Deepcell

The Deepcell Human Foundation Model (HFM)

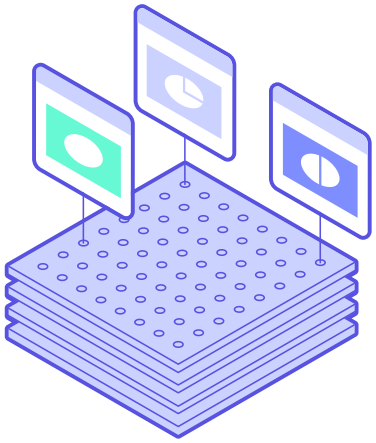
Deep learning embeddings and computer vision morphometrics for high-dimensional single cell morphology analysis

Highlights

- The Deepcell HFM combines deep learning and computer vision methods to extract cell features that quantitatively and reproducibly represent cell morphology
  - Deep learning provides quantitative descriptions of cell features using neural networks
  - Computer vision provides a quantitative assessment of cell and biological features using discrete image analysis algorithms
- The deep learning aspect of the HFM is based on a self-supervised learning algorithm
  - Trained using millions of cell images from many biological sample types
  - Detects differences in cell morphology without labeled training data
- 115 cell morphology dimensions are generated by the HFM
- The HFM enables real-time cell characterization & classification and sorting capabilities on Deepcell's REM-I platform

Introduction

Morphology is a fundamental cell property associated with identity, state, and function, but it is usually characterized in a few standard dimensions such as diameter, perimeter, or area, or with qualitative descriptions. Advanced machine learning algorithms allow for extracting and interpreting cell morphology features with a multidimensional, unbounded, and quantitative assessment. The Deepcell REM-I platform combines label-free imaging, deep learning, computer vision morphometrics, and gentle cell sorting to leverage multidimensional single cell morphology as a quantitative readout. The Deepcell Human Foundation Model, which is compatible with the REM-I platform, is a hybrid architecture that combines self-supervised learning (SSL) and morphometrics (computer vision) to extract 115 dimensional embedding vectors representing cell morphology from high-resolution brightfield cell images.



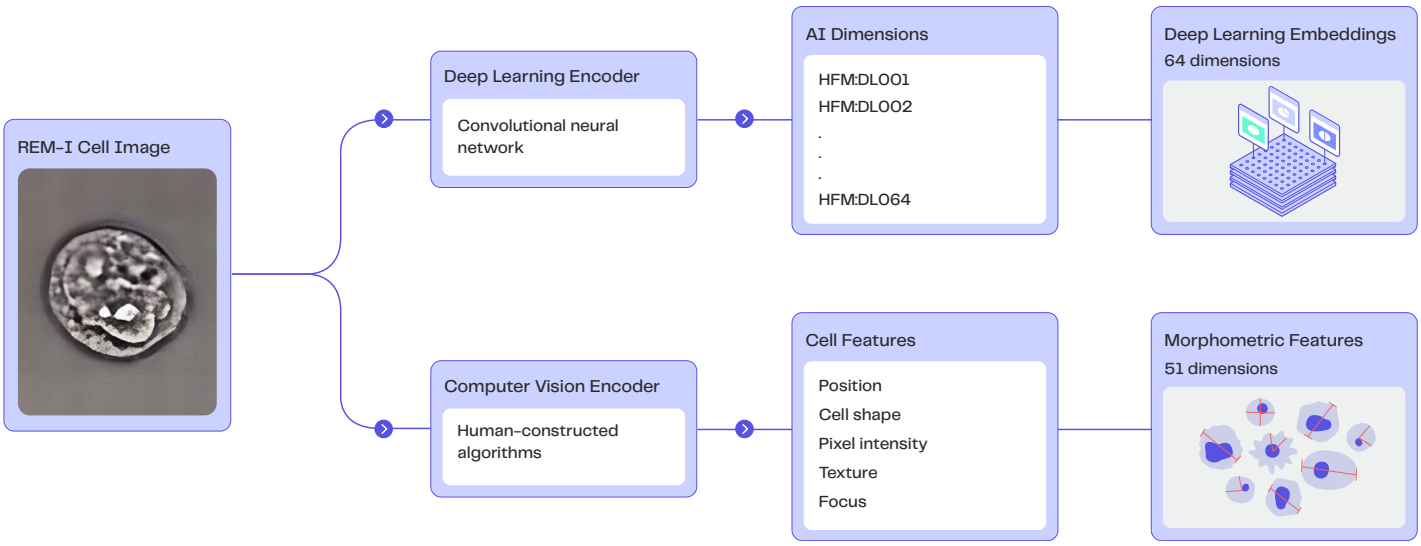


Figure 1. The Deepcell HFM is a feature extractor. *The HFM combines deep learning and computer vision encoders to extract 64 deep learning embeddings and 51 morphometric features from cell images and encode them into multidimensional numerical vectors.*

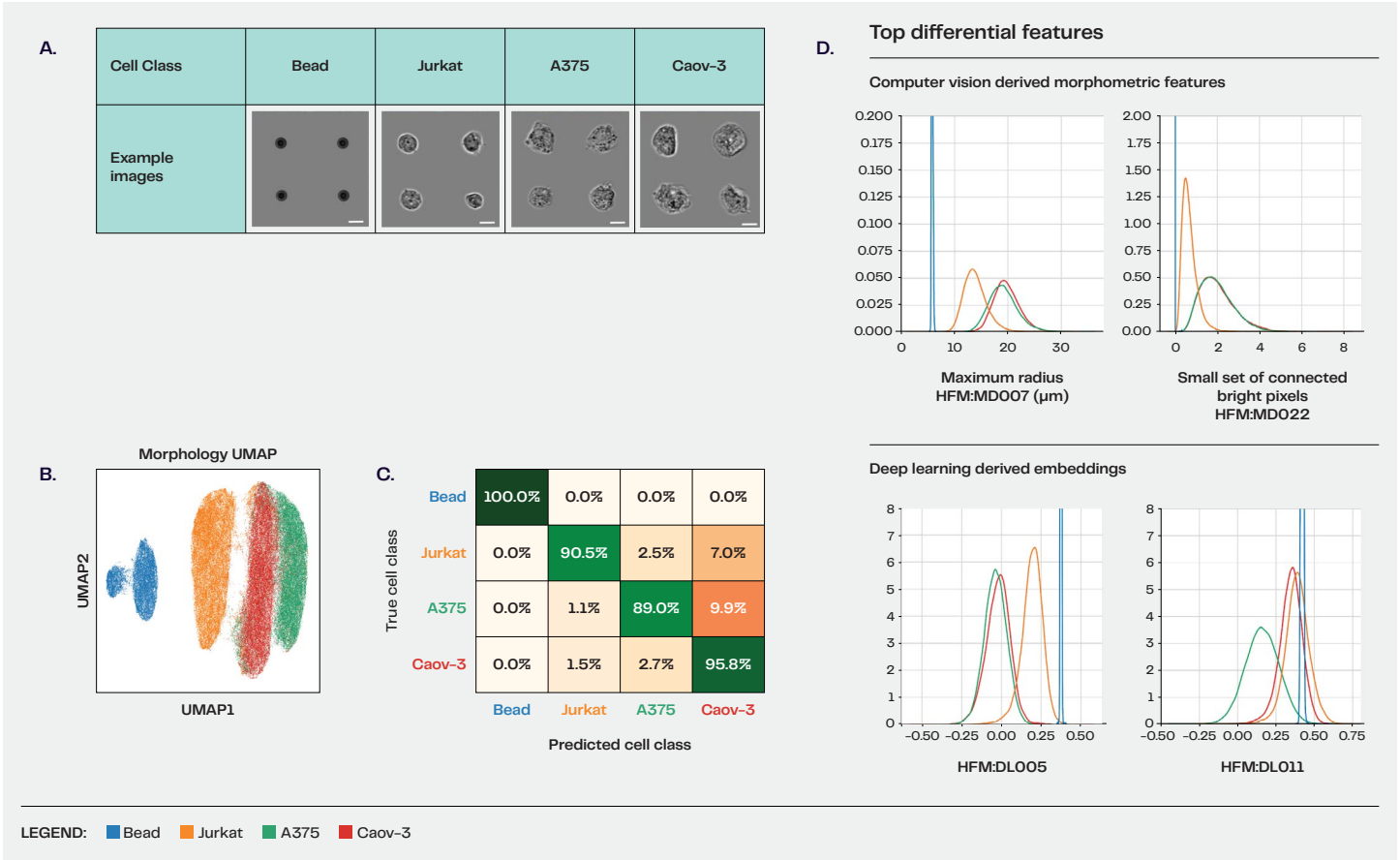


Figure 2. Evaluation of the Deepcell HFM with polystyrene beads and cell lines. *6µm polystyrene beads and cancer (A375 and Caov-3) and immune (Jurkat) cell lines were imaged on the Deepcell REM-I platform, then combined in silico to evaluate the classification performance of the HFM.* ***(A)*** *Table of cell classes and corresponding representative images (scale bar, 10 µm).* ***(B)*** *Morphology UMAP of beads (blue) and cell lines (Jurkat in orange, Caov-3 in red, A375 in green).* ***(C)*** *Confusion matrix showing that the HFM predicts cell classes with high accuracy (89% or greater).* ***(D)*** *Density plots of the top four differential features (HFM:MD007, HFM:MD022, HFM:DL005, HFM:DL011; see Tables 1 and 2 for description) indicated for beads (blue) and cell lines (Jurkat in orange, Caov-3 in red, A375 in green).*

Extracting morphological information from cell images in two ways

Morphometric features are human-interpretable, quantitative metrics of cell morphology including cell size, shape, texture, and intensity. Deep learning embeddings are information-rich metrics of cell morphology with powerful discriminative capabilities, but may not be human-interpretable. Combining deep learning embeddings and morphometric features in the Deepcell HFM provides both accuracy and interpretability and is a fast and effective cell feature extractor, enabling real-time cell classification and sorting. The Deepcell HFM has strong generalization capabilities that enable hypothesis-free sample exploration and efficient generation of application-specific models. The Deepcell HFM contains 115 combined embedding dimensions, including 64 deep learning derived embeddings (Figure 1, Table 1) and 51 computer vision derived morphometrics (Figure 1, Table 2). Computationally, morphometrics can be generated by discrete computer vision algorithms. However, some morphometrics are too computationally intensive to compute in real time while cells are in flow. To overcome this limitation, Deepcell has also incorporated deep learning methods to impute the most computationally intensive morphometrics into our foundation model.

Training the Deepcell HFM

Foundation models require training on a large quantity of data. The Deepcell HFM backbone model, which extracts image features, is based on a convolutional neural network architecture. Model training applies a self-supervised learning approach, which learns image features without labels and generates deep learning embeddings that are orthogonal to each other and orthogonal to morphometric features. The Deepcell HFM includes over 11 million parameters and was trained using millions of cell images taken across many biological sample types.

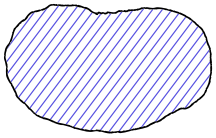
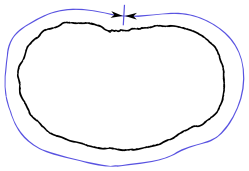
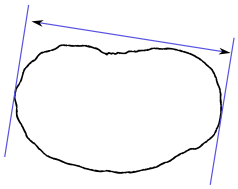
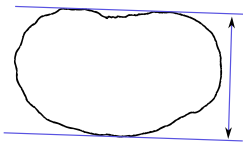
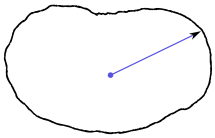
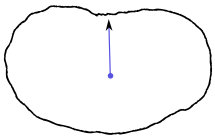
Evaluation of the Deepcell HFM performance

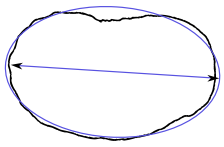
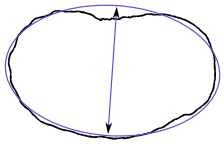
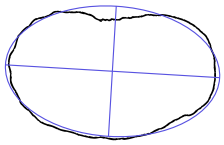
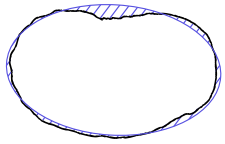
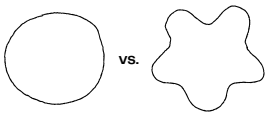
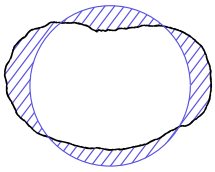
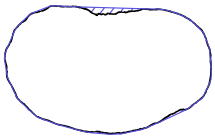
To evaluate Deepcell HFM classification performance, polystyrene beads and cancer (A375 and Caov-3) and immune (Jurkat) cell lines with distinct morphological features were imaged on the Deepcell REM-I platform. Images were combined *in silico*, then deep learning embeddings and morphometric features were extracted and combined for embeddings analysis based on the HFM (Figure 2A). Features were standardized and data were projected into a lower dimensional principal components analysis (PCA) basis. Nearest neighbors were computed in the PCA space, then used to compute 2D Uniform Manifold Approximation and Projections (UMAP). The morphology UMAP shows a clear separation between polystyrene beads and cell lines (Figure 2B). In cell line populations, while the Jurkat immune cell population was clearly distinct from the cancer cell lines (A375 and Caov-3) (Figure 2A and 2B), the two cancer lines show overlap, suggesting that the immune and the cancer lines are morphologically distinct (Figure 2A and 2B). The HFM characterized and classified beads and the three cell lines with high accuracy (100%, 90.5%, 89.0% and 95.8% for beads, Jurkat, A375, and Caov-3, respectively) (Figure 2C). Interestingly, A375 and Caov-3 cell lines were nearly identical in size by maximum radius (HFM:MD007), small set of connected bright pixels (HFM:MD022) (Figure 2D, top panels), and deep learning embedding 5 (HFM:DL005) (Figure 2D, bottom left panel), indicating that cell features other than size contribute to differentiating A375 and Caov-3 such as deep learning embedding 11 (HFM:DL011) (Figure 2D, bottom right panel).

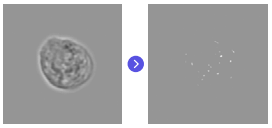
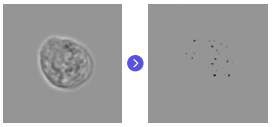
Table 1. Panel of the deep learning derived embeddings in the Deepcell HFM. *Current list of deep learning derived embeddings in the Deepcell HFM.*

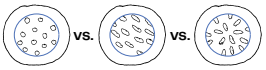
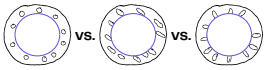
Deepcell Deep Learning Dimension #	Description
HFM:DL001 ... HFM:DL064	64 quantitative descriptions of cell features derived from deep learning models based on neural networks

Table 2. Panel of the morphometric features in the Deepcell HFM. *Current list of morphometric features in the Deepcell HFM. More features may be added to this list over time as we continue to develop the model.*
**Denotes metrics that may also be referred to as blobs or granules in the literature.*

Category	Deepcell Morphometric Dimension #	Morphometric Name	Description	Schematic
Position Features	HFM:MD001	Centroid X axis (μm)	X axis position of the cell relative to the camera's field of view	
	HFM:MD002	Centroid Y axis (μm)	Y axis position of the cell relative to the camera's field of view	
Cell Shape Features	HFM:MD003	Area (μm²)	Cell area	
	HFM:MD004	Perimeter (μm)	Length of the cell outline	
	HFM:MD005	Maximum caliper distance (μm)	Width of widest possible box around the cell	
	HFM:MD006	Minimum caliper distance (μm)	Width of narrowest possible box around the cell	
	HFM:MD007	Maximum radius (μm)	Largest radius from center of cell to cell border	
	HFM:MD008	Minimum radius (μm)	Shortest radius from center of cell to cell border	

Category	Deepcell Morphometric Dimension #	Morphometric Name	Description	Schematic
	HFM:MD009	Long ellipse axis (μm)	Long axis of best fit ellipse	
	HFM:MD010	Short ellipse axis (μm)	Short axis of best fit ellipse	
	HFM:MD011	Ellipse elongation (unitless)	Aspect ratio of best fit ellipse Metric: 0 indicates a circle, 1 indicates a line	
	HFM:MD012	Ellipse similarity (unitless)	Deviation from an elliptical shape Metric: 0 indicates a perfectly elliptical shape	
	HFM:MD013	Roundness (unitless)	Roundness is a measure for circularity or compactness of the shape Metric: Range is 0 to 1, where 1 is a perfect circle	
	HFM:MD014	Circle similarity (unitless)	Deviation from a circular shape Metric: 0 indicates a perfectly circular shape	
	HFM:MD015	Convex shape (unitless)	Ratio of the area of the convex hull of the cell to the total area of the cell	

Category	Deepcell Morphometric Dimension #	Morphometric Name	Description	Schematic
Pixel Intensity Features	HFM:MD016	Mean pixel intensity (arbitrary units)	Mean pixel grayscale value, refers to how much light a cell absorbs and/or scatters Metric: Range is -1 to 1	
	HFM:MD017	Standard deviation of pixel intensity	Standard deviation of pixel grayscale values gives an indication of the uniformity of pixel intensity within the cell	
	HFM:MD018	Pixel intensity 25th percentile	25th percentile of pixel grayscale values	
	HFM:MD019	Pixel intensity 75th percentile	75th percentile of pixel grayscale values	
	HFM:MD020	Positive fraction	Fraction of pixels with a grayscale value significantly above background	
	HFM:MD021	Negative fraction	Fraction of pixels with a grayscale value significantly below background	
Texture Features	HFM:MD022	Small set of connected bright pixels*, integral	Sum of pixel intensities in regions identified as being small bright structures Small is described as ~8 pixels but smaller sets could be detected	
	HFM:MD023	Small set of connected dark pixels*, integral	Sum of pixel intensities in regions identified as being small dark structures Small is defined as ~8 pixels but smaller sets could be detected	
	HFM:MD024	Large set of connected bright pixels*, integral	Sum of pixel intensities in regions identified as being large bright structures Large is defined as ~32 pixels but larger sets could be detected	

Category	Deepcell Morphometric Dimension #	Morphometric Name	Description	Schematic
	HFM:MD025	Large set of connected dark pixels*, integral	Sum of pixel intensities in regions identified as being large dark structures Large is defined as ~32 pixels but larger sets could be detected	
	HFM:MD026 – HFM:MD027	Image moments (2 moments)	Two values that describe the weighted average or distribution of the image pixel intensities within the cell in a rotation and scale-invariant manner	
	HFM:MD028 – HFM:MD037	Local binary patterns – center (10 patterns)	The 10 Local Binary Pattern (LBP) – Center features determine the texture inside the cell and describe the appearance near the center of the cell	
	HFM:MD038 – HFM:MD047	Local binary patterns – periphery (10 patterns)	The 10 Local Binary Pattern (LBP) – Periphery features determine the texture at the periphery of the cell and describe the appearance near the edge of the cell	
Focus Features	HFM:MD048	Image sharpness (arbitrary units)	A measure of how sharp or smooth the image is. Typically an out-of-focus image is less sharp than an in-focus image.	
	HFM:MD049	Image focus (µm)	Estimate of the distance of the cell to the focal plane of the microscope	

Category	Deepcell Morphometric Dimension #	Morphometric Name	Description	Schematic
	HFM:MD050	Ring width (arbitrary units)	Our imaging modality creates a dark or bright ring to appear around the cell, which is larger and more intense the more out of focus the cell is. This feature estimates the width of the ring.	
	HFM:MD051	Ring intensity (arbitrary units)	Our imaging modality creates a dark or bright ring to appear around the cell, which is larger and more intense the more out of focus the cell is. This feature estimates the intensity of the ring.	

Summary

The Deepcell Human Foundation Model combines deep learning and computer vision to efficiently extract features that represent cell morphology with high accuracy and interpretability, enabling real-time cell classification and sorting. Applications include high-throughput sample characterization, disease detection, target cell enrichment, drug and functional screening, and multi-omic integration.

References

1. Salek M. et al. Real time morphological characterization and sorting of unlabeled viable cells using deep learning. bioRxiv, 2022. DOI: 10.1101/2022.02.28.482368.
2. Mavropoulos A. et al. Artificial-intelligence-driven morphology-based enrichment of malignant cells from body. Mod Pathol, 2023. DOI: 10.1016/j.modpat.2023.100195.
3. Masaeli M. et al. Multiparameter mechanical and morphometric screening of cells. Scientific Reports, 2016. DOI: 10.1038/srep37863.

Resources

- → REM-I Product Sheet
- → REM-I Platform Brochure
- → Public Datasets and Axon Demo
- → Other Resources

Contact us

INFO@DEEPCELLBIO.COM · DEEPCELL.COM

Deepcell