Published on Dec 01, 2023 by Arcadia Science **DOI**: 10.57844/arcadia-c021-e384

Automated classification of time-course imaging data applied to nematode embryogenesis

Machine learning is a powerful tool for classifying images in a time series, such as the developmental stages of embryos. We built a classifier using only bright-field microscopy images to infer nematode embryonic stages at high throughput.

Contributors (A-Z)

Prachee Avasthi, Feridun Mert Celebi, Keith Cheveralls, Seemay Chou, Amro Hamdoun, Megan L. Hochstrasser, Ilya Kolb, David Q. Matus, Shalin Mehta, David G. Mets, Harper Wood

Version 4 · Mar 31, 2025

Purpose

We're broadly interested in extracting biological function from label-free, high-throughput imaging data. As a first pass, we tested the effectiveness of a deep-learning framework that incorporates temporal information in classifying the developmental stage of the well-studied nematode, *Caenorhabditis elegans.* We trained a classifier that you can use to identify nematode embryo stages from time-course datasets captured using bright-field microscopy. We hope this tool will be

immediately useful to interrogate embryonic development, reproductive success, or developmental outcomes following perturbations in *C. elegans* or other free-living nematode species. More broadly, you can adapt our approach to any category of classifiable microscopy time-course data. To this end, we provide a PyTorch-based pipeline for training and evaluating your own models.

The tool lets you go from imaging nematode embryos to classifying developmental stages and quantifying the frequency of successful versus unsuccessful developmental outcomes. It's about 80% accurate in calling the correct stage. We're not pursuing this project further but welcome your input and encourage others to incorporate user feedback to improve the functionality of the classifier if it's useful to you.

- Our **code** in Python is available in this GitHub repository.
- The **data** we used in our training, validation, and experimentation are on Zenodo.

We've put this effort on ice! ☒

#StrategicMisalignment

We considered building on this initial effort by predicting likelihood of embryo hatching, but we'd mostly intended the project as an internal exercise in how to build and train a classifier on a visual dataset. We ultimately decided that continuing this particular project wasn't a high priority relative to other work.

Learn more about the Icebox and the different reasons we ice projects.

Background

For most organisms, the effort and expense of genetic or antibody-based labeling for the purpose of imaging is very high, requiring dedicated team effort and resources. We

want to develop tools that we can readily apply to many organisms, allowing new understanding of evolutionary solutions to biological problems. In line with this overarching goal, we set out to leverage the information in label-free images for phenotyping in a scalable, automated fashion. More broadly, we'd like to understand the extent to which we can use label-free imaging across as large a swath of the tree of life as possible to extract phenotypic information and map traits wherever we find novelty.

Combining deep learning with high-throughput live imaging has the potential for broad impacts on many fields of biology, scaling from cells to organisms. For example, deep learning approaches to cell type identification **[1][2]** and cell health **[3]** have the potential to be transformative. Applying these methods to label-free data **[4][5][6]** decreases experimental cost and increases our ability to explore organismal diversity.

Developmental biology is ripe for the application of deep learning approaches to facilitate discovery and unlock translational potential **[7]**. The study of developmental biology has provided fundamental insights into multicellular life at the intersection of genetics, molecular, cellular, and evolutionary biology. Seeing where development goes awry allows us to understand the molecular underpinnings of disease, from developmental disorders leading to birth defects **[8][9]** to the origins of cancer **[10]**. Not surprisingly, there are concerted efforts at improving the outcomes of *in vitro* fertilization by applying deep learning strategies to human embryo health and viability **[11][12]**.

During embryogenesis, multicellular organisms pass through discrete developmental stages, including fertilization, cleavage, morphogenesis, and organogenesis, ultimately hatching into their environment. Animal development is characterized by sets of shared and species-specific features. For example, following fertilization, most animal embryos undergo a series of rapid cell divisions. At some point during this cleavage period, cells undergo a suite of morphogenetic changes as embryo patterning results in tissue-layer organization through the process of gastrulation. While embryos from many different organisms may share similar-looking cleavage stages, within specific lineages there are often unique morphologies characteristic of distinct taxonomic groups — animal embryos that look similar at cleavage stages might look very different during gastrulation. These species-specific differences only compound as development continues. Thus, there is a need for automated tools to classify key embryonic stages to unlock high-throughput approaches to developmental biology.

Finally, to fully understand the development of a particular organism, we need to be both descriptive and mechanistic. The most common way of accomplishing this is by perturbing the system, from traditional mutagenesis to drug screening. If we can devise approaches that take advantage of inherent properties, such as the data we get from label-free light microscopy methods (e.g., bright-field, DIC, phase contrast, etc), we can maximize our ability to perform these experiments at scale, across the tree of life, as we don't have to invest in bespoke genetic labeling tools for each new research organism.

The strategy

As a first step toward a longer-term goal of high-throughput image-based phenotyping across species, we decided to develop an image analysis pipeline for training a neural network to classify developmental stages with high accuracy and minimal human intervention from bright-field movies.

We selected the well-studied nematode, *Caenorhabditis elegans*, as a test case for automated phenotyping because it has a well-defined embryonic lineage and easily observable morphological stages, undergoes rapid embryonic development to a free-living larva, and has a large scientific community that leverages these many strengths for biological discovery. Despite many differences in early embryo patterning between nematode species, key developmental stages appear conserved **[13][14]**, so we also wanted to explore whether a classifier trained on *C. elegans* development would work out-of-the-box on related nematode species, unlocking future evolutionary comparative studies.

The problem

To fully leverage high-throughput experimental approaches that involve imaging, we need automated image processing and analysis workflows. High-throughput experiments involving time courses often consist of large (100+ GB) datasets that require lengthy data curation and annotation before analyses can even begin. We set out to establish a method for classifying embryonic stages from bright-field image data; a modality that does not require the use of species-specific labeling tools.

Previous attempts to generate a nematode classifier required technological innovations in microfluidic approaches to isolate individual embryos and relied on reporter transgenes to properly orient the embryo **[15]**. In contrast, we wanted to create a classifier that performs robustly irrespective of embryo orientation and solely using bright-field microscopy, to allow for comparative studies in organisms lacking genetic tools.

Our solution

As a proof-of-principle, we built an automated, high-throughput, experimental, and computational workflow to image and classify the embryonic stages of *C. elegans*.

Our workflow includes (1) optimizing high-throughput embryo collection and imaging, (2) embryo segmentation, and (3) classification of known stages of nematode development as well as the detection of unfertilized oocytes and embryonic lethality. In constructing this pipeline, we've built a trained classifier to recognize label-free bright-field images of *C. elegans* embryonic stages, based on the original descriptions by John Sulston (1942–2018), who generated the first embryonic lineage map of a multicellular organism. We trained a model based on manual ground truth annotations using the ResNet-18 neural network architecture **[16]**. Our classifier achieves approximately 80% accuracy, accounting for class imbalances, for classifying embryo developmental stages, independent of embryo orientation. To test our nematode classifier and to extend the utility of this tool, we used it to quantify the embryonic lethality associated with induced environmental stress from heat shock and osmotic stress and tested its ability to correctly classify embryonic stages of related nematode species.

We anticipate that this pipeline will be useful in collecting population-level details related to reproductive success or embryonic lethality in phenotyping following perturbation (e.g., RNA interference or traditional mutagenesis screens). More generally, we're excited by the potential of taking this approach to classify other kinds of time-course data. We hope you'll be able to apply our workflow for your own time-course data and would love to hear how it goes, so please drop us a comment if you try it!

The resource

Building a classifier for nematode embryo stages

We created a classifier to facilitate the characterization of *C. elegans* embryonic phenotypes in high-throughput time-course imaging data. In this pub, we summarize how we trained our model. We also describe the CLI that you can use to adapt the model to your imaging data acquired with different contrast, resolution, or magnification (see pipeline documentation).

This first section provides a brief overview of the workflow (Figure 1), from collecting the data to using the computational pipeline to classify and extract labeled time-course data for downstream analyses. To see the classifier in action, jump to "Using the classifier for high-throughput studies of nematode development."

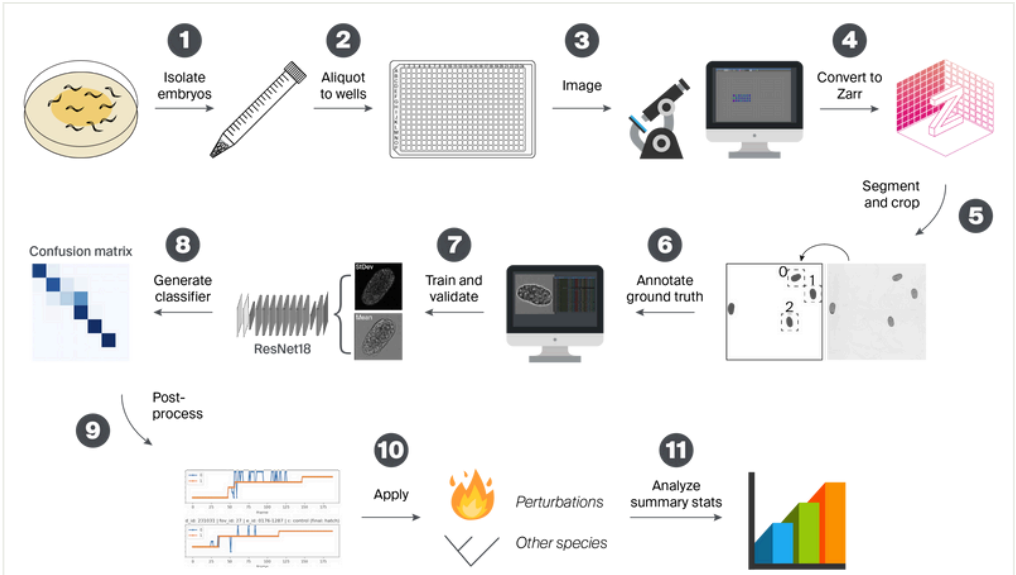


Figure 1

Overview of workflow to train and use our nematode classifier.

(1) We collect embryos by hypochlorite treatment and dispense into 384-well glass-bottom plates (2) for imaging (3). Post-acquisition, we convert files into the ome-zarr format (4) and then segment and crop individual FOVs (5) in preparation for ground truth annotation (6). We then train and validate (7) input data from movies that have ground truth annotation using a ResNet-18 architecture to generate a trained classifier (8). A post-processing filtration step removes transient errors (9) in state calls. We can then process new experimental data and apply the classifier to that processed data (10) to generate summary statistics for high-throughput experiments (11).

Embryo isolation and applying “smart” microscopy to optimize data collection

We isolated embryos from gravid adults by hypocholorite treatment **[17]** and added them to a 384-well glass-bottom plate (Cellvis).

To reduce the collection of empty fields of view (FOVs), we used the “JOBS” function in Nikon NIS Elements software (version 54203) to perform threshold-based object detection (Nikon Elements script available here) in a first round of imaging, tiling over

each well (Figure 2, B). FOVs that passed a minimum object detection criteria (usually > 3 embryos detected) moved to a second round of imaging (Figure 2, C and Figure 3). We then imaged these embryo-containing FOVs every five minutes for a minimum of 14–16 hours, a length of time that should allow for wild-type *C. elegans* to hatch into the L1 larval stage **[18]**.

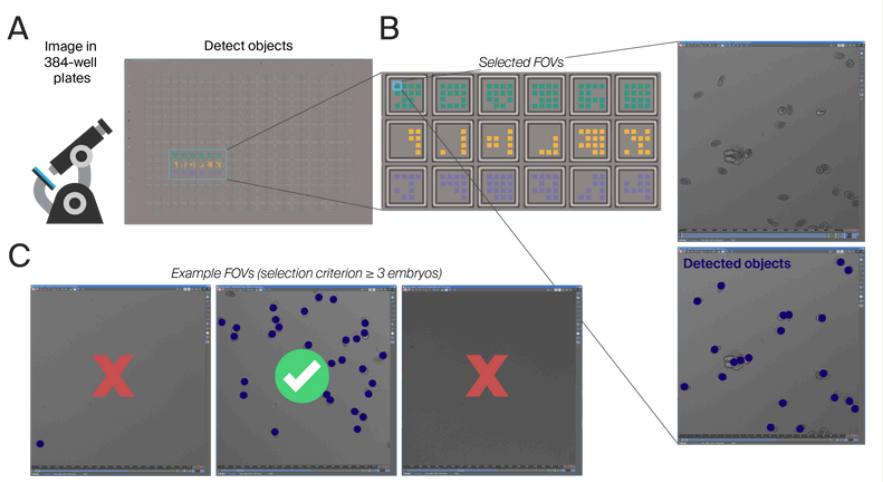


Figure 2

Smart imaging to collect high-throughput nematode embryogenesis data.

(A) We perform imaging in 384-well plates in a subset of wells. In this example, the colored wells denote potential different treatments in a given experiment.

(B) We perform a round of object detection to limit data collection to FOVs that have a minimum number of embryos. Example FOV is shown as raw image (top) and after object detection (bottom), where the dark blue dot indicates a detected object.

(C) Example FOVs that either fail (✖) or pass (✓) object detection.

Image processing

After image acquisition, we preprocess the raw FOVs to crop around each embryo to obtain images of uniform size that are centered on a single embryo. This preprocessing step significantly reduces the complexity of subsequent annotation and analysis — this two-fold approach transforms the problem of object detection and classification into the problem of image classification. We segmented the embryos from the temporal fluctuations in intensity by computing the standard deviation of the raw bright-field movies across the time dimension, then using Otsu thresholding to generate a background mask (Figure 3, A). We then filtered the foreground regions in the background mask using morphological criteria to exclude regions that did not correspond to a single isolated embryo. Finally, we obtained movies of single embryos by cropping square bounding boxes of a size equal to the length of the embryos around each foreground region (Figure 3, B).

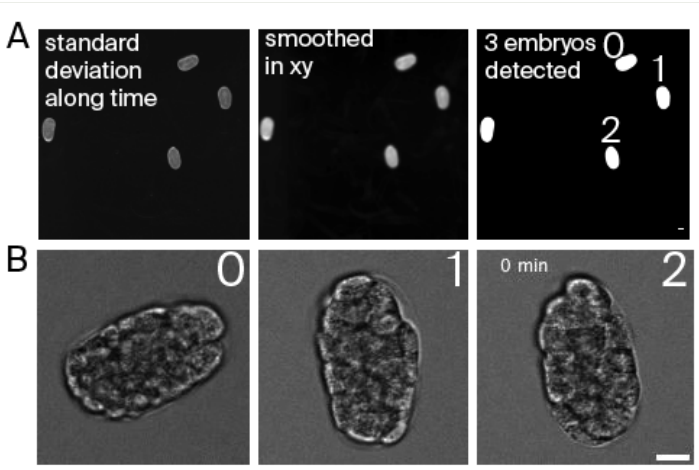


Figure 3

Image processing workflow to generate single cropped embryo time-lapses.

(A) Example segmentation workflow for an FOV. We segment data based on standard deviation over time (left) and perform smoothening in xy (middle), which results in detection of a subset of embryos in a given FOV. Detected embryos are given a unique ID. As an example, we've numbered starting from 0 in this figure.

(B) Representative, cropped time-lapses from the FOV in the example shown in (A). Scale bar, in this and all subsequent images, is 10 μm.

Key frame annotation for ground truth

To build a classifier for nematode embryogenesis, we first had to decide on a core set of developmental stages that would be useful to encode as ground truth. *C. elegans* embryogenesis is highly stereotyped, with a defined cell lineage and rapid development, as embryos hatch in ~12–14 hours into a motile larval stage (L1). For our classifier, we selected key developmental stages based on the work of John Sulston

and colleagues, whose groundbreaking efforts led to the first cell lineage map of any animal embryo **[18]**.

While a rare occurrence in our wild-type imaging, we did observe instances of unfertilized embryos, likely stemming from a result of the hypochlorite bleaching treatment or from older hermaphrodites that had exhausted their supply of sperm **[19]**. In experimental manipulations or in experiments involving aging, we expect that recording the frequency of *unfertilized* embryos would be useful, so we annotated images of unfertilized embryos (Figure 4, A and B.0). Next, we binned all of the early cell division events prior to major morphogenetic movements into a *proliferation* stage, which would also include all the events associated with gastrulation (Figure 4, A and B.1). The first major morphogenetic changes in the embryo are observable by bright-field microscopy imaging restricted to a single z-plane, and happen ~six hours into development, when the embryo takes on a characteristic *bean* morphology (Figure 4, A and B.2). The next characteristic stage in nematode development is the *comma* phase (Figure 4, A and B.3), which Sulston et al. precisely defined as “the moment at which the ventral surface of the tail lies perpendicular to the long axis of the egg” **[18]**. In our movies, this stage only represented a 10-minute imaging window (two frames, as our time interval was five minutes). Shortly after the *comma* stage, the embryo begins to move and progresses through three stages, usually defined as one-, two-, and three-fold. We binned these stages together as the *fold* stage (Figure 4, A and B.4). Finally, the larva hatches into its environment, escaping the eggshell, which for purposes of ground truth training we annotated as *hatch* either the moment we saw the larvae escape or more commonly in our imaging data, the first frame without an embryo, though sometimes the eggshell is visible in the frame (Figure 4, A and B.5).

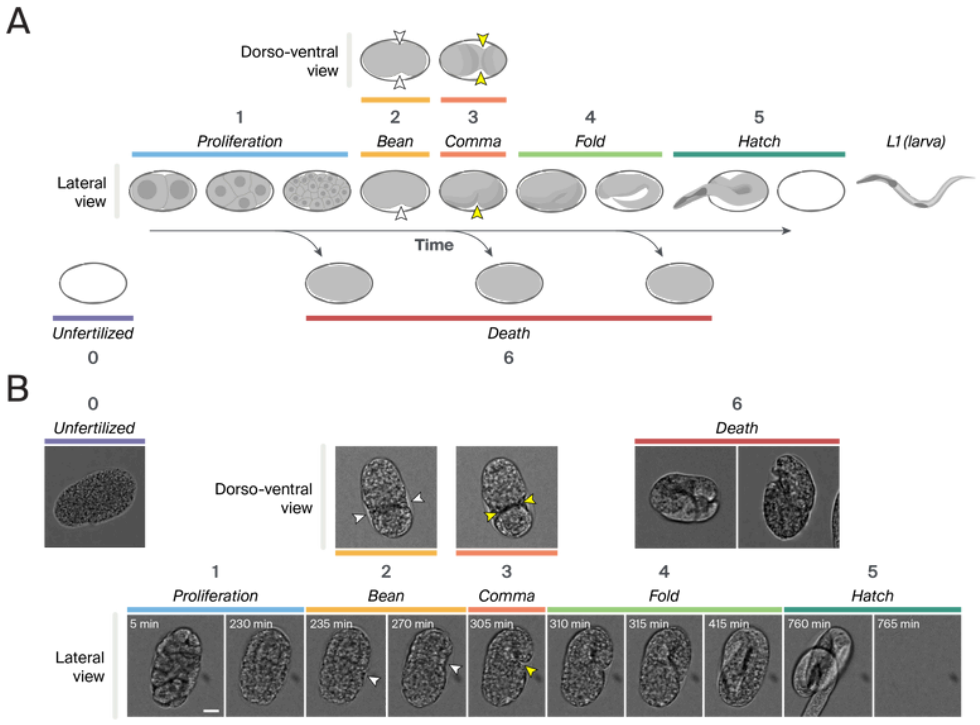


Figure 4

Schematic and representative images of the developmental stages we used to create a nematode classifier.

(A) Schematic and (B) images showing the seven classes we selected for annotation. We annotated images irrespective of orientation so our classifier could correctly identify stages in both lateral and dorso-ventral orientations. White arrows indicate invagination at the *bean* stage and yellow arrows indicate the orientation of the growing tail with respect to the rest of the embryo, as described by Sulston et al.

To add functionality to our classifier for downstream experiments, we wanted to annotate images of embryonic lethality or death (Figure 4, A and B.6). We looked through our original dataset, and not surprisingly, given the high fidelity of *C. elegans* embryogenesis **[20]**, we were only able to find two examples (out of 291) of embryos dying during imaging. In an attempt to generate more images of embryonic lethality, we heat-shocked wild-type L4 stage animals (the last developmental stage before becoming gravid adults) at 37 °C for one hour and collected embryos the following day. However, even in this dataset, we were only able to identify an additional two examples of “death.”

Rather than troubleshoot heat shock conditions, we decided to use a pharmacological perturbation strategy to induce embryonic lethality. A previous attempt to build a *C. elegans* embryo classifier used several perturbation strategies, including high salt **[15]**. We found that at 0.2 M NaCl, the concentration used by Atakan and colleagues, we still observed insufficient incidence of embryonic lethality. Given that Atkan and colleagues found high embryonic lethality in the context of a microfluidic chamber (in addition to 0.2 M NaCl), it's possible that this level of lethality (~30%) depended on other environmental factors in addition to the hyperosmotic stress. In other studies not utilizing a microfluidics chamber, researchers have used higher salt concentrations to induce hyperosmotic stress **[21]**. Thus, we performed an additional round of imaging using 0.5 M and 0.75 M NaCl. At 0.5 M NaCl, we noticed that many embryos were arrested during *fold* stages. At 0.75 M NaCl, we saw pronounced embryonic lethality. We therefore used images from this 0.75 M NaCl dataset as additional ground truth annotations for training a classifier to recognize *death*.

Machine learning using a ResNet-18 architecture to create a nematode classifier

We trained a ResNet-18 **[16]** convolutional neural network (CNN) architecture in PyTorch (Figure 5). We started with a pre-trained ResNet-18 and adapted the model to our task via transfer learning. We replaced the first convolutional layer to allow for multiple input channels. We pooled annotated movies of unperturbed, heat-shocked and osmotically perturbed embryos to train and evaluate a model that generalizes to diverse perturbations. In order to make the model invariant to orientation and small differences in the size of the embryo, we augmented the input images with transforms such as random rotations and random scaling.

We tested several different data transformations when selecting an optimally performing model, comparing model performance on raw data as input (Figure 5, A and B) to measures of temporal fluctuations, such as moving average over time and moving standard deviation over time (Figure 5, C and D). We eventually chose to use the moving standard deviation and the moving mean with a window size of five frames (Figure 5, C and D) as encoding temporal dynamics as input data improved stage classification accuracy for almost all stages as compared to raw data (Figure 5, B and D). The best-performing model classified most stages (*bean*, *fold*, *hatch*, and *death*) with >77% accuracy (Figure 5, D). Confusion resulted during classification of the

comma stage from *bean*-stage embryos, and, to a lesser extent, between *unfertilized* and *dead* embryos (Figure 5, D).

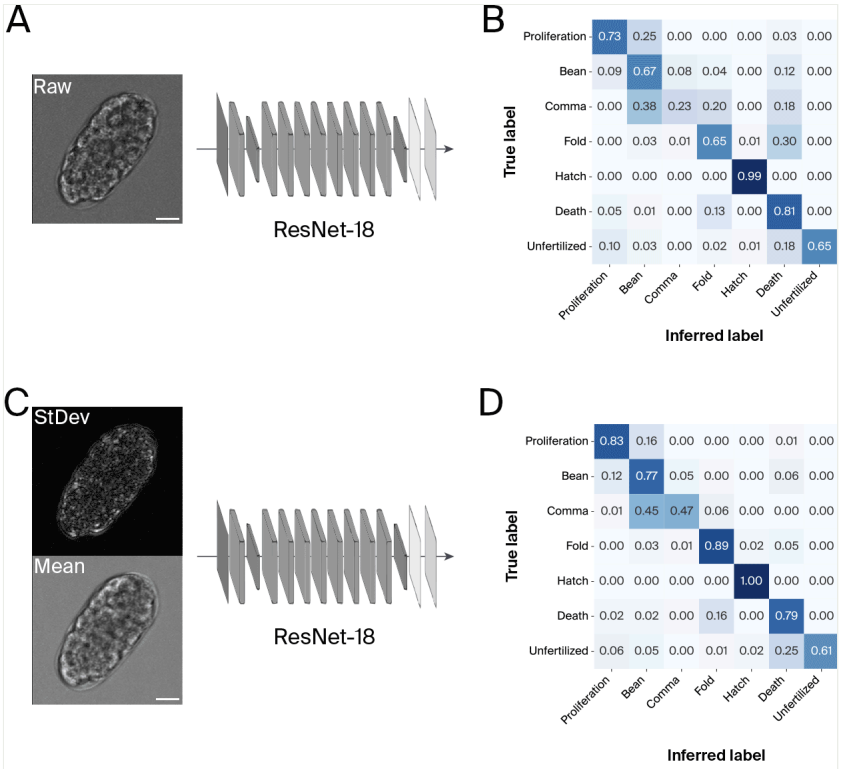


Figure 5

Encoding dynamics into a ResNet-18 CNN improves nematode classifier performance.

We initially trained a ResNet-18 CNN architecture using raw data as input into the model (A) resulting in a confusion matrix (B). Ground truth annotations are shown along the y-axis and classifications along the x-axis. After testing several different input data transformations, we found that using the moving standard deviation and moving mean with a window size of five frames (C) performed better than using raw data (A) as inputs into the neural network (see Materials and methods for additional details), as shown by a test confusion matrix (D) from annotations across experiments. The best-performing model (D) performs at high accuracy for most stages (> 77% for *proliferation*, *bean*, *fold*, *hatch*, and *death*) with poor performance for *comma* (47%) and detection of *unfertilized* embryos (61%).

Improving classifier performance with post-processing

Although our trained network classified developmental stages with reasonable accuracy (Figure 6, A and A′), we noticed that many of the errors in the classification of our test data occurred due to transient confusion between non-sequential stages (e.g., between *proliferation* and *fold*) or confusion between embryonic lethality (*death*) and *fold* stages (Figure 6, B and B′). To correct confusion between non-sequential stages, we first applied a median filter (using a window size of seven frames) to the classified stages to remove transient errors. Then, we eliminated developmentally impossible stage transitions (such as going backward in development or skipping stages). To eliminate confusion between embryonic lethality and the *fold* stage, we took into account the developmental outcome of the individual time-lapse — i.e., if an embryo hatched successfully at the end of the time-lapse, we eliminated any transitions prior to the *death* stage (Figure 6 A′ and B′). Overall, post-processing improves stage classification accuracy for *bean*, *comma*, *fold* and *death* (Figure 5, D versus Figure 6, C).

We were unable to perform post-processing on the stages between *proliferation* and *fold* (*comma* and *bean*), which represent a period of morphogenesis during *C. elegans* development **[22]**. Confusion between *comma* and *bean* is not surprising, as the *comma* stage occurs for ~10 minutes, corresponding to ~two frames in our time-lapse datasets. We used the precise definition of the *comma* stage established by Sulston et al. in our classification, but this stage is easily confused with the previous *bean* stage, even by a trained human annotator. Combining these two stages into a single *morphogenesis* stage, indicative of the cell movements and rearrangements that occur between *proliferation* and the *fold* stage **[22][18]**, would result in > 88% accuracy (e.g., correct *bean* ID = 81% + incorrect ID as *comma* = 7%; Figure 6, C). We expect that experimentally, it would be useful to broadly classify *bean* and *comma* together, as a means of quantifying phenotypic responses that might result in changes in some of the major tissue level rearrangements that occur during this phase of development, including dorsal intercalation and ventral enclosure **[22][23]**.

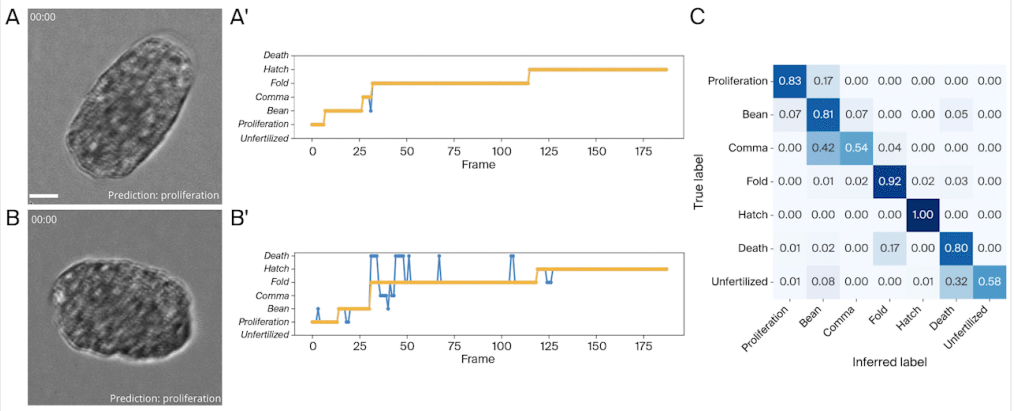


Figure 6

Post-processing to improve classification.

(A, B) Example embryo time-lapse movies with classifier annotations appended to the movie.

(A′, B′) Corresponding classifications per frame. Blue traces are the original classifications and orange traces are the post-processed classifications.

(C) Confusion matrix comparing post-processed annotations to ground truth annotations for embryos not used during training.

Our **code** in Python is available in this GitHub repository (DOI: 10.5281/zenodo.10247028).

Using the classifier for high-throughput studies of nematode development

Identifying developmental outcomes in high-throughput imaging experiments

In this section, we summarize the results of using our classifier to aid in the analysis of high-throughput, time-course imaging data. First, we examined the final state classifications from imaging wild-type embryos and embryos whose mothers experienced a brief 37 °C heat shock (Figure 7, A and B). These data supported our initial observations when we were annotating images for ground truth, as there were few (1%, n = 3/291 embryos) instances of embryonic lethality in wild-type embryos, and only a slight increase (8%, n = 11/137 embryos) in embryonic lethality in embryos following heat shock.

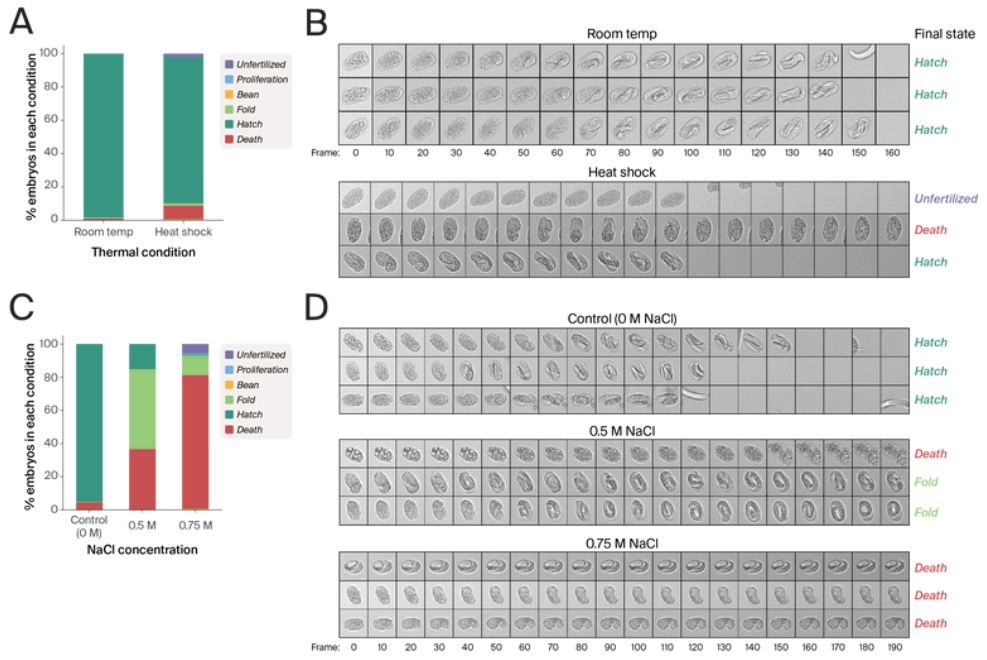


Figure 7

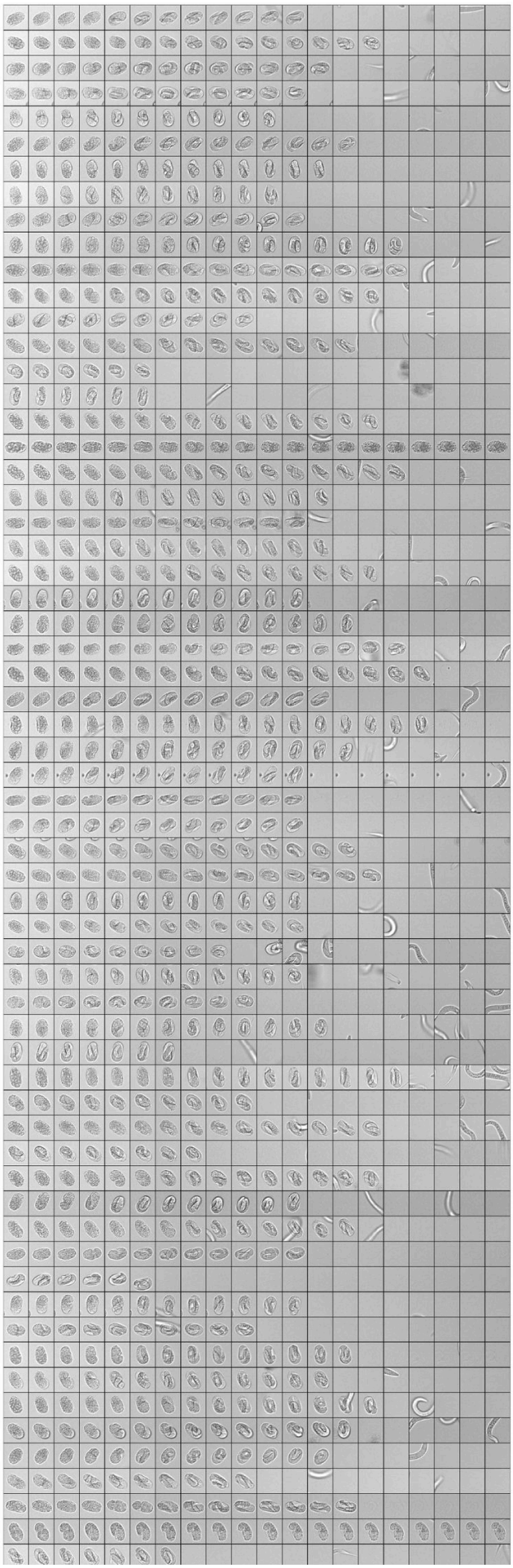
Classifier identifications of developmental outcomes from high-throughput imaging.

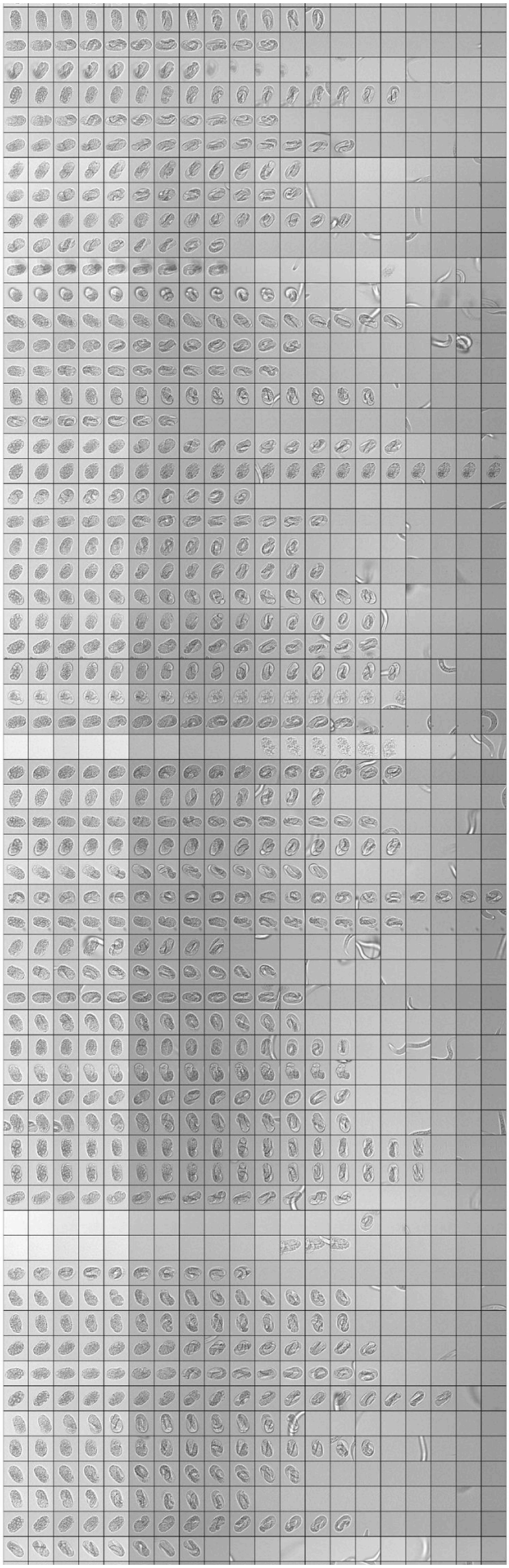
(A, C) Stacked bar graphs depicting the percentage of embryos per developmental stage for each experiment [control and heat shock (A) and osmotic shock experiment (C)]. Color key indicates final developmental state.

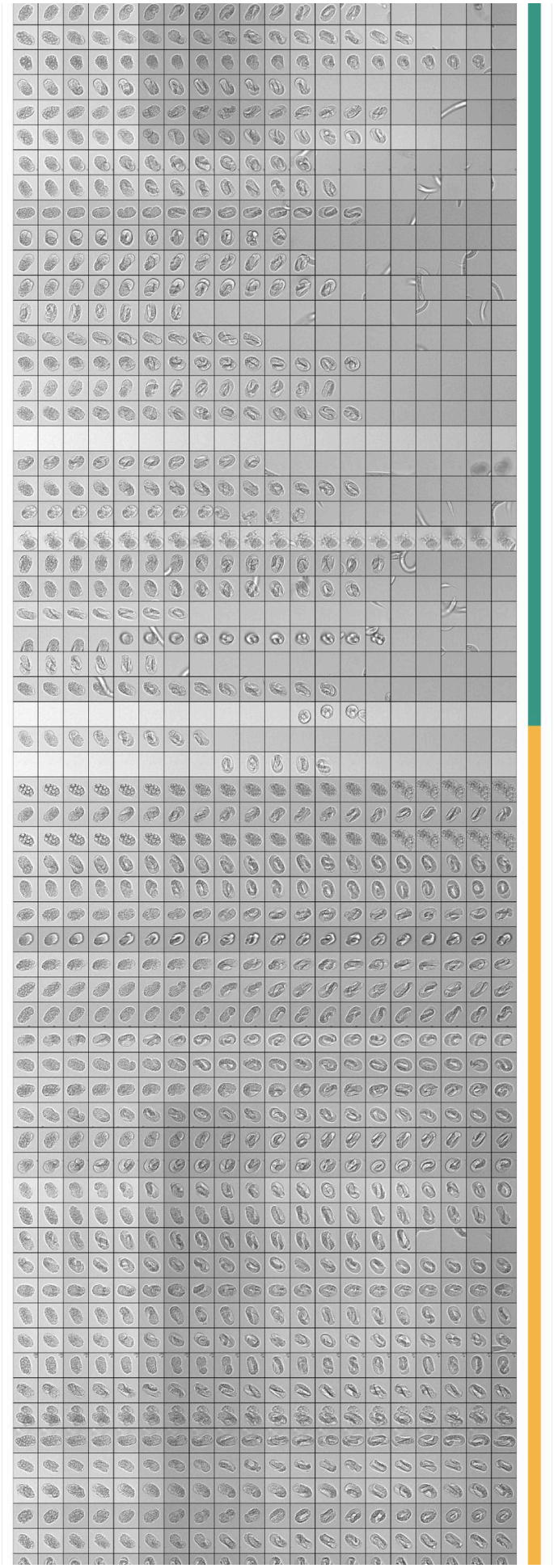
(B, D) Representative frames from individual time-lapses for each experimental condition, as indicated with final state classification shown at the end of each image series.

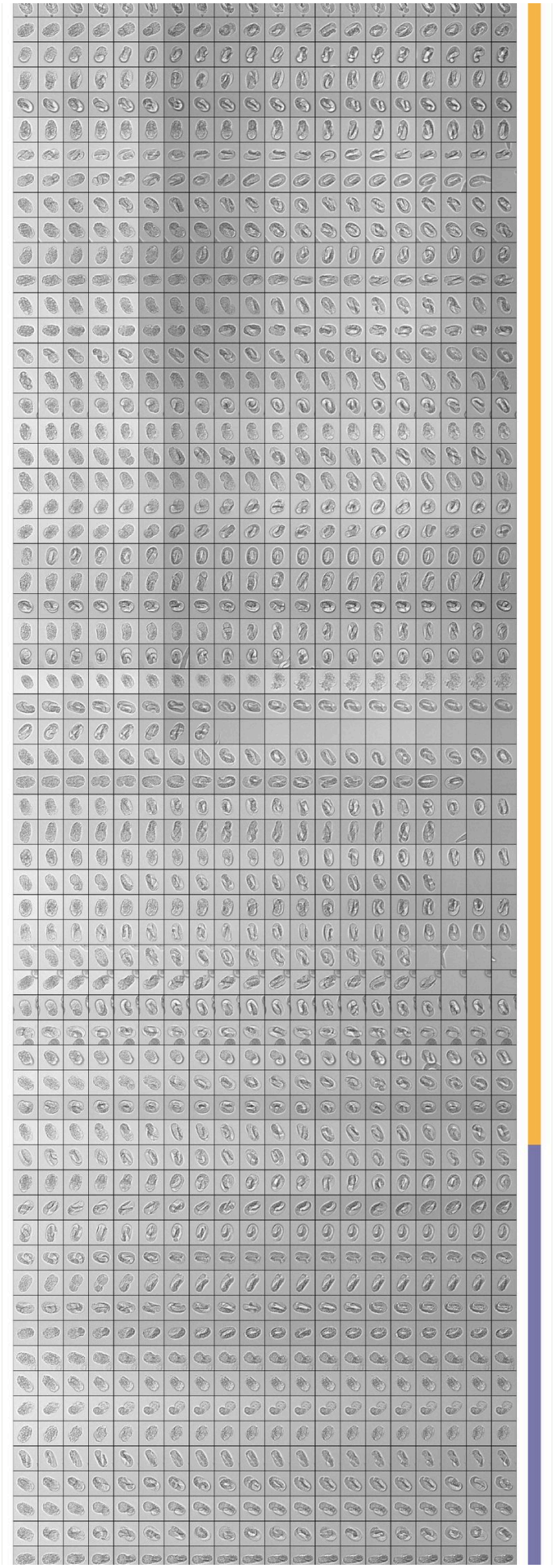
Next, we wanted to analyze the results of the osmotic stress dataset (Figure 7, C and D), which we performed to collect examples of embryonic lethality ("*death*") for our classifier, given the low occurrence of embryonic lethality in our wild-type and heat shock datasets (Figure 7, A). To interact with our data visually, we generated a filmstrip of every 10th frame for every other embryo in our datasets (Figure 8). We treated embryos with either 0 M (control), 0.5 M, or 0.75 M NaCl solution and allowed them to develop for 16 hours. While *C. elegans* is capable of adapting to high-salt environments **[24][21]**, embryos treated with high salt solutions without pre-adaptation results in embryonic lethality at varying penetrances **[21][15]**. We selected two concentrations, 0.5 M and 0.75 M NaCl, as these treatments robustly resulted in embryonic

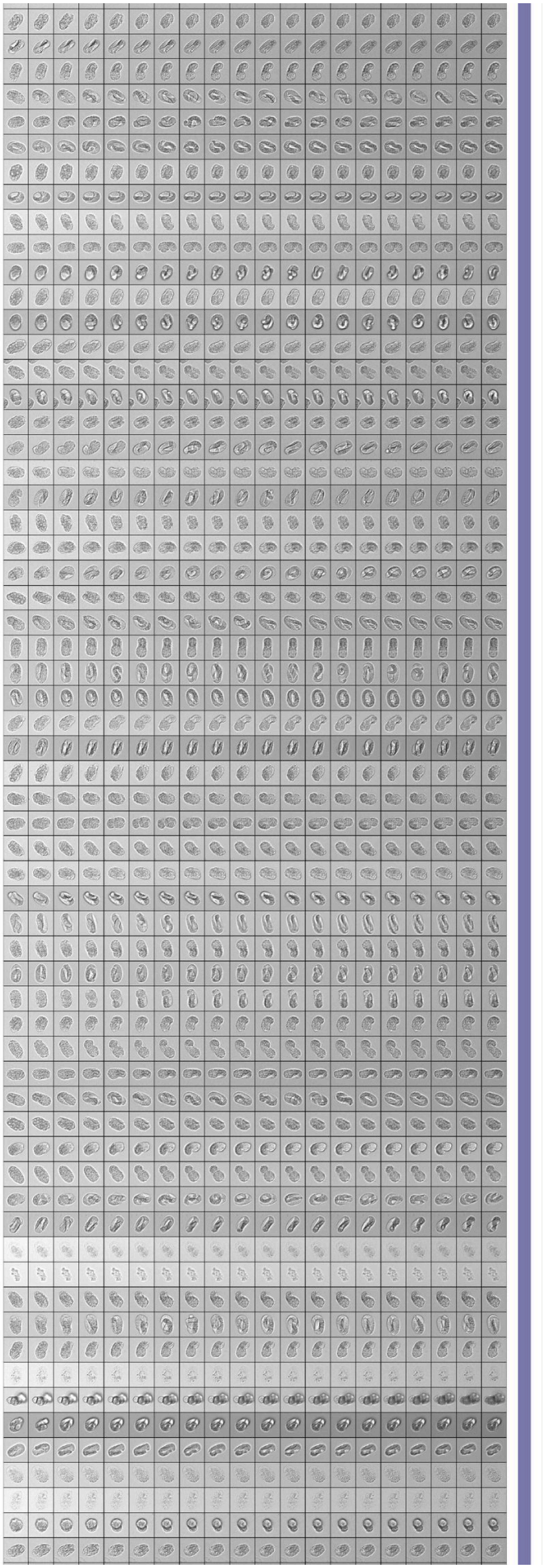
phenotypes during our imaging. Our classifier was able to identify developmental outcomes from this perturbation experiment (Figure 7, C). Specifically, we observed that embryos treated with 0.5 M NaCl either arrested in the *fold* stage (48%, n = 76/158 embryos) or died (37%, n = 58/158 embryos) during the time-lapse. At higher salt concentration (0.75 M NaCl), the majority of embryos died during imaging (81%, n = 129/160 embryos).











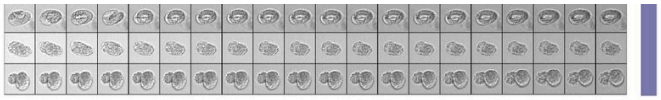


Figure 8

Micrograph time series data from the osmotic shock experiment.

We extracted every tenth frame from the time-course experiment where we exposed *C. elegans* embryos to 0 (green), 0.5 M (gold), and 0.75 M (purple) NaCl.

Reliably classifying other nematode species requires species-specific ground truth annotations

As a final test of our nematode classifier, we imaged embryonic development of two additional species of free-living rhabditid nematodes: an additional *Caenorhabditid* species, *Caenorhabditis portoensis,* and a more distantly related species, *Oscheius tipulae* (Figure 10, A, phylogeny based on **[15]**). We annotated 15 movies (Figure 9, B for representative images) of each species and used the trained model to classify the images from these experiments. The original model performed well at classifying *proliferation* (90% for *C. portoensis*, 92% for *O. tipulae*), *fold* (78%, 94%) and *hatch* (100% for both) in these data, but, as was the case with the *C. elegans* data, struggled to correctly classify morphogenesis stages [*bean (22%, 48%)* and *comma (15%, 7%*)] (Figure 9, C). During annotation, we noticed that *O. tipulae* failed to hatch during the imaging window of 16 hours. These data support observations that *O. tipulae* develops at a slower rate than *C. elegans* **[25]**, accounting for the absence of *hatch* in our confusion matrix (Figure 9, C).

Given the low performance and high confusion on morphogenesis stages (*bean* and *comma*) we next asked if we could improve classification by training a model that included ground truth annotations of data from the other two nematode species. We retrained the network with this additional data, and performance for all stages increased (e.g., *bean* correctly classified at 80% and 71% in *C. portoensis* and *O. tipulae*, respectively; Figure 9, C).

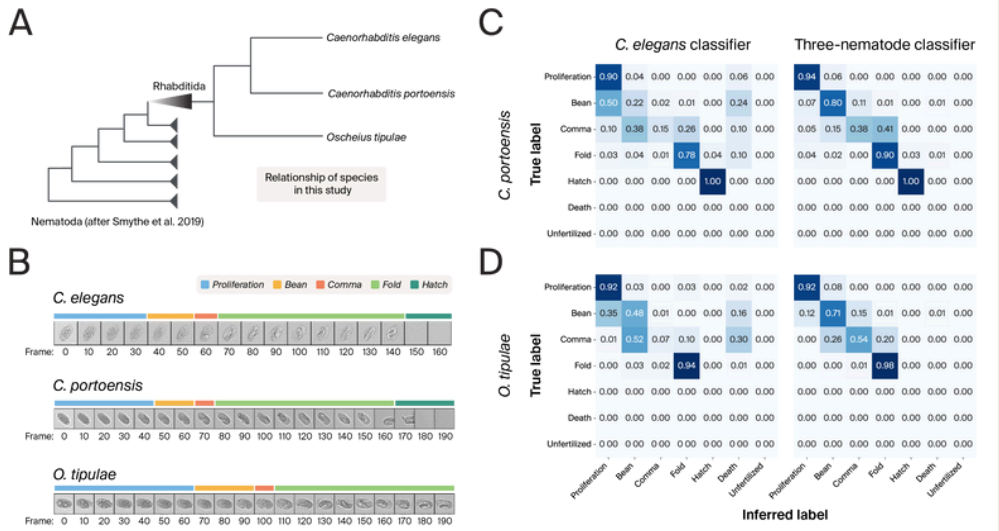


Figure 9

Extending our classifier to other nematode species.

(A) Phylogenetic relationship of the three nematodes studied in this project.

(B) Representative time-lapse images from each species, shown as a filmstrip containing every 10th frame of the original cropped movie.

(C) Confusion matrices for each species based on a classifier model trained only on *C. elegans* (left) and a new model trained on ground truth annotations from the additional nematode species (right).

Finally, we asked whether our model trained with images from additional nematode species performed better or worse when classifying our original *C. elegans* data. The addition of images for other nematode species resulted in improved performance for some of the stages, specifically *proliferation* (83% to 91%) and *death* (79% to 94%) (Figure 10, A–B). While there was improvement at classifying *comma* stage (47% to 65%), identification of the *bean* stage was poorer in the general model (77% to 56%) (Figure 10, A–B).

We're interested in seeing if these trends might improve with the addition of more data, and have included all of the documentation necessary to train new models. If you want to classify developmental outcomes from your own high-throughput imaging

experiments, we suggest using the model trained on all three species, as it performed better at classifying *hatch* and embryonic lethality (*death*).

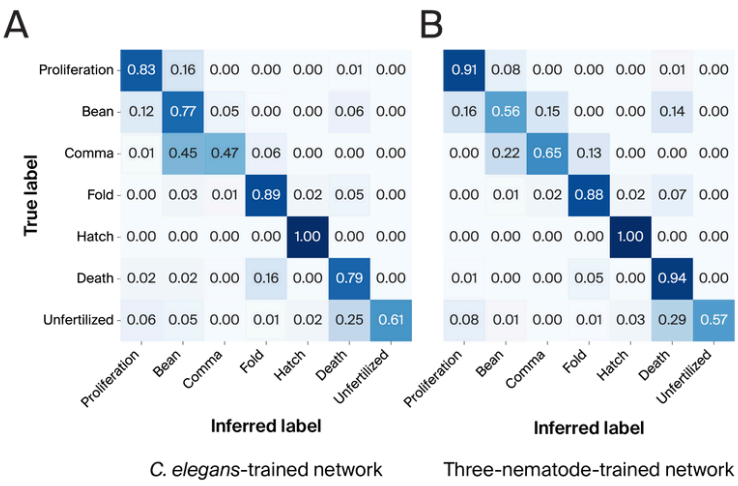


Figure 10

A more general model improves classification accuracy of *C. elegans* images.

Confusion matrices of validation data for the original network trained on only *C. elegans* images (A) versus the network trained on *C. elegans* and the two additional nematode species (B).

Materials and methods

Species and strains

The following strains were used in this study: *C. elegans*: N2 (wild-type), DQM327 (bmd75[*eef-1A.1p*::*his-58*::dendra::3xHA::*tbb-2* 3'UTR]) I; cpIs80 [*eef-1A.1p*::mKate2-C1::mKate2-GLO::PH::3xHA::tbb-2 3'UTR] II. *O. tipulae*: CEW1. *C. portoensis*: EG4788. We maintained all nematode strains used in this study on 60 mm NGM plates on an OP50 *E. coli* lawn using standard methods **[26]**.

Embryo isolation

We isolated nematode embryos by hypochlorite treatment of a minimum of three 60 cm NGM plates of gravid adults using a standard protocol **[17]**. Briefly, we washed gravid hermaphrodites off NGM plates using M9 media, then concentrated and treated with hypochlorite for 6–8 min, then washed repeatedly with M9 to remove the unreacted hypochlorite. To concentrate embryos following the final M9 wash for dispensing into 384-well plates for imaging, we decanted the M9 wash and examined 1 µl of embryo suspension. Our target concentration was ~50–75 embryos/µl. If too concentrated, we added an appropriate volume of M9, usually ~50–100 µl. We added 1 µl of embryo suspension to individual wells in a 384-well glass-bottom plate (Cellvis) containing 50 µl of M9 per well. For hyperosmotic perturbation experiments, we added embryos to the appropriate NaCl concentration (0.5 M or 0.75 M). To disperse embryos throughout the well, we gently pipetted the suspension up and down using a 200 µl pipette. We settled embryos to the bottom of the well in preparation for imaging by performing a brief centrifugation (1 min, 600 × g) in a table-top centrifuge (Sorvall X Pro Series) at room temperature (~21°C).

Microscopy

We performed all imaging experiments on a Nikon Ti2-E compound inverted microscope, equipped with an ORCA-Fusion BT digital sCMOS camera and configured for widefield imaging. We collected all data using a Plan Apo 20× 0.75 NA Air objective. We performed acquisition using High Content Analysis NIS-Elements software (version 54203). We performed object detection to select FOVs that contained a minimum number of embryos by designing a custom JOBS script to perform thresholding (script available here). Following tiled scans of wells containing embryos, we then imaged FOVs that met the object detection criteria every five minutes for 14–16 hours, to allow for embryos to complete development and hatch as L1 larvae.

> **SHOW ME THE DATA**: All of the cropped images used in this pub are available on Zenodo (DOI: 10.5281/zenodo.10211684)

Image processing and model training

We performed all image processing in Python. Briefly, we converted raw images from each dataset from Nikon's ND2 format to Zarr format, cropped embryos from each raw FOV, and calculated the moving mean and moving standard deviation for all cropped embryos.

We used PyTorch with PyTorch Lightning to facilitate dataset loading and model training. We wrote a custom dataloader to aggregate the time-lapse frames from all annotated cropped embryos and split the aggregated frames (from 95 *C. elegans* movies) into training, validation, and test sets. After training, we used the model checkpoint with the highest validation accuracy to infer (use the tool to provide a best guess for) stage labels for all cropped embryos. Finally, we post-processed the inferred labels (as described in Figure 6) to generate the final summary statistics shown in Figure 7. To calculate the confusion matrices, we generated an independent set of manually annotated embryos (from 55 *C. elegans* movies and 15 movies from *C. portoensis* and *O. tipulae*) that were not among the embryos used during training. For re-training a network on all three species of nematodes, we annotated additional frames (from 15 movies per species) for training, validation, and test sets as above.

> All our **code** in Python and additional documentation is available in this GitHub repository (DOI: 10.5281/zenodo.10247028).

We wrote a separate CLI script to perform each of these steps (e.g., ND2 conversion, embryo cropping, model training, label classification, post-processing, etc). Please see the README in our GitHub repo for more details and examples of how to use each of these scripts. We used ChatGPT and GitHub Copilot to write some code.

We added timestamps for figures using a Napari plugin (napari-timestamper).

Key takeaways

We trained a ResNet-18 neural network to identify key developmental stages of nematode embryos and classify endpoint results from high-throughput imaging experiments, distinguishing between embryonic lethality and successful hatching. We

chose a deep learning model that relied on supervised learning and human annotation of key frames, but trained a model that took advantage of the dynamic nature of the time-course data. While the model performed well at identifying most of the developmental stages as well as classifying lethality and hatching, we found it classified the subtle differences that make up the key morphogenesis phases of nematode development less robustly. Finally, we found that we needed to add image data from other species to train a new model that could perform well in identifying stages of nematodes beyond *C. elegans*.

We hope that *C. elegans* researchers who want to phenotype mutants at scale or use forward or reverse genetic approaches at high throughput will find this tool useful. More broadly, we hope that our workflow and approach might be useful to anyone wanting to apply deep learning to time-course data.

Next steps

While we're not pursuing this work further, as our scientific strategy has shifted, we'd be interested in hearing whether this approach is useful for building classifiers for other time-course imaging data. We hope that the basic tools we've included in our GitHub repository will be a useful starting point for anyone interested in building a classifier with their own imaging data. We're particularly curious if researchers who would find this tool useful for their own science have the required computational expertise to use it based on the documentation we've provided. If you do use this resource, we'd love to hear about your experience.

References

1 Yao K, Rochman ND, Sun SX. (2019). Cell Type Classification and Unsupervised Morphological Phenotyping From Low-Resolution Images Using Deep Learning. https://doi.org/10.1038/s41598-019-50010-9

2 Amitay Y, Bussi Y, Feinstein B, Bagon S, Milo I, Keren L. (2023). CellSighter: a neural network to classify cells in highly multiplexed images. https://doi.org/10.1038/s41467-023-40066-7

3 Zahedi A, On V, Phandthong R, Chaili A, Remark G, Bhanu B, Talbot P. (2018). Deep Analysis of Mitochondria and Cell Health Using Machine Learning. https://doi.org/10.1038/s41598-018-34455-y

4 Choi J, Kim H-J, Sim G, Lee S, Park WS, Park JH, Kang H-Y, Lee M, Heo WD, Choo J, Min H, Park Y. (2021). Label-free three-dimensional analyses of live cells with deep-learning-based segmentation exploiting refractive index distributions. https://doi.org/10.1101/2021.05.23.445351

5 Cross-Zamirski JO, Mouchet E, Williams G, Schönlieb C-B, Turkki R, Wang Y. (2022). Label-free prediction of cell painting from brightfield images. https://doi.org/10.1038/s41598-022-12914-x

6 Zaritsky A, Jamieson AR, Welf ES, Nevarez A, Cillay J, Eskiocak U, Cantarel BL, Danuser G. (2021). Interpretable deep learning uncovers cellular properties in label-free live cell images that are predictive of highly metastatic melanoma. https://doi.org/10.1016/j.cels.2021.05.003

7 Villoutreix P. (2021). What machine learning can do for developmental biology. https://doi.org/10.1242/dev.188474

8 Duncan AR, Khokha MK. (2016). Xenopus as a model organism for birth defects—Congenital heart disease and heterotaxy. https://doi.org/10.1016/j.semcdb.2016.02.022

9 Truong BT, Artinger KB. (2021). The power of zebrafish models for understanding the co-occurrence of craniofacial and limb disorders. https://doi.org/10.1002/dvg.23407

10 Aiello NM, Stanger BZ. (2016). Echoes of the embryo: using the developmental biology toolkit to study cancer. https://doi.org/10.1242/dmm.023184

11 Theilgaard Lassen J, Fly Kragh M, Rimestad J, Nygård Johansen M, Berntsen J. (2023). Development and validation of deep learning based embryo selection across multiple days of transfer. https://doi.org/10.1038/s41598-023-31136-3

12 Marsh P, Radif D, Rajpurkar P, Wang Z, Hariton E, Ribeiro S, Simbulan R, Kaing A, Lin W, Rajah A, Rabara F, Lungren M, Demirci U, Ng A, Rosen M. (2022). A proof of concept for a deep learning system that can aid embryologists in predicting blastocyst survival after thaw. https://doi.org/10.1038/s41598-022-25062-z

13 Voronov DA, Panchin YV. (1998). Cell lineage in marine nematode *Enoplus brevis*. https://doi.org/10.1242/dev.125.1.143

14 Schulze J, Schierenberg E. (2011). Evolution of embryonic development in nematodes. https://doi.org/10.1186/2041-9139-2-18

15 Baris Atakan H, Alkanat T, Cornaglia M, Trouillon R, Gijs MAM. (2020). Automated phenotyping of Caenorhabditis elegans embryos with a high-throughput-screening microfluidic platform. https://doi.org/10.1038/s41378-020-0132-8

16 He K, Zhang X, Ren S, Sun J. (2015). Deep Residual Learning for Image Recognition. https://doi.org/10.48550/ARXIV.1512.03385

17 Porta-de-la-Riva M, Fontrodona L, Villanueva A, Cerón J. (2012). Basic Caenorhabditis elegans Methods: Synchronization and Observation. https://doi.org/10.3791/4019

18 Sulston JE, Schierenberg E, White JG, Thomson JN. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. https://doi.org/10.1016/0012-1606(83)90201-4

19 Luo S, Kleemann GA, Ashraf JM, Shaw WM, Murphy CT. (2010). TGF-β and Insulin Signaling Regulate Reproductive Aging via Oocyte and Germline Quality Maintenance. https://doi.org/10.1016/j.cell.2010.09.013

20 Andux S, Ellis RE. (2008). Apoptosis Maintains Oocyte Quality in Aging Caenorhabditis elegans Females. https://doi.org/10.1371/journal.pgen.1000295

21 Lamitina ST, Morrison R, Moeckel GW, Strange K. (2004). Adaptation of the nematode *Caenorhabditis elegans* to extreme osmotic stress. https://doi.org/10.1152/ajpcell.00381.2003

22 Williams-Masson EM, Malik AN, Hardin J. (1997). An actin-mediated two-step mechanism is required for ventral enclosure of the *C. elegans* hypodermis. https://doi.org/10.1242/dev.124.15.2889

23 Drewnik ED, Wiesenfahrt T, Smit RB, Park Y-J, Pallotto LM, Mains PE. (2021). Tissue-specific regulation of epidermal contraction during*Caenorhabditis elegans*embryonic morphogenesis. https://doi.org/10.1093/g3journal/jkab164

24 Solomon A, Bandhakavi S, Jabbar S, Shah R, Beitel GJ, Morimoto RI. (2004). Caenorhabditis elegans OSR-1 Regulates Behavioral and Physiological Responses to Hyperosmotic Environments. https://doi.org/10.1534/genetics.167.1.161

25 Félix M.-A. Oscheius tipulae. (2006). https://www.doi.org/10.1895/wormbook.1.119.1

26 Brenner S. (1974). THE GENETICS OF *CAENORHABDITIS ELEGANS*. https://doi.org/10.1093/genetics/77.1.71