# Predicting cellular capacity for growth after starvation from phase contrast microscopy images

Spencer Cesar

Department of Microbiology and Immunology
Stanford University
scesar@stanford.edu

June 10, 2020

## 1 Abstract

With prolonged starvation there are three different classes of genetically identical *E. coli* cells cells that emerge that are defined by their different responses to fresh nutrients. There is a population that resumes growth immediately, a population that resumes growth with some delay, and a final population that never appears to resume growth. I used a densely connected convolutional network (DenseNet) to classify phase microscopy images of cells coming from stationary phase (but before they had been allowed to resume growth) into the three different subpopulations based on how they would grow when exposed to nutrients. I was able to get relatively high performance at predicting the growing and nongrowing subpoulations, however the model performed much worse on the delayed fraction of cells. This is likely because his subpopulation is much rarer in the data set and also physiologically represents a space between the two other classes that could be harder to learn to classify especially with the limited training examples I had.

## 2 Introduction

Many bacteria live in complex natural environments that are often stressful and nutrient limited [10]. Within many environments such as the mammalian gut, leaf litter in soil, and whale falls in the ocean, food is provided only periodically, and hence bacteria encounter cycles of feast and famine. Thus, the transition from starvation to rapid growth can serve as a selective pressure for evolution [7, 9].

When coming from starvation conditions, *E. coli* (and many other bacterial species) have phenotypic heterogeneity where even though the population is genetically identical and all cells have seen the exact same culture conditions not all cells behave similarly. When exposed to fresh nutrients some cells immediately resume growth, some never resume growth (potentially because they have died) and some cells resume growth only after some delay which varies among cells, potentially being as short as a few minutes or as long as several hours.

I have built a convolutional neural network that will classify cells coming from starvation conditions into these three different subpopulations of *E. coli* cells based on their appearance in microscopy. The network takes cropped grayscale phase microscopy images of single cells and predicts what growth characteristic it will have when exposed to nutrients. This is interesting because currently there is no method that can predict how a cell will react to being exposed to nutrients after starvation by simply looking at its phase contrast signal, which is a relatively simple metric that doesn't require genetically altering the cells to produce a fluorescent reporter or staining them with any dye and also leaves them alive for further experimentation.

## 3 Related work

A significant amount of work has been done to extract biological information from different types of microscopy images. Classification [6,8], segmentation [2,3,12], tracking [1,14], and refocusing [15–17] problems have been tackled using deep learning models, particularly convolutional neural networks (CNNs). CNNs typically take images as inputs and learn hierarchical feature representations. Over recent years, many different neural network architectures, such as ResNet [4], Inception [13], DenseNet [5], and a Capsule Network [11], and different training techniques, such as Dropout and Batch Normalization have been developed.

Previously the only ways to differentiate between these different types of cells was to let them grow and observe how they behave. This method represents a large step forward as the cells can be classified based on appearance rather than needing to explicitly measure growth, this also allows for sorting and experimentation on the populations in starvation conditions which previously would have been impossible.

## 4 Dataset

To generate the dataset I took time-lapse phase microscopy images (which are 2048x2048 and 16-bit depth) of wild-type *E. coli* cells coming from prolonged starvation conditions. Cells were imaged every 2 minutes for 2-4 hours while growing on an agarose pad made of LB media in an environmental chamber that sets the temperature to 37° C which is ideal for *E. coli* growth. Next the images were segmented using an algorithm previously developed by our lab and various parameters were extracted from the segmented data. To determine which class a particular cell belonged to, I measured its single-cell growth rate. If a cell immediately resumed growth (median growth rate of .004 or greater in the first 20 minutes) it was classed as growing, if it resumed growth some time after that it was classed as delayed growth, and if the cell had an average growth rate of 0 for the duration of the imaging experiment it was classified as nongrowing (Figure 1).

Next, I took those labels and cell locations and went back to the first image in the tif stack of that specific position and experiment and cropped out a 64x64 image centered around the cell to get an image of the cell right as it was put onto the microscope. Figure 2 contains representative example images of each classification. Generally growing cells are the largest and nongrowing cells are the smallest, there is also a prominent bright spot in many nongrowing cells and a less prominent bright spot in delayed cells. The data used in this analysis is from 9 individual imaging experiments each of which contained about 100 different positions with 60-120 images in each stack.

One difficulty of this project is that delayed cells are relatively rare when compared to the two other classes which could lead to problems with classification. Another difficulty is the fact that currently it is not possible to collect more data, so I have a smaller amount of samples than is ideal. Currently there are 788 unique images of the delayed growth subclass of cells, 2333 unique images of the immediate growth class of cells, and 3225 unique images of the no growth class of cells for a total of 6346 unique examples. To augment the data for each sample I generated a mirrored image and 90, 180 and 270 degrees rotated images of both the original an mirrored images, this resulted in 8x as many samples. Next, I hand curated the data removing cell debris, aberrations in the pad, and cells that were out of focus. This is important as the nongrowing classification had all the non-cell objects sorted into it which would alter the learning process. This leads to a 0.13/0.37/0.50 split between the delayed, growing and non-growing subclasses. I used a 80-10-10 split between training, development, and test. Another difficulty is that while there are clearly different classes of cells based on behavior, it is difficult to know what boundaries are appropriate to use to differentiate between these cell types and my definition may not match up perfectly with the actual physiological boundaries between the subclass which would limit maximum performance if overfitting is correctly controlled for.

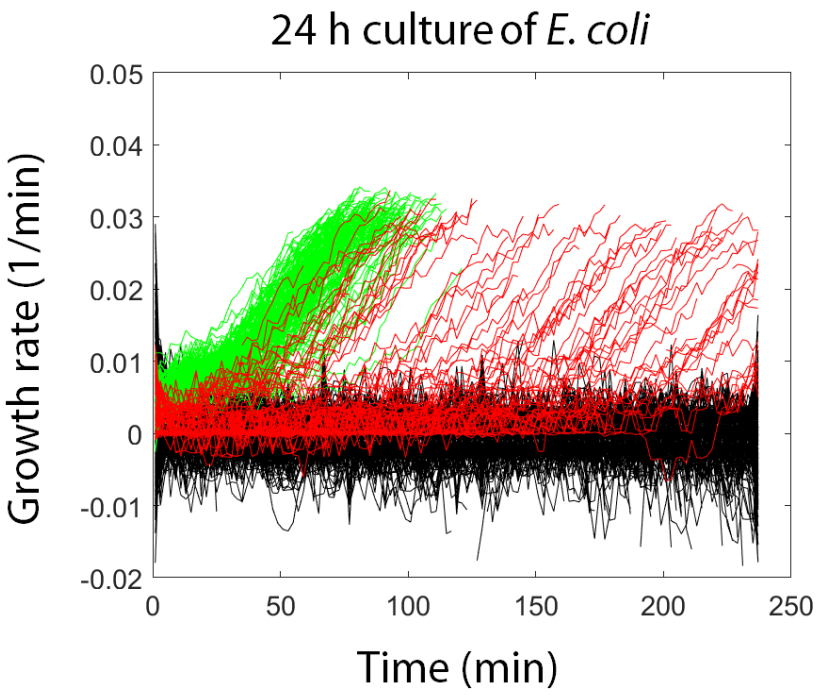


Figure 1: Example of growth rates extracted from single cells and what classification they lead to (green = growing, red = delayed, black = nongrowing).

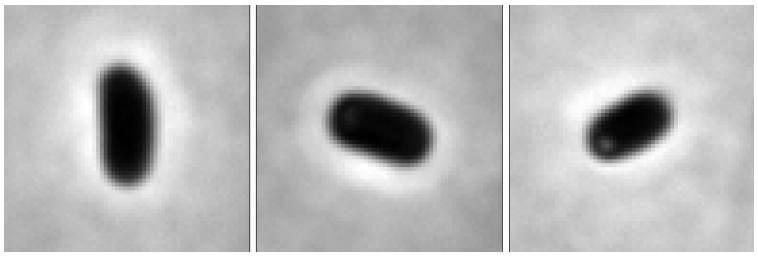

Figure 2: Representative examples of (from left to right) a growing cell, a delayed growth cell, and a nongrowing cell.

## 5 Methods

I implemented the DenseNet-121 architecture using the tensorflow package and pre-built tensorflow architecture for the network. The network had a initial convolution layer, then a pooling layer, then 7 layers of alternating dense blocks and transition layers, and finally a classification layer. The dense blocks are fully connected layers. The full description of the model and architecture can be found in Huang *et al* [5]. I also added l2 regularization to each layer of the model that had the capacity to combat the rapid overfitting that was observed without it.

# 6 Results and Analysis

After 230 epochs the current model reached nearly 100% accuracy on the training dataset. At this point, the model had an accuracy of 53% percent on the dev set (Figure 3). The highest performance on the dev set was at after 40 epochs (87% accuracy) and it was what was used for further evaluation. Seeing the large disparity between training and dev performance supports that the model is currently overfitting the training set. To combat this I tried to increase the L2 hyperparameter. Increasing L2 didnt fix the problem. It did was slow down the rate of learning and lower the max performance of the model on the training data but the performance on the dev set was not significantly affected by different L2 values (0.01, 0.001, and 0.0001 were all tested). Training the model for 200 epochs takes approximately 12 hours.

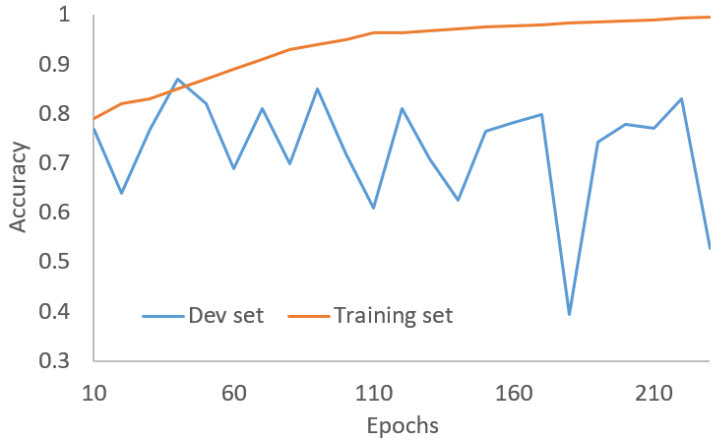


Figure 3: Accuracy over epochs of training for the previous model train (blue) and dev (orange) sets.)

```
4776/4776 - 11s - loss: 0.5804 - accuracy: 0.8237

Test accuracy: 0.82370186
              precision    recall  f1-score   support

 non-growing       0.87      0.98      0.92      2376
     delayed       0.39      0.45      0.41       592
     growing       0.95      0.75      0.83      1808

    accuracy                           0.82      4776
   macro avg       0.73      0.72      0.72      4776
weighted avg       0.84      0.82      0.83      4776
```

Figure 4: Performance on the test set with class specific performance measurments.

Figure 4 demonstrates the problem that arises due to the dataset imbalance where the delayed growth class only makes up 13 percent of the overall sample data. while the performance for the two more prevalent classes is relatively good (F1 of 0.92 for nongrowing and 0.83 for growing) the score for the delayed growth population is only 0.41 which is only slightly better than random when sorting between three classes.

The class activation maps helped explain what the network was doing. For a correctly identified growing and nongrowing cells (Figure 5) we see that the network is largely focusing on the perimeter of the cell, which is could support that it is using cell size/shape as a metric with an intense focus at the cell poles which could be the network looking for the characteristic bright spot that is present in the delayed and

Figure 5: Class activation maps for one growing, delayed, and nongrowing cell respectively.

nongrowing population but absent in the growing population. One flaw that the map for the delayed population demonstrates could be present is that the network could be learning to recognize the background which would indicate which experiment a specific cell was present. Certain experiments had higher levels of specific classes of cells so if the network is learning to recognize different levels of illumination that are unavoidable between different microscopy runs that is a problem.

Activation maps also demonstrated that the vast majority of the architecture was being unused. When looking at activation maps throughout the different layers it quickly became apparent that all layers in the dense blocks after the first 2 layers of the second block never showed anything in the class activation maps. This means that almost 75% of the network was not being utilized. This is not entirely unexpected given that my images are small, simple, and only utilize one channel.

# 7 Conclusions and Future Work

The network was able to differentiate between growing and nongrowing cells very well but it had problems with the delayed population of cells. Overfitting is a problem with the current architecture as the dev set performance drops very soon after training begins. The class activation maps are encouraging for the growing and nongrowing classes as the model seems to be focusing on the things that a human would focus on to differentiate between cells and is able to do so with high performance.

In the future, collecting more data could significantly improve class specific and overall performance. It could also be useful to set all pixels that are not part of the cell to 0 to prevent the network from learning to identify the different backgrounds from different experiments which have different levels of the three populations and using that information to predict which class cells belong to rather than cellular appearance. Additionally, the class activation maps seem to indicate that the vast majority of the current architecture isn't being utilized. A smaller model would run faster and could also help prevent the over fitting that occurs with longer training times in the current model. I am also interested in tracking even more metrics. The samples were collected from 10 different experiments that all had different experimental conditions (the biggest difference being in the age of the cells) that could lead to variations in cellular physiology and I am interested to see if there are measurable differences in the models ability to correctly predict cells from specific experiments/time points. This could be an indicator that cells within a class change throughout time in starvation conditions.

# References

[1] F. Buggenthin, F. Buettner, P. S. Hoppe, M. Endele, M. Kroiss, M. Strasser, M. Schwarzfischer, D. Loeffler, K. D. Kokkaliaris, O. Hilsenbeck, T. Schroeder, F. J. Theis, and C. Marr. Prospective identification of hematopoietic lineage choice by deep learning. *Nature Methods*, 14(4):403–406, 2017.

[2] J. C. Caicedo, A. Goodman, K. W. Karhohs, B. A. Cimini, J. Ackerman, M. Haghighi, C. Heng, T. Becker, M. Doan, C. McQuin, M. Rohban, S. Singh, and A. E. Carpenter. Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. *Nature Methods*, 16(12):1247–1253, 2019.

[3] T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald, A. Dovzhenko, O. Tietz, C. Dal Bosco, S. Walsh, D. Saltukoglu, T. L. Tay, M. Prinz, K. Palme, M. Simons, I. Diester, T. Brox, and O. Ronneberger. U-Net: deep learning for cell counting, detection, and morphometry. *Nature Methods*, 16(1):67–70, 2019.

[4] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[5] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017.

[6] M. S. Manak, J. S. Varsanik, B. J. Hogan, M. J. Whitfield, W. R. Su, N. Joshi, N. Steinke, A. Min, D. Berger, R. J. Saphirstein, G. Dixit, T. Meyyappan, H.-M. Chu, K. B. Knopf, D. M. Albala, G. R. Sant, and A. C. Chander. Live-cell phenotypic-biomarker microfluidic assay for the risk stratification of cancer patients via machine learning. *Nature Biomedical Engineering*, 2(10):761–772, 2018.

[7] R. Monds, T. Lee, A. Colavin, T. Ursell, S. Quan, T. Cooper, and K. Huang. Systematic perturbation of cytoskeletal function reveals a linear scaling relationship between cell geometry and fitness. *Cell Reports*, 9(4):1528 – 1537, 2014.

[8] W. Ouyang, C. F. Winsnes, M. Hjelmare, A. J. Cesnik, L. Åkesson, H. Xu, D. P. Sullivan, S. Dai, J. Lan, P. Jinmo, S. M. Galib, C. Henkel, K. Hwang, D. Poplavskiy, B. Tunguz, R. D. Wolfinger, Y. Gu, C. Li, J. Xie, D. Buslov, S. Fironov, A. Kiselev, D. Panchenko, X. Cao, R. Wei, Y. Wu, X. Zhu, K.-L. Tseng, Z. Gao, C. Ju, X. Yi, H. Zheng, C. Kappel, and E. Lundberg. Analysis of the Human Protein Atlas Image Classification competition. *Nature Methods*, 16(12):1254–1261, 2019.

[9] J. M. Peters, A. Colavin, H. Shi, T. L. Czarny, M. H. Larson, S. Wong, J. S. Hawkins, C. H. Lu, B.-M. Koo, E. Marta, A. L. Shiver, E. H. Whitehead, J. S. Weissman, E. D. Brown, L. S. Qi, K. C. Huang, and C. A. Gross. A comprehensive, crispr-based functional analysis of essential genes in bacteria. *Cell*, 165(6):1493 – 1506, 2016.

[10] D. B. Roszak and R. R. Colwell. Survival strategies of bacteria in the natural environment. *Microbiology and Molecular Biology Reviews*, 51(3):365–379, 1987.

[11] S. Sabour, N. Frosst, and G. E. Hinton. Dynamic routing between capsules. *CoRR*, abs/1710.09829, 2017.

[12] A. Samacoits, R. Chouaib, A. Safieddine, A.-M. Traboulsi, W. Ouyang, C. Zimmer, M. Peter, E. Bertrand, T. Walter, and F. Mueller. A computational framework to study sub-cellular RNA localization. *Nature Communications*, 9(1):4584, 2018.

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016.

[14] V. Ulman, M. Maška, K. E. G. Magnusson, O. Ronneberger, C. Haubold, N. Harder, P. Matula, P. Matula, D. Svoboda, M. Radojevic, I. Smal, K. Rohr, J. Jaldén, H. M. Blau, O. Dzyubachyk, B. Lelieveldt, P. Xiao, Y. Li, S.-Y. Cho, A. C. Dufour, J.-C. Olivo-Marin, C. C. Reyes-Aldasoro, J. A. Solis-Lemus, R. Bensch, T. Brox, J. Stegmaier, R. Mikut, S. Wolf, F. A. Hamprecht, T. Esteves, P. Quelhas, Ö. Demirel, L. Malmström, F. Jug, P. Tomancak, E. Meijering, A. Muñoz-Barrutia, M. Kozubek, and C. Ortiz-de Solorzano. An objective comparison of cell-tracking algorithms. *Nature Methods*, 14(12):1141–1152, 2017.

[15] H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Günaydın, L. A. Bentolila, C. Kural, and A. Ozcan. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nature Methods*, 16(1):103–110, 2019.

[16] M. Weigert, U. Schmidt, T. Boothe, A. Müller, A. Dibrov, A. Jain, B. Wilhelm, D. Schmidt, C. Broaddus, S. Culley, M. Rocha-Martins, F. Segovia-Miranda, C. Norden, R. Henriques, M. Zerial, M. Solimena, J. Rink, P. Tomancak, L. Royer, F. Jug, and E. W. Myers. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature Methods*, 15(12):1090–1097, 2018.

[17] Y. Wu, Y. Rivenson, H. Wang, Y. Luo, E. Ben-David, L. A. Bentolila, C. Pritz, and A. Ozcan. Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning. *Nature Methods*, 16(12):1323–1331, 2019.