

**Improved Recognition of Protein Subcellular Location Patterns via Classifier Ensembles and New Features Derived From Protein Fluorescence Microscope Images****Huang, Kai<sup>\*</sup>, Murphy, Robert F.*****Departments of Biological Sciences and Biomedical Engineering and Center of Automatic Learning and Discovery, Carnegie Mellon University, Pittsburgh, PA, USA***

Protein location is critical to protein function, and systematic approaches to determine subcellular locations for entire proteomes are needed. An automatic system to recognize protein subcellular location patterns via protein fluorescence microscope images has been described previously, in which a neural network classifier was able to recognize all major subcellular organelles by using numerical features derived from those images. Here, we report improved results on classifying both 2D and 3D protein fluorescence microscope images by introducing new features from Gabor and Daubechies wavelet transformations and new classifiers including support vector machines with linear, radial basis, polynomial, and exponential radial basis kernel functions, and ensemble methods such as AdaBoost, Bagging, and Mixture-of-Experts. The optimal parameters for each classifier were selected via 10-fold cross validation, after which optimal majority-voting classifier ensembles were formed by combining all eight classifiers including neural networks for both 2D and 3D protein fluorescence microscope images. Statistically better classification accuracy was observed over previous results and the overall error rate was reduced by one third. On average, we can achieve 92% accuracy on 10 major protein subcellular location patterns in 2D images and 96% accuracy on 11 patterns in 3D images. The CPU time costs of calculating different feature subsets from the images were analyzed and compared to their corresponding performance. Finally, we conducted classification of image sets by plurality voting and achieved essentially perfect classification accuracy.