COMPUTER-AIDED CLASSIFICATION OF CELLS IN COMPLEX BRAIN TISSUE FROM 5-CHANNEL 3-D CONFOCAL DATASETS

Yousef Al-Kofahi1, Gang Lin1, Christopher Bjorussen2, Karen Smith3, Arunachalam Narayanaswamy1, Badrinath Roysam1, William Shain3

1 Electrical, Computer, and Systems Engineering Department, Rensselaer Polytechnic Institute, Troy, NY 12180-3590
2 Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180-3590
3 Laboratory of Nervous System Disorders, Wadsworth Center, Albany, NY 12201-0509.

Abstract

The cellular organization of brain tissue is truly complex. This work presents a computational method to identify the principal cell types in three-dimensional (3-D) confocal image stacks with multiple fluorescent channels. The cells are classified into four major classes (Neurons, Microglia, Astrocytes and Endothelials) by using a two-step classifier that applies fuzzy c-means clustering followed by Support Vector Machines (SVM). The resulting classification results were validated against a human expert, and the accuracy of the classifier was 95.5% in the correctly segmented nuclei.

Spectral Unmixing and Image Segmentation

- The image is decomposed into 5 channels
- Each channel is segmented independently

Features Extraction

- **Intrinsic Features:**
  - Intrinsic features include a set of morphometric, topological, and intensity-based features for each nucleus.
  - We have a total of 23 intrinsic features

- **Associative Features:**
  - Associative features are computed for each nucleus to quantify relationships with other channels.
  - We have a total of 8 associative features

  - Examples:
    - Neuronal nuclei are surrounded by Nissl signal
    - Microglial nuclei are surrounded by IBA-1 signal
    - Astrocyte nuclei are proximal to a point of convergence of processes
    - Endothelial cells are adjacent vessels

  - **The Feature Vector:**
    - Each Nucleus is represented by a feature vector
    - We have a total of 31 intrinsic and Associative features

Cells Classification

- **Initial Fuzzy c-means clustering**
  - Class 1: Neurons
  - Class 2: Microglia
  - Class 3: Astrocytes + Endothelials

- **Training Set Extraction**
  - (Optional Step): Manual Correction of Training Set Errors

- **SVM Training** / **SVM Classification**
  - Final k-means clustering of Astrocytes and Endothelials

- **Summary cell classification results using all the cells**

<table>
<thead>
<tr>
<th>Class</th>
<th>Neurons</th>
<th>Microglia</th>
<th>Astrocytes + Endothelials</th>
<th>Total Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>348</td>
<td>324</td>
<td>197</td>
<td>246</td>
</tr>
<tr>
<td></td>
<td>9.4%</td>
<td>8.1%</td>
<td>24.5%</td>
<td></td>
</tr>
</tbody>
</table>

- **Summary cell classification results after 5 manual corrections to the training set (using all the cells)**

<table>
<thead>
<tr>
<th>Class</th>
<th>Neurons</th>
<th>Microglia</th>
<th>Astrocytes + Endothelials</th>
<th>Total Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>341</td>
<td>324</td>
<td>197</td>
<td>246</td>
</tr>
<tr>
<td></td>
<td>11.1%</td>
<td>8.1%</td>
<td>24.5%</td>
<td></td>
</tr>
</tbody>
</table>

- **The Validation Process**
  - The classification results were validated against a human expert validation
  - The validation process starts with validating the nuclear segmentation
  - Some of the segmentations errors include over/under-segmented nuclei and nuclei near the image edge
  - We used a dataset of 1019 cells. Among them, 888 are correctly segmented.

**Classification Results**

- **Initial Clustering**
  - Fuzzy c-means clustering is used to divide the cells into 3 clusters
  - Each cell \( j \) has a degree of membership to each cluster:
    \[ d_j = \int_{0,1} \sum d_j = 1 \]
  - Each cell \( j \) is initially assigned to the cluster with the highest degree of membership \( d_j^{\text{max}} \)

- **Training Set Extraction**
  - Most of the clustering errors are expected to be at the boundaries of the clusters
  - Cells close to the centroid of the cluster tend to have higher degrees of membership
  - The training set for class \( j \) contains all the cells with:
    \[ d \geq (\text{max}(d_j) - \text{median}(d_j)) / 2 \]
  - The training set contained 289 cells

- **SVM Training and classification**
  - A multi-category SVM classifier is first trained using the training set extracted in the previous step
  - The support vectors are then used to classify the unlabeled cells
  - The SVM’s solution approximates the Bayes decision rules without estimating conditional probabilities.

- **Optional Manual corrections of Training Set Errors**
  - We implemented an interactive scheme by incorporating the user’s feedback into our classifier.
  - An expert user can correct some of the all of the errors in the training set to provide better classification results.

- **Final Clustering of Astrocytes and Endothelials**
  - Each cell is represented by 31 features, but only 28 features were used in the classification done so far.
  - In the final clustering step, we used 8 features only, including the 3 unused features
  - In this step, k-means clustering is used.

References


Acknowledgements

Supported in part by NIBIB, EB-000359, NINDS, NS-044287, & NSF, ECC-9986821 is gratefully acknowledged.