived from the Sobel–Feldman operator.

To expand on these classic descriptive variables, we next introduce complexity metrics (Figure 3B) used in image processing and computer vision (51). Compression ratio is a single variable that describes how repetitive an image is, i.e. how repetitive are certain patterns in the pixel distribution and how many times do they occur? Specifically, it is a measurement of the reduction in size of pixel data produced by, in our case, the traditional JPEG compression algorithm. An alternate approach to quantifying image complexity is using the Sobel–Feldman operator, a popular method to detect edges in images due to its ability to detect gradients. At every pixel in an image, it calculates the pixel intensity of the vertical and horizontal adjacent pixels and expresses the results as vectors. You can then plot these gradient vectors creating a unique profile for every colony.

183 These profiles can then be summarized by describing their distribution (mean, median, standard
184 deviation, skew, root mean square, and kurtosis). Unsurprisingly, we find that many of these
185 metrics are highly correlated (Supplemental Figure 1).
186
187 Figure 3 summarizes variation in our metrics across both strains and replicates. In Figure 4, we
188 assess the extent of variation using coefficients of variation (CV = mean / standard deviation),
189 across replicates (black bars) and across strains (grey bars). With the exception of the eccentricity
190 and circularity metrics, we see coefficients of variation across replicates are low (standard
191 deviation << mean), and less than the coefficient of variation across strains. This pattern of
192 variation indicates that colony morphology under common laboratory conditions is a robust,
193 repeatable phenotype on the level of individual strains, and therefore forms a potential basis for
194 strain classification.



195
196 **Figure 4. Morphological metrics are generally stable across replicates.** Coefficient of variation (mean
197 / standard deviation) across replicates (black) and across strains (grey). With the exception of the

198 eccentricity and circularity metrics, coefficients of variation across replicates are low (standard deviation
199 << mean), and less than the coefficient of variation across strains.
200
201 *Classifying strain identity from image data*
202 Considering our initial results indicating robust morphological traits for each strain, we reasoned
203 that morphological data could be used to classify colonies via a deep learning approach. Deep
204 convolutional neural networks (D-CNNs) typically require larger datasets than those found in
205 biology- usually a minimum of thousands of examples to train properly (52). To address this issue,
206 we applied transfer learning (50) in the form of the VGG-19 D-CNN pre-trained on the ImageNet
207 (25) dataset with our own custom top layers catered to our colony dataset (outlined in
208 Supplemental Figure 2) and data augmentation techniques (see methods for details) (48,49) that
209 increase our original 266 images into a dataset with a total 38,304 images. This dataset was split
210 into 74% training and 26% testing with three replicate colonies in the training pool and one
211 replicate colony in the testing. We then test a variety of training approaches by varying the layers
212 trained, the learning rates, and the size of the dataset. Our first approach is to only train the added
213 custom layers of our deep learning architecture and use a default learning rate of 10^{-3} (Table 1A,
214 "feature-extractor"). Our second approach is to train all the layers with a smaller learning rate of
215 10^{-5} (Table 1A, "fine-tuning") as the larger the learning rate, the more unstable the training. In
216 addition to these, we also tested how dependent performance was on the amount of data provided
217 by using only half of the dataset (using only the colonies with even numbered labels; Table 1B,
218 "reduced-feature-extractor" and "reduced-fine-tuning"). To determine the best approach for our
219 dataset, we measure and compare the maximum performance metrics during validation testing,
220 and, by training these groups five times each, measure how consistent the results are (Table 1).
221 See Methods for more detail on our data preparation and model development. Our performance

metrics are accuracy (the number of correct predictions divided by the total number of predictions x100) and loss (a summation of the errors, calculated via cross-entropy, made for each sample). We chose to use cross-entropy for loss instead of negative log-likelihood since it automatically transforms the raw outputs of the neural network into probabilities (via a softmax activation).

We find that the "fine-tuning" approach (whole architecture training with lower learning rate) produced the best performance across all metrics and on both the whole and reduced dataset, producing significantly higher accuracy (Two Sample t-test, $t(18) = 15.3$, $p = 9e\text{-}12$) and lower loss (Two Sample t-test, $t(18) = -8.7$, $p = 7.4e\text{-}8$) (Table 1). Comparing the whole versus the reduced dataset for the "fine-tuning" approach, we see no significant difference in validation accuracy (Two Sample t-test, $t(8) = 1.2564$, $p = 0.2444$) or in validation loss (Two Sample t-test, $t(8) = -0.47707$, $p = 0.6461$), indicating that our approach is robust to a halving of our sample size.

Table 1. Performance metrics for different training approaches to classifying strains via colony images. We tested two training methods: the "feature-extractor" method only trains the added custom layers of our deep learning architecture and uses a default learning rate of 10^{-3}; the "fine-tuning" method trains all the layers and uses a smaller learning rate of 10^{-5}. Additionally, we tested how reliant these results were on the number of colonies in the dataset by training on the whole dataset (A) versus a reduced dataset (B) that consists of half the number of colonies. The model was then trained five independent times per approach to access consistency and performance was quantified via accuracy (the number of correct predictions divided by the total number of predictions x100) and loss (a summation of the errors made for each sample). 'Top' indicates the single-best performance among the five sessions. Asterisks indicate significantly higher average performance (Two Sample t-test, df = 18, $p < 0.01$) for 'fine-tuning' compared to 'feature extractor' methods, for both the whole (A) and reduced (B) datasets.

A) Whole strain dataset						
Method	**Top Accuracy (%)**	**Average Accuracy (%)**	**Standard deviation of Accuracy (%)**	**Top Loss**	**Average Loss**	**Standard deviation of Loss**
feature-extractor	85.69	84.27	1.219	0.7160	0.7890	0.0531
fine-tuning	94.77	94.18*	0.413	0.1448	0.2352*	0.0495
B) Reduced strain dataset						
Method	**Top Accuracy (%)**	**Average Accuracy (%)**	**Standard deviation of Accuracy (%)**	**Top Loss**	**Average Loss**	**Standard deviation of Loss**
reduced-feature-extractor	86.90	84.38	1.561	0.6011	0.7318	0.0725
reduced-fine-tuning	95.69	93.86*	1.092	0.1376	0.2303*	0.0640

The performance metrics in Table 1 indicate that while most classification calls are accurate, there remain a number of mistaken calls in our validation tests. To look more closely at these errors, we present in Figure 4.5 a reduced confusion matrix from our most successful and most consistent model (fine-tuning using all of the data, top accuracy run (94.77%), Table 1A). Figure 5 is a reduced confusion matrix as we only show strains involved in mistaken calls. Strains 25 and 344 had the lowest validation accuracy (44% and 42%, respectively). The confusion with strain 25 was likely due to the qualitative difference between the training and testing data. Due to chance, the colony replicate assigned to the testing set varied more in appearance than the three replicate colonies assigned to the training. In contrast, the confusion with classifying strain 334 was only with strain 337 and is likely due to variation across all colony replicates.

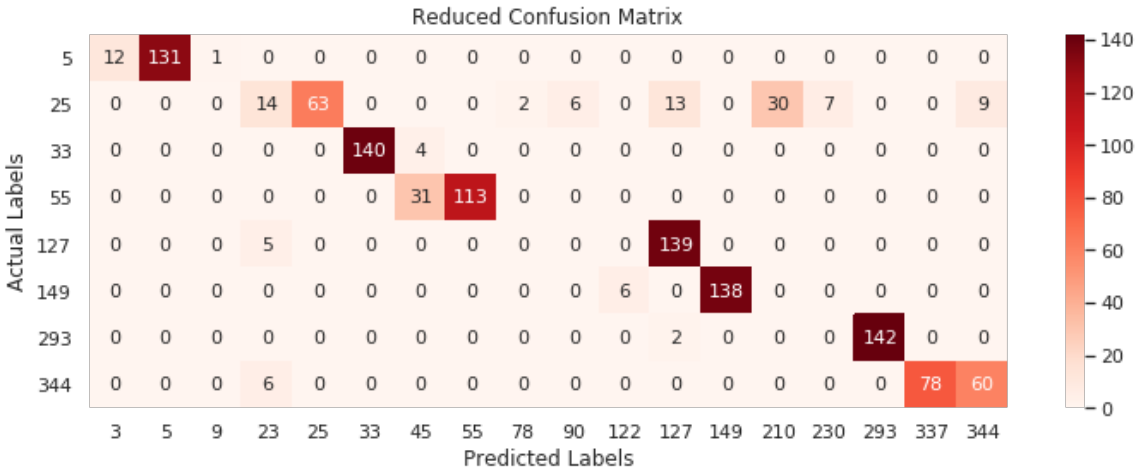


Figure 5. Reduced confusion matrix for the best performing 'fine-tuning' D-CNN model predictions. Of the 69 strains in this study, 51 strains had 100% precision when classified. For brevity, this table includes only the strains that were misclassified by our model instead of the whole 69 x 69 matrix.

Discussion

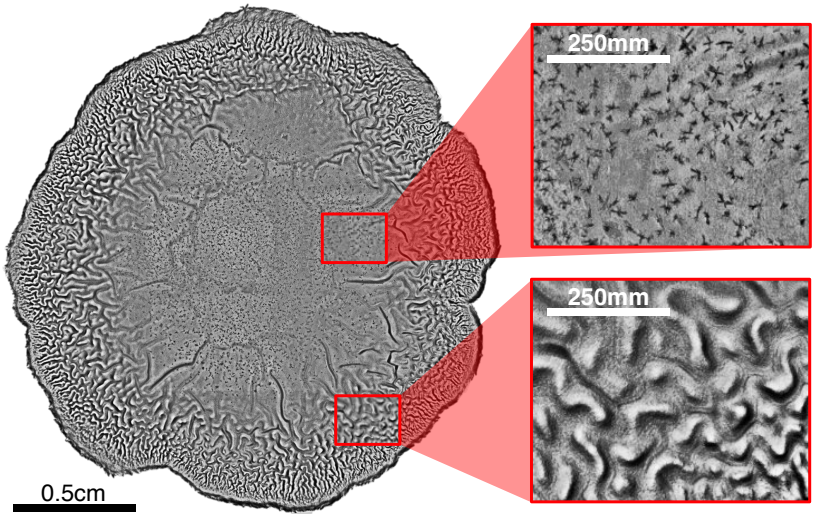
Our results show that *P. aeruginosa* strains have a characteristic morphological 'fingerprint' (Figures 2-4) that enables accurate strain classification via a custom deep convolutional neural network (Table 1, Figure 5). When trained on morphological data, our model can classify strain identity with a validation accuracy that repeatably exceeds 94% despite what is considered a "data starved" dataset. Our work points to extensions towards predicting phenotypes of interest (e.g. antibiotic resistance, virulence), and suggests that sample size limitations may be less restrictive than previously thought for deep learning applications in biology. In the following paragraphs we will review these implications, while addressing caveats and next steps for this research.

The existence of a morphological 'fingerprint' that can be classified via machine learning architecture opens the potential to relate strain identity to key phenotypes of interest, such as antimicrobial resistance (AMR) or virulence. While species differ in their levels and types of

276 virulence (53), there are also significant differences in virulence between strains of the same
277 species (54,55). In so far as AMR or virulence is a stable and repeatable property of a strain, our
278 classification model could simply "look up" the recorded values in a dataset based on the strain
279 prediction. Yet this "look up" approach is vulnerable to variation within strains due for instance to
280 gain or loss in virulence factor and AMR genes, or mutations in the regulatory control of these
281 genes.
282
283 To address the challenge of predicting phenotypes such as virulence or AMR, a future direction
284 would be to extend our deep learning approach to directly predict phenotypes of interest, either
285 through a classification (qualitative trait) or regression (quantitative trait) framework. Our current
286 analysis shows that morphological 'fingerprints' are sufficient to identify individual strains, but
287 this does not directly imply that we can use image data to successfully predict disease traits
288 (analogously, human fingerprints are sufficient to identify individuals, but are not considered
289 predictive of human disease traits). We speculate that colony 'fingerprints' are likely to allow for
290 trait prediction, due to paths of common causality from bacterial modes of growth to colony
291 morphology and disease phenotypes. Given the established connection between biofilm growth
292 and antibiotic tolerance (56), it is possible that our default growth settings (Figure 4.2) will reveal
293 differences in patterning that correlate with differences in antibiotic resistance (and/or tolerance).
294 We further anticipate that our approach would become more discriminatory if imaging data was
295 also collected in the presence/absence of relevant stressors of interest (e.g. antibiotics or immune
296 effectors).
297

298 Deep learning methods are commonly described as 'black box' methodologies (57), presenting
299 challenges for model interrogation and mathematical justification for sample size requirements.
300 Even when classification through deep learning has a monotonically increasing prediction
301 accuracy, the nature of the image collection process in biological research will limit the total size
302 of the dataset. This is a contrast to machine learning applications in other fields that have
303 continually and often exponentially increasing datasets (e.g., spending habits and surveillance
304 cameras). Considering these challenges, arguably our best alternatives for increasing the size of
305 biological datasets for image-based prediction in microbiology are data augmentation and transfer
306 learning, which we combined in this study. Our algorithm transfers learning from the canonical
307 ImageNet dataset while also using standard data augmentation techniques (rotations, flips,
308 reflections.). Our results in Table 1 indicate that, with our combination of data augmentation and
309 transfer learning, we could have reached a similar success with classifying even if we only had
310 half the number of strains. Future tests to explore the amount of data needed for accurate
311 classification could be run by reducing the size of the dataset by decreasing the amount of data
312 augmentation performed or by reducing the number of replicate colonies used. Moving forward,
313 generative adversarial networks, a class of machine learning frameworks that learns to generate
314 new data by competing two neural networks, is promising technique that could supplement current
315 data augmentation efforts (58,59).
316
317 Prior to our data augmentation, we downsampled the original dataset in order to match the image
318 dimensions of the ImageNet dataset which we pretrained on and to minimize computational time.
319 The original dataset contains colony images of 12 million to 183 million pixels (depending on
320 colony size) while our downsized colony images each contained a standardized 60 thousand pixels.

321 The success of strain classification based on these heavily down-sampled images suggests that
322 classification is possible based only on more 'macro' scale morphological features that are still
323 discriminable at this resolution. Yet this down-sampling is essentially throwing away a rich array
324 of more micro-scale image features (Figure 6), and so opens the possibility for future analyses to
325 explore these features. The inset panels in Figure 6 illustrate striking local and finer-scale features
326 that are present across strains. These panels suggest an additional data augmentation option, where
327 each colony image is processed into multiple 60K chunks, sampling at different locations within
328 the image and at different spatial scales. Comparing model performance using 60K images taken
329 at different spatial scales could provide valuable information on what are the most diagnostic
330 spatial scales.



331

332 **Figure 1. Avenues for finer scale image segmentation and augmentation.** A) Higher resolution chunks
333 of a full-sized colony.

334

335 Our analysis to date implicitly assumes that our image data represent independent samples of *P.*
336 *aeruginosa* morphology, which due to phylogenetic relationships among these strains is not the
337 case. A simple null expectation given these phylogenetic relationships is that strains that are most

338 morphologically similar (i.e., similar morphological metrics, Figure 3; and more commonly
339 confused in classification, Figure 6) will tend to have a shorter phylogenetic distance. Under this
340 phylogenetic null model, strains with minimal phylogenetic differences will have overlapping
341 morphological metrics and will not provide dependable classification results. We do not expect
342 this pattern to hold uniformly, as we know that single-gene changes can generate large-scale
343 changes to colony morphology (11,12,44). Future work can systematically explore the impacts of
344 single gene mutations through automated screening of transposon mutagenesis libraries (60),
345 potentially aiding in the discovery of previously unrecognized genes involved in colony / biofilm
346 formation. In summary, the level of classification accuracy achieved in this work illustrates the
347 potential of image-based prediction tools, deep learning, transfer learning, and data augmentation
348 in the characterization of bacterial strains.
349

350 **Methods**

351 *Strain Culturing*

352 This study uses a collection of 69 clinical and environmental *P. aeruginosa* isolates (45–47), each
353 grown and imaged in quadruplicate. We used Luria-Bertani (LB) agar (1.5%) plates supplemented
354 with 2% Congo Red, a dye commonly used to characterize biofilm formation which binds to
355 compounds in the extracellular biofilm matrix. In an effort to minimize variation in plate
356 preparation and the age of the plates, 20ml volume plates were made 24 hours before each
357 experiment with a large, motorized pipette and kept in a sealed plastic sleeve at 4C. Strains were
358 inoculated into LB broth and incubated shaking overnight at 37C. The day of the experiment, each
359 plate was then sliced into 4 sections using a sterilized razor and tweezers, which both limited
360 colony outgrowth and prevented the diffusion of small molecules between colonies. 5µl of the

361 overnight culture was then spotted onto pre-prepared, sectioned LB agar Congo Red plates. After
362 the spots dried, plates were parafilmed to retain moisture and placed in a 37C static incubator. All
363 strains were incubated at 37C for 72 hours, which was determined during the pilot experiment to
364 be enough time to reveal major morphological differences while minimizing outward expansion.
365
366 *Colony Imaging.*
367 After 72 hours, strains were imaged on a Nikon Eclipse Ti inverted microscope at 4X using the
368 DS-Fi2 colored camera. This results in a pixel-to-size ratio is 2.07uM per pixel. Large, stitched
369 images were required to scan an entire colony and were generated with Nikon's built-in software.
370 Some replicate colonies had to be excluded from analysis due to error (strains 13, 47, 78, 86, 122,
371 124, 132, 210 have 3 replicates and strain 329 has 2 replicates). This resulted in a dataset of 266
372 *P. aeruginosa* colony images that ranged in size from 12,644,652 pixels (approximately 0.74cm x
373 0.74cm) to 183,329,280 pixels (approximately 2.8cm x 2.8cm).
374
375 *Morphological and Complexity Metrics*
376 For the morphological and complexity measurements, colony images were analyzed in
377 Mathematica. Images were scaled down by a factor of four, to make computations timelier, and
378 the backgrounds were cropped out by using a watershed transformation, binarization, and selection
379 of the contiguous non-zero area surrounding the colony. The morphological metrics (Area,
380 EquivalentDiskRadius, PerimeterLength, MeanCentroidDistance, Eccentricity, Circularity,
381 BoundingDiskCenter, and MeanCaliperDiameter) were all calculated using the native
382 Mathematica commands. For the complexity metrics, CompRatio was calculated by comparing
383 the number of bytes in the uncompressed and compressed (using the native Mathematica

384 command) colony images. A Sobel-Freidman operator was applied in both x and y dimensions and
385 Sobel histograms were generated. MeanSobel, MedianSobel, SDSobel, RootMSSobel,
386 SkewSobel, and KurtosisSobel were all then calculated using the native Mathematica commands.
387
388 *Machine Learning: Data processing, Augmentation, and Splitting*
389 To exploit existing transfer learning tools and to conform to Google Colab data processing
390 limitations, we downsampled all images to size 244x244. This size was selected to allow use of
391 the pretrained ImageNet, which is trained on standardized images of size 244×244. To standardize
392 input shape, we used inter-area interpolation. One of the main challenges with deep-based
393 prediction from images of this sort lies in the restricted number of available samples. Traditionally
394 in deep learning, sufficient datasets require thousands of examples to train properly. For instance,
395 the ImageNet was trained on 14 million images distributed across 1000 classes (52). To address
396 this issue, we applied several data augmentation techniques to inflate our original 266 images into
397 a dataset with a total 38,304 images. Images were taken and rotated 10 degrees counterclockwise,
398 then flipped vertically, horizontally, and mirrored, producing a total of 4 images. This process was
399 repeated until the image rotated back to its original starting point after 36 iterations. With 4 images
400 for each iteration, 144 total images were produced from a single image. Recent surveys of data
401 augmentation techniques have found rotation and mirror to be a simple, well-evaluated and reliable
402 technique (48). Several of these surveys have also addressed an additional benefit of data
403 augmentation: its ability to decrease likelihood of model overfitting without making modifications
404 to the existing architecture (61,62).
405

406 In deep learning, it is customary to split the total dataset into training and validation bins. The
407 standard practice is to take 70-90% of the available samples into a training bin and the remaining
408 into the validation bin. Due to slight variation in samples per strain and the use of data
409 augmentation, a custom procedure was used to curate our training and validation datasets. All but
410 one of each image and their corresponding augmentations within a strain's samples were placed
411 into the training bin and the remainder placed in the validation bin. Given 4 samples for a strain,
412 3 samples were extracted from the set, then augmentation was performed creating a total of 432
413 samples. These 432 samples were then added to the training dataset. Augmentation was then
414 performed on the remaining image, creating 144 samples which were then added to the validation
415 dataset. This assures that the final sample and its augmentations are unique to the validation
416 dataset. Fortunately, this results in a fairly optimal split of 74.06% training and 25.94% validation.
417

418 *Deep Learning Architecture*

419 Our deep learning architecture (Supplemental Figure 4.2) is built using standard transfer learning
420 methodology. The Keras API VGG-19 model pretrained on ImageNet is loaded with the top
421 removed. The top is replaced with a new flatten layer followed by two pairs of dense and dropout
422 layers and a final softmax dense output layer. The new dense layers were changed to size 1024
423 from the original 4096. This choice was made due to the size of our dataset. ImageNet is comprised
424 of 1000 classes whereas our dataset is only comprised of 69, thus motivating a major reduction in
425 layer size. The dense layers were built using Rectified Linear Unit activations, inspired by the
426 dense layers of the original VGG-19 paper (52). In addition to smaller dense layers, the dropout
427 layers were added to further narrow the network into the output layer. The choice to add dropout
428 layers was additionally motivated by the transfer learning research performed by (63). The final

429 dense layer used a softmax activation function to convert the output into a probability distribution
430 across 69 neurons. Each neuron corresponds to a unique strain identification within the dataset.
431 When the output layer is activated, the resulting probability distribution is commonly referred to
432 as the confidence vector.
433
434 *Transfer Learning and Training*
435 We investigate a combination of transfer learning and data augmentation methods and evaluate
436 their respective average performance. The experiments are divided among four groups, and each
437 was done with five independent training sessions. The first group, "feature-extractor", is trained
438 with a feature extractor approach which freezes all layers preceding the new layers and uses the
439 Adam optimizer with a default learning rate of 10^-3 (Keras default batch size documentation,
440 available at: https://keras.io/models/model). The second group, "fine-tuning", is trained in fine
441 tuning approach which trains all layers of the network and uses the Adam optimizer but with a
442 smaller learning rate of 10^-5. The two additional groups are set in the same modes respectively
443 but receive a reduced dataset, containing only the samples with even indexes (a 50% reduction),
444 "reduced-feature-extractor" and "reduced-fine-tuning". Each group is trained 5 times with
445 uniquely random weight initialization for the new layers. Through these repeated training sessions,
446 we measure and compare the maximum performance metrics and repeatability between each
447 group.
448
449 Before conducting these experiments, we hand tune the model in fixed extractor mode to tune
450 dropout rates (with respect to new layers), motivated by (16). The intuition for this is motivated
451 by the Keras API dropout layers only effecting the connections of the previous layer. Thus, the

452 dropout rates are relatively independent of the transfer learning in which the network is trained
453 and only effect the newly added dense layers. Since it is significantly faster to train in a fixed
454 feature extractor setting, it is more optimal to first tune the dropout rates in this setting before
455 performing further experiments. The dropout rates were adjusted to a final rate of 55% for both
456 dropout layers. We use a categorical cross entropy loss function to calculate loss and chose to use
457 cross-entropy for loss instead of negative log-likelihood since it automatically transforms the raw
458 outputs of the neural network into probabilities (via a softmax activation). The paper by (63)
459 motivated our use of the Adam optimizer for back propagation. Batch size was left unmodified,
460 initialized with the Keras default value of 32. A Keras early stopping callback was initialized to
461 halt training when validation loss increased across an epoch.
462
463 Our choice of model approach was primarily motivated by the work of (64)who used a pretrained
464 network to classify PAP-smear (65) and 2D-Hela (66) datasets and share the small number of
465 samples (small by traditional deep learning standards with fewer than 100 images per class, yet an
466 order of magnitude more than in our work). In particular, (64) used a VGG-16 architecture
467 pretrained on the approximately 14 million images (distributed across 1000 classes) of the
468 ImageNet dataset (25). This architecture had been previously shown to be able to outperform
469 humans when classifying a collection of images (63). On the other hand, VGG-19 has a slight
470 increase in performance on the ImageNet dataset (52). Consequently, we decided to use the VGG-
471 19 architecture instead of the VGG-16.
472
473 **Acknowledgements**

We thank Jennifer Farrell for culturing the mixed community in Figure 1 and members of the Center for Microbial Dynamics and Infection (CMDI), in particular Dr. Peter Yunker, Dr. Conan Zhao, and Dr. Yifei Wang, for valuable comments and discussion on this work. This research was supported by the National Science Foundation Graduate Research Fellowship Program under grant no. DGE-1650044.

References

1. Koch R. Ueber den augenblicklichen Stand der bakteriologischen Choleradiagnose. Zeitschrift für Hygiene und Infektionskrankheiten 1983 14:1. 1893 Dec. Available from: https://link.springer.com/article/10.1007/BF02284324

2. Madigan MT, Brock TD. Brock biology of microorganisms. 12th editi. Biology of Microorganisms. San Francisco, CA: Pearson/Benjamin Cummings; 2009.

3. Bergey D, Krieg NR, Holt JG. Bergey's manual of systematic bacteriology. Baltimore, MD: Williams & Wilkins; 1984.

4. Rainey PB, Travisano M. Adaptive radiation in a heterogeneous environment. Nature 1998 394:6688 . 1998 Jul 1;394(6688):69–72. Available from: https://www.nature.com/articles/27900

5. Woo PCY, Lau SKP, Teng JLL, Tse H, Yuen KY. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. Clin Microbiol Infect . 2008 [cited 2022 Jul 19];14(10):908–34. Available from: https://pubmed.ncbi.nlm.nih.gov/18828852/

6. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer KH, et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol . 2014 [cited 2022 Jul 19];12(9):635–45. Available from: https://pubmed.ncbi.nlm.nih.gov/25118885/

7. Woese CR. Bacterial evolution. Microbiol Rev . 1987 Jun [cited 2022 Jul 19];51(2):221. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC373105/

8. Woese C. The universal ancestor. Proceedings of the National Academy of Sciences . 1998 Jun 9 [cited 2022 Jul 19];95(12):6854–9. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.95.12.6854

9. Güell M, Yus E, Lluch-Senar M, Serrano L. Bacterial transcriptomics: what is beyond the RNA horiz-ome? Nature Reviews Microbiology 2011 9:9 . 2011 Aug 12 [cited 2022 Jul 19];9(9):658–69. Available from: https://www.nature.com/articles/nrmicro2620

10. Kalziqi A, Yanni D, Thomas J, Ng SL, Vivek S, Hammer BK, et al. Immotile Active Matter: Activity from Death and Reproduction. Phys Rev Lett . 2018 Jan 5 [cited 2022 Jul 17];120(1):018101. Available from: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.120.018101

11. Starkey M, Hickman JH, Ma L, Zhang N, de Long S, Hinz A, et al. Pseudomonas aeruginosa rugose small-colony variants have adaptations that likely promote persistence in the cystic fibrosis lung. J Bacteriol . 2009 Jun [cited 2016 May 18];191(11):3492–503. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19329647

12. Recinos DA, Sekedat MD, Hernandez A, Cohen TS, Sakhtah H, Prince AS, et al. Redundant phenazine operons in Pseudomonas aeruginosa exhibit environment-dependent expression and differential roles in pathogenicity. Proc Natl Acad Sci U S A . 2012 Nov 20 [cited 2022 Jul 3];109(47):19420–5. Available from: www.pnas.org/cgi/doi/10.1073/pnas.1213901109

13. Banada PP, Guo S, Bayraktar B, Bae E, Rajwa B, Robinson JP, et al. Optical forward-scattering for detection of Listeria monocytogenes and other Listeria species. Biosens Bioelectron. 2007 Mar 15;22(8):1664–71.

14. Tang Y, Kim H, Singh AK, Aroonnual A, Bae E, Rajwa B, et al. Light Scattering Sensor for Direct Identification of Colonies of Escherichia coli Serogroups O26, O45, O103, O111, O121, O145 and O157. PLoS One . 2014 Aug 19 [cited 2022 Jul 17];9(8). Available from: /pmc/articles/PMC4138183/

15. Singh AK, Sun X, Bai X, Kim H, Abdalhaseib MU, Bae E, et al. Label-free, non-invasive light scattering sensor for rapid screening of Bacillus colonies. J Microbiol Methods. 2015 Feb 1;109:56–66.

16. Alsulami TS, Zhu X, Abdelhaseib MU, Singh AK, Bhunia AK. Rapid detection and differentiation of Staphylococcus colonies using an optical scattering technology. Anal Bioanal Chem . 2018 Sep 1 [cited 2022 Jul 17];410(22):5445–54. Available from: https://pubmed.ncbi.nlm.nih.gov/29796901/

17. Zieliński B, Plichta A, Misztal K, Spurek P, Brzychczy-Włoch M, Ochońska D. Deep learning approach to bacterial colony classification. PLoS One . 2017 Sep 1 [cited 2022 Jul 17];12(9):e0184554. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184554

18. Sousa AM, Pereira MO, Lourenço A. MorphoCol: An ontology-based knowledgebase for the characterisation of clinically significant bacterial colony morphologies. J Biomed Inform . 2015 Jun 1 [cited 2022 Jul 17];55:55–63. Available from: https://pubmed.ncbi.nlm.nih.gov/25817920/

19. Fratamico PM, Strobaugh TR, Medina MB, Gehring AG. Detection of Escherichia coli 0157:H7 using a surface plasmon resonance biosensor. Biotechnology Techniques 1998 12:7 . 1998 [cited 2022 Aug 23];12(7):571–6. Available from: https://link.springer.com/article/10.1023/A:1008872002336

20. Perkins EA, Squirrell DJ. Development of instrumentation to allow the detection of microorganisms using light scattering in combination with surface plasmon resonance. Biosens Bioelectron . 2000 Jan [cited 2022 Aug 23];14(10–11):853–9. Available from: https://pubmed.ncbi.nlm.nih.gov/10945460/

21. Turra G, Arrigoni S, Signoroni A. CNN-Based Identification of Hyperspectral Bacterial Signatures for Digital Microbiology. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) . 2017 [cited 2022 Jul 17];10485 LNCS:500–10. Available from: https://link.springer.com/chapter/10.1007/978-3-319-68548-9_46

22. Andreini P, Bonechi S, Bianchini M, Mecocci A, Scarselli F. A deep learning approach to bacterial colony segmentation. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) . 2018 [cited 2022 Jul 17];11141 LNCS:522–33. Available from: https://link.springer.com/chapter/10.1007/978-3-030-01424-7_51

23. Geirhos R, Janssen DHJ, Schütt HH, Rauber J, Bethge M, Wichmann FA. Comparing deep neural networks against humans: object recognition when the signal gets weaker. 2017 Jun 21 [cited 2022 Jul 17]; Available from: https://arxiv.org/abs/1706.06969v2

24. Buetti-Dinh A, Galli V, Bellenberg S, Ilie O, Herold M, Christel S, et al. Deep neural networks outperform human expert's capacity in characterizing bioleaching bacterial biofilm composition. Biotechnology Reports. 2019 Jun 1;22:e00321.

25. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis . 2014 Sep 1 [cited 2022 Jul 17];115(3):211–52. Available from: https://arxiv.org/abs/1409.0575v3

26. Deng J, Berg AC, Li K, Fei-Fei L. What Does Classifying More than 10,000 Image Categories Tell Us? In: Proceedings of the 11th European Conference on Computer Vision: Part V. Berlin, Heidelberg: Springer-Verlag; 2010. p. 71–84. (ECCV'10).

27. Russakovsky O, Fei-Fei L. Attribute learning in large-scale datasets. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) . 2012 [cited 2022 Jul 18];6553 LNCS(PART 1):1–14. Available from: https://collaborate.princeton.edu/en/publications/attribute-learning-in-large-scale-datasets

28. Russakovsky O, Deng J, Huang Z, Berg AC, Fei-Fei L. Detecting avocados to Zucchinis: What have we done, and where are we going? Proceedings of the IEEE International Conference on Computer Vision . 2013 [cited 2022 Jul 18];2064–71. Available from: https://collaborate.princeton.edu/en/publications/detecting-avocados-to-zucchinis-what-have-we-done-and-where-are-w

29. Balki I, Amirabadi A, Levman J, Martel AL, Emersic Z, Meden B, et al. Sample-Size Determination Methodologies for Machine Learning in Medical Imaging Research: A Systematic Review. Canadian Association of Radiologists Journal . 2019 Nov 1 [cited

584 2022 Jul 17];70(4):344–53. Available from:
585 https://journals.sagepub.com/doi/10.1016/j.carj.2019.06.002?url_ver=Z39.88-
586 2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub++0pubmed

587 30. Diggle SP, Whiteley M. Microbe profile: Pseudomonas aeruginosa: Opportunistic
588 pathogen and lab rat. Microbiology (United Kingdom) . 2020 Oct 10 [cited 2022 Jul
589 17];166(1):30–3. Available from:
590 https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.000860

591 31. Moradali MF, Ghods S, Rehm BHA. Pseudomonas aeruginosa Lifestyle: A Paradigm for
592 Adaptation, Survival, and Persistence. Front Cell Infect Microbiol . 2017 Feb 15 [cited
593 2022 Jul 17];7(FEB):39. Available from: /pmc/articles/PMC5310132/

594 32. Nathwani D, Raman G, Sulham K, Gavaghan M, Menon V. Clinical and economic
595 consequences of hospital-acquired resistant and multidrug-resistant Pseudomonas
596 aeruginosa infections: A systematic review and meta-analysis. Antimicrob Resist Infect
597 Control . 2014 Oct 20 [cited 2022 Jul 17];3(1):1–16. Available from:
598 https://aricjournal.biomedcentral.com/articles/10.1186/2047-2994-3-32

599 33. Elborn JS. Cystic fibrosis. The Lancet. 2016 Nov 19;388(10059):2519–31.

600 34. Pohl S, Klockgether J, Eckweiler D, Khaledi A, Schniederjans M, Chouvarine P, et al.
601 The extensive set of accessory Pseudomonas aeruginosa genomic components. FEMS
602 Microbiol Lett . 2014 Jul 1 [cited 2022 Jul 18];356(2):235–41. Available from:
603 https://academic.oup.com/femsle/article/356/2/235/542446

604 35. Freschi L, Vincent AT, Jeukens J, Emond-Rheault JG, Kukavica-Ibrulj I, Dupont MJ, et
605 al. The Pseudomonas aeruginosa Pan-Genome Provides New Insights on Its Population
606 Structure, Horizontal Gene Transfer, and Pathogenicity. Genome Biol Evol . 2019 Jan 1
607 [cited 2022 Jul 17];11(1):109–20. Available from:
608 https://pubmed.ncbi.nlm.nih.gov/30496396/

609 36. Poulsen BE, Yang R, Clatworthy AE, White T, Osmulski SJ, Li L, et al. Defining the core
610 essential genome of Pseudomonas aeruginosa. Proc Natl Acad Sci U S A . 2019 May 14
611 [cited 2022 Jul 17];116(20):10072–80. Available from: /pmc/articles/PMC6525520/

612 37. Lebreton F, Snesrud E, Hall L, Mills E, Galac M, Stam J, et al. A panel of diverse
613 Pseudomonas aeruginosa clinical isolates for research and development. JAC Antimicrob
614 Resist . 2021 Sep 30 [cited 2022 Jul 18];3(4). Available from:
615 https://academic.oup.com/jacamr/article/3/4/dlab179/6458700

616 38. Kiyaga S, Kyany'a C, Muraya AW, Smith HJ, Mills EG, Kibet C, et al. Genetic Diversity,
617 Distribution, and Genomic Characterization of Antibiotic Resistance and Virulence of
618 Clinical Pseudomonas aeruginosa Strains in Kenya. Front Microbiol. 2022 Mar 14;13:699.

619 39. Kirisits MJ, Prost L, Starkey M, Parsek MR. Characterization of Colony Morphology
620 Variants Isolated from Pseudomonas aeruginosa Biofilms. Appl Environ Microbiol . 2005
621 Aug [cited 2022 Jul 17];71(8):4809. Available from: /pmc/articles/PMC1183349/

622 40. Ikeno T, Fukuda K, Ogawa M, Honda M, Tanabe T, Taniguchi H. Small and rough colony
623 pseudomonas aeruginosa with elevated biofilm formation ability isolated in hospitalized
624 patients. Microbiol Immunol . 2007 [cited 2022 Jul 17];51(10):929–38. Available from:
625 https://pubmed.ncbi.nlm.nih.gov/17951982/

626 41. Rakhimova E, Munder A, Wiehlmann L, Bredenbruch F, Tümmler B. Fitness of Isogenic
627 Colony Morphology Variants of Pseudomonas aeruginosa in Murine Airway Infection.
628 PLoS One . 2008 Feb 27 [cited 2022 Jul 17];3(2):1685. Available from:
629 /pmc/articles/PMC2246019/

630 42. Azimi S, Roberts AEL, Peng S, Weitz JS, McNally A, Brown SP, et al. Allelic
631 polymorphism shapes community function in evolving Pseudomonas aeruginosa
632 populations. The ISME Journal 2020 14:8 . 2020 Apr 27 [cited 2022 Jul 17];14(8):1929–
633 42. Available from: https://www.nature.com/articles/s41396-020-0652-0

634 43. Vanderwoude J, Fleming D, Azimi S, Trivedi U, Rumbaugh KP, Diggle SP. The
635 evolution of virulence in Pseudomonas aeruginosa during chronic wound infection.
636 Proceedings of the Royal Society B . 2020 Oct 28 [cited 2022 Jul
637 17];287(1937):20202272. Available from:
638 https://royalsocietypublishing.org/doi/10.1098/rspb.2020.2272

639 44. Boucher JC, Yu H, Mudd MH, Deretic V. Mucoid Pseudomonas aeruginosa in cystic
640 fibrosis: characterization of muc mutations in clinical isolates and analysis of clearance in
641 a mouse model of respiratory infection. Infect Immun . 1997 [cited 2022 Jul
642 20];65(9):3838–46. Available from: https://pubmed.ncbi.nlm.nih.gov/9284161/

643 45. Pirnay JP, Bilocq F, Pot B, Cornelis P, Zizi M, van Eldere J, et al. Pseudomonas
644 aeruginosa Population Structure Revisited. PLoS One . 2009 Nov 13 [cited 2022 Jul
645 17];4(11):e7740. Available from:
646 https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0007740

647 46. Dettman JR, Rodrigue N, Aaron SD, Kassen R. Evolutionary genomics of epidemic and
648 nonepidemic strains of Pseudomonas aeruginosa. Proc Natl Acad Sci U S A . 2013 Dec 24
649 [cited 2022 Jul 17];110(52):21065–70. Available from:
650 www.pnas.org/cgi/doi/10.1073/pnas.1307862110

651 47. Shrestha SD, Guttman DS, Perron GG. Draft Genome Sequences of 10 Environmental
652 Pseudomonas aeruginosa Strains Isolated from Soils, Sediments, and Waters. Genome
653 Announc . 2017 [cited 2022 Jul 17];5(34). Available from:
654 https://journals.asm.org/doi/10.1128/genomeA.00804-17

655 48. Mikołajczyk A, Grochowski M. Data augmentation for improving deep learning in image
656 classification problem. 2018 International Interdisciplinary PhD Workshop, IIPhDW
657 2018. 2018 Jun 18;117–22.

658 49. Wong SC, Gatt A, Stamatescu V, McDonnell MD. Understanding Data Augmentation for
659 Classification: When to Warp? 2016 International Conference on Digital Image
660 Computing: Techniques and Applications, DICTA 2016. 2016 Dec 22;

50. Pratt LY. Discriminability-Based Transfer between Neural Networks. In: Hanson S, Cowan J, Giles C, editors. Advances in Neural Information Processing Systems . Morgan-Kaufmann; 1992. Available from: https://proceedings.neurips.cc/paper/1992/file/67e103b0761e60683e83c559be18d40c-Paper.pdf

51. Yu H, Winkler S. Image complexity and spatial information. 2013 5th International Workshop on Quality of Multimedia Experience, QoMEX 2013 - Proceedings. 2013;12–7.

52. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings . 2014 Sep 4 [cited 2022 Jul 17]; Available from: https://arxiv.org/abs/1409.1556v6

53. Shaw LP, Wang AD, Dylus D, Meier M, Pogacnik G, Dessimoz C, et al. The phylogenetic range of bacterial and viral pathogens of vertebrates. Mol Ecol . 2020 Sep 1 [cited 2022 Jul 17];29(17):3361–79. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/mec.15463

54. Subedi D, Vijay AK, Kohli GS, Rice SA, Willcox M. Comparative genomics of clinical strains of Pseudomonas aeruginosa strains isolated from different geographic sites. Scientific Reports 2018 8:1 . 2018 Oct 23 [cited 2022 Jul 17];8(1):1–14. Available from: https://www.nature.com/articles/s41598-018-34020-7

55. Medina-Rojas M, Stribling W, Snesrud E, Garry BI, Li Y, Gann PM, et al. Comparison of Pseudomonas aeruginosa strains reveals that Exolysin A toxin plays an additive role in virulence. Pathog Dis . 2020 Feb 1 [cited 2022 Jul 17];78(1):10. Available from: https://academic.oup.com/femspd/article/78/1/ftaa010/5804881

56. Sharma D, Misba L, Khan AU. Antibiotics versus biofilm: an emerging battleground in microbial communities. Antimicrobial Resistance & Infection Control 2019 8:1 . 2019 May 16 [cited 2022 Jul 25];8(1):1–10. Available from: https://aricjournal.biomedcentral.com/articles/10.1186/s13756-019-0533-3

57. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence 2019 1:5 . 2019 May 13 [cited 2022 Jul 20];1(5):206–15. Available from: https://www.nature.com/articles/s42256-019-0048-x

58. Tolstikhin I, Bousquet O, Schölkopf B, Thierbach K, Bazin PL, de Back W, et al. Generative Adversarial Networks. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) . 2014 Jun 10 [cited 2022 Jul 17];11046 LNCS(NeurIPS):1–9. Available from: https://arxiv.org/abs/1406.2661v1

59. Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. 4th International Conference on

Learning Representations, ICLR 2016 - Conference Track Proceedings . 2015 Nov 19 [cited 2022 Jul 17]; Available from: https://arxiv.org/abs/1511.06434v2

60. Cabeen MT, Leiman SA, Losick R. Colony-morphology screening uncovers a role for the Pseudomonas aeruginosa nitrogen-related phosphotransferase system in biofilm formation. Mol Microbiol . 2016 Feb 1 [cited 2022 Jul 17];99(3):557. Available from: /pmc/articles/PMC5130288/

61. Perez L, Wang J. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. 2017 Dec 13 [cited 2022 Jul 17]; Available from: https://arxiv.org/abs/1712.04621v1

62. Hernández-García A, König P. Further advantages of data augmentation on convolutional neural networks. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) . 2019 Jun 26 [cited 2022 Jul 17];11139 LNCS:95–103. Available from: https://arxiv.org/abs/1906.11052v1

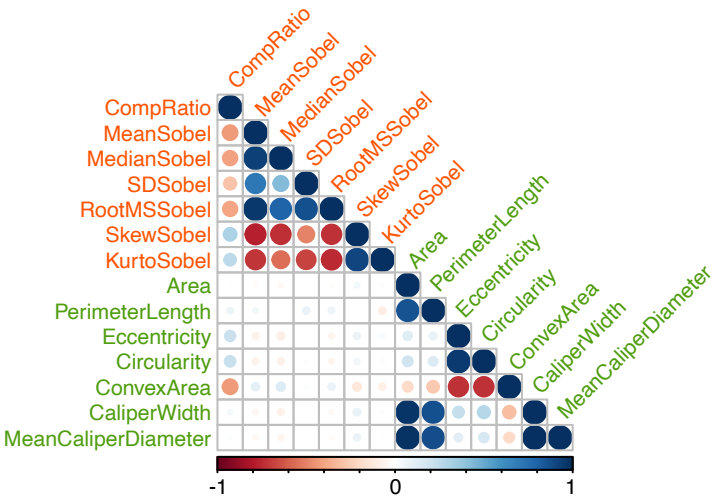
63. Oei RW, Hou G, Liu F, Zhong J, Zhang J, An Z, et al. Convolutional neural network for cell classification using microscope images of intracellular actin networks. PLoS One . 2019 Mar 1 [cited 2022 Jul 17];14(3):e0213626. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0213626

64. Nguyen LD, Lin D, Lin Z, Cao J. Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation. Proceedings - IEEE International Symposium on Circuits and Systems. 2018 Apr 26;2018-May.

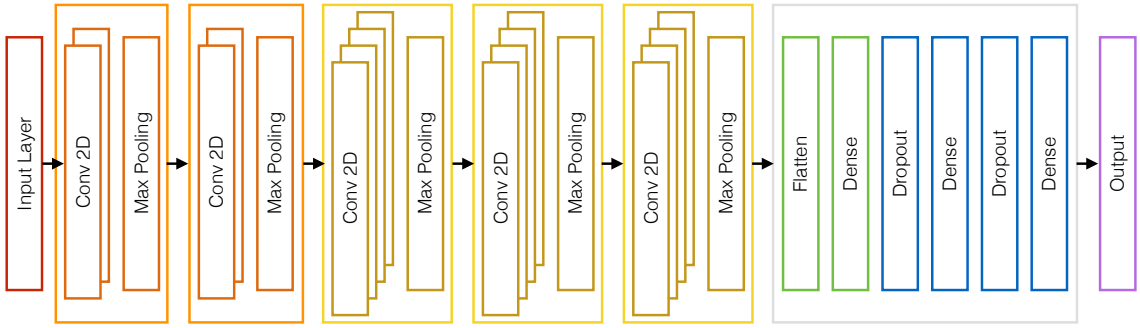
65. William W, Ware A, Basaza-Ejiri AH, Obungoloch J. A pap-smear analysis tool (PAT) for detection of cervical cancer from pap-smear images. Biomed Eng Online . 2019 Feb 12 [cited 2022 Jul 17];18(1). Available from: /pmc/articles/PMC6373062/

66. Boland M v., Murphy RF. A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells. Bioinformatics . 2001 [cited 2022 Jul 17];17(12):1213–23. Available from: https://pubmed.ncbi.nlm.nih.gov/11751230/

Supplemental Materials



Supplemental Figure 1. Correlation matrix for all morphological and complexity metrics. Morphological (green) and complexity (orange) metrics.



Supplemental Figure 2. Deep learning architecture, the Keras API VGG-19 model pretrained on ImageNet with a custom top.