

## Chapter 4.5

# Ensemble of Neural Networks for Automated Cell Phenotype Image Classification

**Loris Nanni**

*Università di Bologna, Italy*

**Alessandra Lumini**

*Università di Bologna, Italy*

### **ABSTRACT**

Subcellular location is related to the knowledge of the spatial distribution of a protein within the cell. The knowledge of the location of all proteins is crucial for several applications ranging from early diagnosis of a disease to monitoring of therapeutic effectiveness of drugs. This chapter focuses on the study of machine learning techniques for cell phenotype image classification and is aimed at pointing out some of the advantages of using a multi-classifier system instead of a stand-alone method to solve this difficult classification problem. The main problems and solutions proposed in this field are discussed and a new approach is proposed based on ensemble of neural networks

trained by local and global features. Finally, the most used benchmarks for this problem are presented and an experimental comparison among several state-of-the-art approaches is reported which allows to quantify the performance improvement obtained by the approach proposed in this chapter.

### **INTRODUCTION**

Nowadays, as a consequence of having obtained the sequence of numerous genomes and subsequent identification of the encoded proteome, understanding the function of proteins at the cellular level has become one of the most important objectives in biological sciences (Chebira et al., 2007). One of the main aspects to deeply under-

DOI: 10.4018/978-1-60960-818-7.ch4.5

stand the function of a protein is the knowledge of its spatial distribution within the cell, known as subcellular location.

A cell is a complete functional biological unit with many different internal structures named organelles that perform various functions important to the cell's survival. Cell phenotype image classification is a bio-imaging problem where an image representing a region of a cell containing a protein is given, and such image should be classified to its correct class (i.e. mitochondrion, Golgi complex, and so on).

According to a recent survey (Peng, 2008), the automated cell phenotype image classification problem can be categorized in the area of bioinformatics that is denoted as "bio-image informatics" and that comprises all the bioinformatics applications/problems which require various image data analysis and informatics techniques to extract, compare, search and manage the biological knowledge hidden inside the images. Due to the great complexity and information content in bioimages, such as the very high density of cells and their high resolution, the methods for automated cell phenotype image classification cannot simply reuse existing image analysis methods from the medical field, but requires the study of novel ad hoc techniques to analyze its complicated image objects.

Classifying protein subcellular patterns is important since the knowledge of the subcellular location of a protein is useful to understand its specific function and to describe the cell behavior under different conditions (Boland & Murphy, 2001). Moreover, the knowledge of the location of all proteins is crucial for several practical computational biology applications, ranging from early diagnosis of a disease to the design of high-throughput screening systems for drug discovery and monitoring of its therapeutic effectiveness, and to conduct experiments to determine the effect of various treatments on the synthesis and behavior of proteins within a cell.

The protein location changes are also an important factor in cancer. For example in (Glory et al., 2008) a new automated approach to identify cancer biomarkers is presented, which involves automated learning methods to compare subcellular location patterns between normal and cancerous tissues. The proteins whose locations were observed to change between cancerous and normal tissue are all candidates for potential biomarkers that could be used to diagnose or monitor cancer.

This chapter focuses on the study of machine learning techniques for cell phenotype image classification. Machine learning is a subset of the Pattern Recognition techniques where the parameters of a given approach (e.g. the parameters of an enhancement algorithm or of a neural network) are obtained analyzing a given dataset. Duin et al. (2002) have defined Pattern Recognition as: "the scientific discipline that studies theories and methods for designing machines that are able to recognize patterns in noisy data... Pattern recognition has an engineering nature as its final goal is the design of machines". The main reason why machine learning methods should be used in this problem is that they can be employed to extract hidden relationships and correlations among the data.

In this chapter, after a short introduction concerning the basic problem definition and notation, a detailed review of the existing literature about automated methods for cell phenotype image classification is provided, by considering the most salient approaches proposed in the literature. A generic architecture of a Pattern Recognition System may be defined in five steps:

- Data Collection (input is obtained by a sensor, etc);
- Pre-Processing (noise filtering, illumination normalization, etc);
- Feature Extraction (identify the characteristics of the patterns);
- Classification (classify the data);

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/ensemble-neural-networks-automated-cell/56175](http://www.igi-global.com/chapter/ensemble-neural-networks-automated-cell/56175)

## Related Content

---

### Image Analysis and Understanding Based on Information Theoretical Region Merging Approaches for Segmentation and Cooperative Fusion

Felipe Calderero and Ferran Marqués (2013). *Handbook of Research on Computational Intelligence for Engineering, Science, and Business* (pp. 75-121).

[www.irma-international.org/chapter/image-analysis-understanding-based-information/72490](http://www.irma-international.org/chapter/image-analysis-understanding-based-information/72490)

### History of Water Management: Historical Perspectives and Technological Dynamics

Taoufik Ait Bouchgour, Khalid Lahyani and Ghizlane Moukhliiss (2026). *Computational Intelligence and Optimization Methods for Sustainable Water Management* (pp. 1-38).

[www.irma-international.org/chapter/history-of-water-management/392458](http://www.irma-international.org/chapter/history-of-water-management/392458)

### Mankind at a Crossroads: The Future of Our Relation With AI Entities

Neantro Saavedra-Rivano (2020). *International Journal of Software Science and Computational Intelligence* (pp. 28-37).

[www.irma-international.org/article/mankind-at-a-crossroads/258864](http://www.irma-international.org/article/mankind-at-a-crossroads/258864)

### Cognitive Memory: Human Like Memory

Bernard Carlos Widrow and Juan Aragon (2010). *International Journal of Software Science and Computational Intelligence* (pp. 1-15).

[www.irma-international.org/article/cognitive-memory-human-like-memory/49128](http://www.irma-international.org/article/cognitive-memory-human-like-memory/49128)

### Image Classification Using CNN With Multi-Core and Many-Core Architecture

Debajit Datta and Saira Banu Jamal Mohammed (2021). *Applications of Artificial Intelligence for Smart Technology* (pp. 233-266).

[www.irma-international.org/chapter/image-classification-using-cnn-with-multi-core-and-many-core-architecture/265589](http://www.irma-international.org/chapter/image-classification-using-cnn-with-multi-core-and-many-core-architecture/265589)