Deep learning models capture high-dimensional features for cell morphology analysis from brightfield images

**Methods continued**

**Image Feature Extraction and Classification Architecture**

**A** Training architecture

- **Model parameters**: -1 million
- **Cell images**: 4,055,556
- **Imaging run**: B14
- **Biological samples**: 
- **Embedding dimensions**: 115 total
- **Forward pass latency**: 64 AI + 51 morphometrics
- **Embedding dimensions**: 115 total
- **Forward pass latency**: 64 AI + 51 morphometrics

**B** Less terms

- **Invariance Loss**: The distance between two embeddings from augmented images of the same cell is maintained.
- **Variance Loss**: The variance of each embedding dimension over a training batch of images is minimized.
- **Covariance Loss**: The covariances between pairs of embedding and morphometric dimensions over a training batch of images are maintained.

**C** Evaluation architecture

- **Grayscale**
- **Color**
- **Texture**
- **Intensity**
- **Granules**

**D** HFM details

- **Backbone deep learning model**: ResNet18
- **Model parameters**: -1 million
- **Cell images**: 4,055,556
- **Imaging run**: B14
- **Biological samples**: 
- **Embedding dimensions**: 115 total
- **Forward pass latency**: 64 AI + 51 morphometrics

**Figure 3. Deepcell HFM training and inference architectures.** (A) HFM training architecture. The backbone model, which extracts image features, is based on the publicly available convolutional deep neural network architecture. Training combines VICReg self-supervised learning, which learns image feature without labels, with orthogonal morphometry features to improve model performance and interpretability. (B) Less terms: Invariance Loss, Variance Loss, and Covariance Loss. The backbone architecture and the morphometric feature encoder are agglomerated with custom trained classification heads to generate embedding vectors and classify cells. Calculations are performed on the HFM backbone using Foward (Xception, Pay Top, Deep, VGG), and a custom FCNN to identify and sort cells in dataset. (B) HFM hyperparameters, training dataset details, and inference latency.

**Model Performance Evaluation**

- **HFM performance was evaluated using two simple datasets comprising multiple known cell types mixed in artificial cell populations.**
- **Figure 4**: mixture of four synthetic cell types ranging from 20-40 µm diameter and 20-50 µm diameter, which relate to cell structure like vesicles.
- **Classification accuracy is 100% with several highly interpretable features driving classification.**

**Results**

**A** Cell Class

<table>
<thead>
<tr>
<th>Diameter</th>
<th>Small</th>
<th>Large w/ particles</th>
<th>Large collapsed</th>
<th>Large w/ granules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training images</td>
<td>5,677</td>
<td>3,158</td>
<td>4,129</td>
<td>5,716</td>
</tr>
<tr>
<td>Test images</td>
<td>7,175</td>
<td>5,950</td>
<td>4,129</td>
<td>7,175</td>
</tr>
</tbody>
</table>

**B** Morphology UMAP

**C** Synthetic cell classification

**D** Top differential morphometric features

**Figure 4. Evaluation of HFM synthetic cell classification.** Synthetic cells with diverse structural features were imaged on the REM-I platform. (A) Table of cell classes, number of images used for training and testing, representative images of each class. (B) Morphometry: UMAP of HFM embeddings colored by ground truth. (C) Confusion matrix: HFM predicts synthetic cell class with 100% accuracy in all 4 classes. (D) Top differential morphometric and deep learning features. Utilizing both AI-derived features and computer vision features enhances model explainability.

**Figure 5. Evaluation of mixed cell types.** 4 µm polystyrene beads and three cancer cell lines with distinct features were imaged on the REM-I platform and combined in silico to evaluate HFM classification performance. (A) Table of cell classes, number of cell images analyzed, and representative images. (B) Morphometry: UMAP of 3 µm beads and cancer cell lines colored by ground truth. (C) Confusion matrix: HFM predicts cell classes with 100% accuracy. (D) Density plot of top differential features by cell class. All cell lines have overlapping cell size distributions. A375 and Caov-3 are nearly identical in size, indicating that cell features other than size must be utilized to accurately differentiate A375 and Caov-3.

**Conclusions**

- **Combining deep learning and morphometric features improves accuracy and interpretability.**
- **HFM characterizes and classifies synthetic cells and biological cell samples with high accuracy.**
- **Features beyond cell size are utilized to drive classification performance.**
- **The Deepcell Human Foundation Model (HFM) is fast and effective at extracting cell features, enabling real-time cell classification and sorting.**
- **Real-world applications of Deepcell’s HFM on the REM-I platform include hypothesis-free sample exploration and adaptation to specific classification tasks.**