1

INTRODUCTION TO PATTERN RECOGNITION AND DATA MINING

1.1 INTRODUCTION

Pattern recognition is an activity that human beings normally excel in. The task of pattern recognition is encountered in a wide range of human activity. In a broader perspective, the term could cover any context in which some decision or forecast is made on the basis of currently available information. Mathematically, the problem of pattern recognition deals with the construction of a procedure to be applied to a set of inputs; the procedure assigns each new input to one of a set of classes on the basis of observed features. The construction of such a procedure on an input data set is defined as pattern recognition.

A pattern typically comprises some features or essential information specific to a pattern or a class of patterns. Pattern recognition, as per the convention, is the study of how machines can observe the environment, learn to distinguish patterns of interest from their background, and make sound and reasonable decisions about the categories of the patterns. In other words, the discipline of pattern recognition essentially deals with the problem of developing algorithms and methodologies that can enable the computer implementation of many recognition tasks that humans normally perform. The objective is to perform these tasks more accurately, faster, and perhaps more economically than humans and, in many cases, to release them from drudgery resulting from performing routine recognition tasks repetitively and mechanically. The scope of pattern recognition also encompasses tasks at which humans are not good, such as reading bar codes. Hence, the goal

Rough-Fuzzy Pattern Recognition: Applications in Bioinformatics and Medical Imaging,
First Edition. Pradipta Maji and Sankar K. Pal.

of pattern recognition research is to devise ways and means of automating certain decision-making processes that lead to classification and recognition.

Pattern recognition can be viewed as a twofold task, consisting of learning the invariant and common properties of a set of samples characterizing a class, and of deciding that a new sample is a possible member of the class by noting that it has properties common to those of the set of samples. The task of pattern recognition can be described as a transformation from the measurement space $\mathcal{M}$ to the feature space $\mathcal{F}$ and finally to the decision space $\mathcal{D}$; that is,

$$\mathcal{M} \to \mathcal{F} \to \mathcal{D}, \tag{1.1}$$

where the mapping $\delta : \mathcal{F} \to \mathcal{D}$ is the decision function, and the elements $d \in \mathcal{D}$ are termed as *decisions*.

Pattern recognition has been a thriving field of research for the past few decades [1–8]. The seminal article by Kanal [9] gives a comprehensive review of the advances made in the field until the early 1970s. More recently, a review article by Jain et al. [10] provides an engrossing survey of the advances made in statistical pattern recognition till the end of the twentieth century. Although the subject has attained a very mature level during the past four decades or so, it remains green to the researchers because of continuous cross-fertilization of ideas from disciplines such as computer science, physics, neurobiology, psychology, engineering, statistics, mathematics, and cognitive science. Depending on the practical need and demand, various modern methodologies have come into being, which often supplement the classical techniques [11].

In recent years, the rapid advances made in computer technology have ensured that large sections of the world population have been able to gain easy access to computers on account of the falling costs worldwide, and their use is now commonplace in all walks of life. Government agencies and scientific, business, and commercial organizations routinely use computers, not only for computational purposes but also for storage, in massive databases, of the immense volumes of data that they routinely generate or require from other sources. Large-scale computer networking has ensured that such data has become accessible to more and more people. In other words, we are in the midst of an information explosion, and there is an urgent need for methodologies that will help us to bring some semblance of order into the phenomenal volumes of data that can readily be accessed by us with a few clicks of the keys of our computer keyboard. Traditional statistical data summarization and database management techniques are just not adequate for handling data on this scale and for intelligently extracting information, or rather, knowledge that may be useful for exploring the domain in question or the phenomena responsible for the data, and providing support to decision-making processes. This quest has thrown up a new phrase, called *data mining* [12–14].

The massive databases are generally characterized by the numeric as well as textual, symbolic, pictorial, and aural data. They may contain redundancy, errors, imprecision, and so on. Data mining is aimed at discovering natural structures within such massive and often heterogeneous data. It is visualized as being capable of knowledge discovery using generalizations and magnifications of existing and new pattern recognition algorithms. Therefore, pattern recognition plays a significant role in the data mining process. Data mining deals with the process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. Hence, it can be viewed as applying pattern recognition and machine learning principles in the context of voluminous, possibly heterogeneous data sets [11].

One of the important problems in real-life data analysis is uncertainty management. Some of the sources of this uncertainty include incompleteness and vagueness in class definitions. In this background, the possibility concept introduced by the fuzzy sets theory [15] and rough sets theory [16] have gained popularity in modeling and propagating uncertainty. Both the fuzzy sets and rough sets provide a mathematical framework to capture uncertainties associated with the data [17]. They are complementary in some aspects. The generalized theories of rough-fuzzy sets and fuzzy-rough sets have been applied successfully to feature selection of real-valued data [18, 19], classification [20], image processing [21], data mining [22], information retrieval [23], fuzzy decision rule extraction, and rough-fuzzy clustering [24, 25].

The objective of this book is to provide some results of investigations, both theoretical and experimental, addressing the relevance of rough-fuzzy approaches to pattern recognition with real-life applications. Various methodologies are presented, integrating fuzzy logic and rough sets for clustering, classification, and feature selection. The emphasis of these methodologies is given on (i) handling data sets which are large, both in size and dimension, and involve classes that are overlapping, intractable and/or having nonlinear boundaries; (ii) demonstrating the significance of rough-fuzzy granular computing in soft computing paradigm for dealing with the knowledge discovery aspect; and (iii) demonstrating their success in certain tasks of bioinformatics and medical imaging as an example. Before describing the scope of the book, a brief review of pattern recognition, data mining, and application of pattern recognition algorithms in data mining problems is provided.

The structure of the rest of this chapter is as follows: Section 1.2 briefly presents a description of the basic concept, features, and techniques of pattern recognition. In Section 1.3, the data mining aspect is elaborated, discussing its components, tasks involved, approaches, and application areas. The pattern recognition perspective of data mining is introduced next and related research challenges are mentioned. The role of soft computing in pattern recognition and data mining is described in Section 1.4. Finally, Section 1.5 discusses the scope and organization of the book.

1.2 PATTERN RECOGNITION

A typical pattern recognition system consists of three phases, namely, *data acquisition*, *feature selection or extraction*, and *classification or clustering*. In the data acquisition phase, depending on the environment within which the objects are to be classified or clustered, data are gathered using a set of sensors. These are then passed on to the feature selection or extraction phase, where the dimensionality of the data is reduced by retaining or measuring only some characteristic features or properties. In a broader perspective, this stage significantly influences the entire recognition process. Finally, in the classification or clustering phase, the selected or extracted features are passed on to the classification or clustering system that evaluates the incoming information and makes a final decision. This phase basically establishes a transformation between the features and the classes or clusters [1, 2, 8].

1.2.1 Data Acquisition

In data acquisition phase, data are gathered via a set of sensors depending on the environment within which the objects are to be classified. Pattern recognition techniques are applicable in a wide domain, where the data may be qualitative, quantitative, or both; they may be numerical, linguistic, pictorial, or any combination thereof. Generally, the data structures that are used in pattern recognition systems are of two types: object data vectors and relational data. Object data, sets of numerical vectors of m features, are represented as $X = \{x_1, \ldots, x_i, \ldots, x_n\}$, a set of n feature vectors in the m-dimensional measurement space $\Re^m$. The ith object observed in the process has vector x_i as its numerical representation; x_{ij} is the jth ($j = 1, \ldots, m$) feature associated with the ith object. On the other hand, relational data are a set of n^2 numerical relationships, say r_{ij}, between pairs of objects. In other words, r_{ij} represents the extent to which objects x_i and x_j are related in the sense of some binary relationship ρ. If the objects that are pairwise related by ρ are called $O = \{o_1, \ldots, o_i, \ldots, o_n\}$, then $\rho : O \times O \rightarrow \Re$.

1.2.2 Feature Selection

Feature selection or extraction is a process of selecting a map by which a sample in an m-dimensional measurement space is transformed into a point in a d-dimensional feature space, where $d < m$ [1, 8]. Mathematically, it finds a mapping of the form $y = f(x)$, by which a sample $x = [x_1, \ldots, x_j, \ldots, x_m]$ in an m-dimensional measurement space $\mathcal{M}$ is transformed into an object $y = [y_1, \ldots, y_j, \ldots, y_d]$ in a d-dimensional feature space $\mathcal{F}$.

The main objective of this task is to retain or generate the optimum salient characteristics necessary for the recognition process and to reduce the dimensionality of the measurement space so that effective and easily computable algorithms can be devised for efficient classification. The problem of feature selection or extraction has two aspects, namely, formulating a suitable criterion to evaluate the goodness of a feature set and searching the optimal set in terms of the criterion. In general, those features are considered to have optimal saliencies for which interclass (respectively, intraclass) distances are maximized (respectively, minimized). The criterion for a good feature is that it should be unchanging with any other possible variation within a class, while emphasizing differences that are important in discriminating between patterns of different types.

The major mathematical measures so far devised for the estimation of feature quality are mostly statistical in nature, and can be broadly classified into two categories, namely, feature selection in the measurement space and feature selection in a transformed space. The techniques in the first category generally reduce the dimensionality of the measurement space by discarding redundant or least information-carrying features. On the other hand, those in the second category utilize all the information contained in the measurement space to obtain a new transformed space, thereby mapping a higher dimensional pattern to a lower dimensional one. This is referred to as *feature extraction* [1, 2, 8].

1.2.3 Classification and Clustering

The problem of classification and clustering is basically one of partitioning the feature space into regions, one region for each category of input. Hence, it attempts to assign every data object in the entire feature space to one of the possible classes or clusters. In real life, the complete description of the classes is not known. Instead, a finite and usually smaller number of samples are available, which often provide partial information for optimal design of feature selector or extractor or classification or clustering system. Under such circumstances, it is assumed that these samples are representative of the classes or clusters. Such a set of typical patterns is called a *training set*. On the basis of the information gathered from the samples in the training set, the pattern recognition systems are designed. That is, the values of the parameters of various pattern recognition methods are decided.

Design of a classification or clustering scheme can be made with labeled or unlabeled data. When the algorithm is given a set of objects with known classifications, that is, labels, and is asked to classify an unknown object based on the information acquired by it during training, the design scheme is called *supervised learning*; otherwise it is *unsupervised learning*. Supervised learning is used for classifying different objects, while clustering is performed through unsupervised learning. Through cluster analysis, a given data set is divided into a set of clusters in such a way that two objects from the same cluster are as similar as possible and the objects from different clusters are as dissimilar as possible. In effect, it tries to mimic the human ability to group similar objects into classes and categories. A number of clustering algorithms have been proposed to suit different requirements [2, 26, 27].

Pattern classification or clustering, by its nature, admits many approaches, sometimes complementary, sometimes competing, to provide the solution to a given problem. These include decision theoretic approach (both deterministic and probabilistic), syntactic approach, connectionist approach, fuzzy and rough set theoretic approaches and hybrid or soft computing approach. Let $\beta = \{\beta_1, \ldots, \beta_i, \ldots, \beta_c\}$ represent the c possible classes or clusters in a d-dimensional feature space $\mathcal{F}$, and $y = [y_1, \ldots, y_j, \ldots, y_d]$ be an unknown pattern vector whose class is to be identified. In deterministic classification or clustering approach, the object is assigned to only one unambiguous pattern class or cluster β_i if the decision function D_i associated with the class β_i satisfies the following relation:

$$D_i(y) > D_j(y), \qquad j = 1, \ldots, c, \text{ and } j \neq i. \tag{1.2}$$

In the decision theoretic approach, once a pattern is transformed through feature evaluation to a vector in the feature space, its characteristics are expressed only by a set of numerical values. Classification can be done by using deterministic or probabilistic techniques [1, 2, 8]. The nearest neighbor classifier [2] is an example of deterministic classification approach, where it is assumed that there exists only one unambiguous pattern class corresponding to each of the unknown pattern vectors. In most of the practical problems, the features are often noisy and the classes in the feature space are overlapping. In order to model such systems, the features are considered as random variables in the probabilistic approach. The most commonly used classifier in such probabilistic systems is the Bayes maximum likelihood classifier [2].

When a pattern is rich in structural information such as picture recognition, character recognition, scene analysis, that is, the structural information plays an important role in describing and recognizing the patterns, it is convenient to use the syntactic approach [3]. It deals with the representation of structures via sentences, grammars, and automata. In the syntactic method [3], the ability of selecting and classifying the simple pattern primitives and their relationships represented by the composition operations is the vital criterion for making a system effective. Since the techniques of composition of primitives into patterns are usually governed by the formal language theory, the approach is often referred to as a *linguistic approach*. An introduction to a variety of approaches based on this idea can be found in Fu [3]. Other approaches to pattern recognition are discussed in Section 1.4 under soft computing methods.

1.3 DATA MINING

Data mining involves fitting models to or determining patterns from observed data. The fitted models play the role of inferred knowledge. Typically, a data mining algorithm constitutes some combination of three components, namely, model, preference criterion, and search algorithm [13].

The model represents its function (e.g., classification, clustering) and its representational form (e.g., linear discriminants, neural networks). A model contains parameters that are to be determined from the data. The preference criterion is a basis to decide the preference of one model or a set of parameters over another, depending on the given data. The criterion is usually some form of goodness of fit function of the model to the data, perhaps tempered by a smoothing term to avoid overfitting, or generating a model with too many degrees of freedom to be constrained by the given data. On the other hand, the search algorithm is the specification of an algorithm for finding particular models and parameters, given the data, models, and a preference criterion [5].

1.3.1 Tasks, Tools, and Applications

The common tasks or functions in current data mining practice include association rule discovery, clustering, classification, sequence analysis, regression, summarization, and dependency modeling.

The association rule discovery describes association relationship among different attributes. The origin of association rules is in market basket analysis. A market basket is a collection of items purchased by a customer in an individual customer transaction. One common analysis task in a transaction database is to find sets of items or itemsets that frequently appear together. Each pattern extracted through the analysis consists of an itemset and its support, that is, the number of transactions that contain it. Knowledge of these patterns can be used to improve placement of items in a store or for mail-order marketing. The huge size of transaction databases and the exponential increase in the number of potential frequent itemsets with increase in the number of attributes or items make the above problem a challenging one. The a priori algorithm [28] provides an early solution, which is improved by subsequent algorithms using partitioning, hashing, sampling, and dynamic itemset counting.

The clustering technique maps a data item into one of several clusters, where clusters are natural groupings of data items based on similarity metrics or probability density models. Clustering is used in several exploratory data analysis tasks, customer retention and management, and web mining. The clustering problem has been studied in many fields, including statistics, machine learning, and pattern recognition. However, large data considerations were absent in these approaches. To address those issues, several new algorithms with greater emphasis on scalability have been developed in the framework of data mining, including those based on summarized cluster representation called *cluster feature* [29], *sampling* [30], and *density joins* [31].

On the other hand, the classification algorithm classifies a data item into one of several predefined categorical classes. It is used for the purpose of predictive data mining in several fields such as scientific discovery, fraud detection, atmospheric data mining, and financial engineering. Several classification methodologies have been mentioned earlier in Section 1.2.3. Some typical algorithms suitable for large databases are based on Bayesian techniques [32] and decision trees [33, 34].

Sequence analysis [35] models sequential patterns such as time series data. The goal is to model the process of generating the sequence or to extract and report deviation and trends over time. The framework is increasingly gaining importance because of its application in bioinformatics and streaming data analysis. The regression [13, 36] technique maps a data item to a real-valued prediction variable. It is used in different prediction and modeling applications.

The summarization [13] procedure provides a compact description for a subset of data. A simple example would be mean and standard deviation for all fields. More sophisticated functions involve summary rules, multivariate visualization techniques, and functional relationship between variables. Summarization functions are often used in interactive data analysis, automated report generation, and text mining. On the other hand, the dependency modeling [37] describes significant dependencies among variables. Some other tasks required in some data mining applications are outlier or anomaly detection, link analysis, optimization, and planning.

A wide variety and number of data mining algorithms are described in the literature, from the fields of statistics, pattern recognition, machine learning, and databases. They represent a long list of seemingly unrelated and often highly specific algorithms. Some representative groups are statistical models [2, 14], probabilistic graphical dependency models [38], decision trees and rules [39], inductive-logic-programming-based models, example-based methods [40, 41], neural-network-based models [42, 43], fuzzy set theoretic models [12, 44, 45], rough set theory-based models [46–48], genetic-algorithm-based models [49], and hybrid and soft computing models [50].

Data mining algorithms determine both the flexibility of the model in representing the data and the interpretability of the model in human terms. Typically, the more complex models may fit the data better but may also be more difficult to understand and to fit reliably. Also, each representation suits some problems better than others. For example, decision tree classifiers can be very useful for finding structure in high dimensional spaces and are also useful in problems with mixed continuous and categorical data. However, they may not be suitable for problems where the true decision boundaries are nonlinear multivariate functions.

A wide range of organizations including business companies, scientific laboratories, and governmental departments have deployed successful applications of data mining. Although early adopters of this technology have tended to be in information-intensive industries such as financial services and direct mail marketing, the technology is applicable to any company looking to leverage a large data warehouse to better manage their operations. Two critical factors for success with data mining are a large, well-integrated data warehouse and a well-defined understanding of the process within which data mining is to be applied. Several domains where large volumes of data are stored in centralized or distributed databases include financial investment, hospital management systems, manufacturing and production, telecommunication network, astronomical object detection, genomic and biological data mining, and information retrieval [5].

1.3.2 Pattern Recognition Perspective

At present, pattern recognition and machine learning provide the most fruitful framework for data mining [5, 51, 52]. They provide a wide range of linear and nonlinear, comprehensible and complex, predictive and descriptive, instance and rule-based models for different data mining tasks such as clustering, classification, and rule discovery. Also, the methods for modeling probabilistic and fuzzy uncertainties in the discovered patterns form a part of pattern recognition research. Another aspect that makes pattern recognition algorithms attractive for data mining is their capability of learning or induction. As opposed to many statistical techniques that require the user to have a hypothesis in mind first, pattern recognition algorithms automatically analyze the data and identify relationships among attributes and entities in the data to build models that allow domain experts to understand the relationship between the attributes and the class. Several data preprocessing tasks such as instance selection, data cleaning, dimensionality reduction, and handling missing data are also extensively studied in pattern recognition framework. Besides these, other data mining issues addressed by pattern recognition methodologies include handling of relational, sequential, and symbolic data (syntactic pattern recognition; pattern recognition in arbitrary metric spaces); human interaction (knowledge encoding and extraction); knowledge evaluation (description length principle); and visualization.

Pattern recognition is at the core of data mining systems. However, pattern recognition and data mining are not equivalent considering their original definitions. There exists a gap between the requirements of a data mining system and the goals achieved by present-day pattern recognition algorithms. Development of new generation pattern recognition algorithms is expected to encompass more massive data sets involving diverse sources and types of data that will support mixed initiative data mining, where human experts collaborate with the computer to form hypotheses and test them.

1.4 RELEVANCE OF SOFT COMPUTING

A good pattern recognition system should possess several characteristics. These are online adaptation to cope with the changes in the environment, handling nonlinear class separability to tackle real-life problems, handling of overlapping classes or clusters for discriminating almost similar but different objects, real time processing for making a decision in a reasonable time, generation of soft and hard decisions to make the system flexible, verification and validation mechanisms for evaluating its performance, and minimizing the number of parameters in the system that have to be tuned for reducing the cost and complexity. Moreover, the system should be made artificially intelligent in order to emulate some aspects of the human processing system. Connectionist or artificial neural-network-based approaches to pattern recognition are attempts to achieve some of these goals because of their major characteristics such as adaptivity, robustness or ruggedness, speed, and optimality [53–57]. They are also suitable in data-rich environments and are typically used for extracting embedded knowledge in the form of rules, quantitative evaluation of these rules, clustering, self-organization, classification, and regression. They have an advantage, over other types of machine learning algorithms, for scaling [58, 59].

The fuzzy set theoretic classification approach is developed on the basis of the realization that a pattern may belong to more than one class, with varying degrees of class membership. Accordingly, fuzzy decision theoretic, fuzzy syntactic, fuzzy neural approaches are developed [4, 6, 60, 61]. These approaches can handle uncertainties, arising from vague, incomplete, linguistic, and overlapping patterns at various stages of pattern recognition systems [4, 15, 60, 62].

The theory of rough sets [16, 63, 64] has emerged as another major mathematical approach for managing uncertainty that arises from inexact, noisy, or incomplete information. It is turning out to be methodologically significant to the domains of artificial intelligence and cognitive sciences, especially in the representation of and reasoning with vague and/or imprecise knowledge, data classification, data analysis, machine learning, and knowledge discovery [48, 64–66]. This approach is relatively new when compared to connectionist and fuzzy set theoretic approaches.

Investigations have also been made in the area of pattern recognition using genetic algorithms [67, 68]. Similar to neural networks, genetic algorithms [69] are also based on powerful metaphors from the natural world. They mimic some of the processes observed in natural evolution, which include crossover, selection, and mutation, leading to a stepwise optimization of organisms.

There have been several attempts over the past two decades to evolve new approaches to pattern recognition and to derive their hybrids by judiciously combining the merits of several techniques [6, 70] involving mainly fuzzy logic, artificial neural networks, genetic algorithms, and rough set theory, for developing an efficient new paradigm called *soft computing* [71]. Here integration is done in a cooperative, rather than a competitive, manner. The result is a more intelligent and robust system providing a human interpretable, low cost, approximate solution, as compared to traditional techniques. Neuro-fuzzy approach is perhaps the most visible hybrid paradigm [6, 61, 72–75] realized so far in soft computing framework. Besides the generic advantages, neuro-fuzzy approach provides the corresponding application-specific merits [76–82]. Rough-fuzzy [63, 83] and neuro-rough [84–86] hybridizations are also proving to be fruitful frameworks for modeling human perceptions and providing means for computing with words. Rough-fuzzy computing provides a powerful mathematical framework to capture uncertainties associated with the data. Its relevance in modeling the fuzzy granulation (f-granulation) characteristics of the computational theory of perceptions may also be mentioned in this regard [87–89]. Other hybridized models for pattern recognition and data mining include neuro-genetic [90–94], rough-genetic [95–97], fuzzy-genetic [98–103], rough-neuro-genetic [104], rough-neuro-fuzzy [105–109], and neuro-fuzzy-genetic [110–115] approaches.

1.5 SCOPE AND ORGANIZATION OF THE BOOK

This book has nine chapters describing various theories, methodologies, and algorithms, along with extensive experimental results, addressing certain pattern recognition and mining tasks in rough-fuzzy computing paradigm with real-life applications. Various methodologies are described using soft computing approaches, judiciously integrating fuzzy logic and rough sets for clustering, classification, and feature selection. The emphasis is placed on the use of the methodologies for handling both object and relational data sets that are large both in size and dimension, and involve classes that are overlapping, intractable and/or having nonlinear boundaries. The effectiveness of the algorithms is demonstrated on different real-life data sets taken from varied domains such as remote sensing, medical imagery, speech recognition, protein sequence encoding and gene expression analysis with special emphasis on problems in medical imaging and mining patterns in bioinformatics. The superiority of the rough-fuzzy models presented in this book over several related ones is found to be statistically significant.

The basic notions and characteristics of two soft computing tools, namely, fuzzy sets and rough sets are briefly presented in Chapter 2. These are followed by the concept of information granules, f-granulations, emergence of rough-fuzzy computing paradigm, and their relevance to pattern recognition. It also provides a mathematical framework for generalized rough sets incorporating the concept of fuzziness in defining the granules as well as the set. Various roughness and uncertainty measures with properties are introduced. Different research issues related to rough granules are stated.

A generalized hybrid unsupervised learning algorithm, termed as *rough-fuzzy-possibilistic c-means*, is reported in Chapter 3. It comprises a judicious integration of the principles of rough sets and fuzzy sets. Although the concept of lower and upper approximations of rough sets deals with uncertainty, vagueness, and incompleteness in class definition, the membership function of fuzzy sets enables efficient handling of overlapping partitions. It incorporates both probabilistic and possibilistic memberships simultaneously to avoid the problems of noise sensitivity of fuzzy c-means and the coincident clusters of possibilistic c-means. The concept of crisp lower bound and fuzzy boundary of a class, introduced in rough-fuzzy-possibilistic c-means, enables efficient selection of cluster prototypes. The algorithm is generalized in the sense that all the existing variants of c-means algorithms can be derived from this algorithm as a special case. Several quantitative indices are described on the basis of rough sets for evaluating the performance of different c-means algorithms on real-life data sets.

A rough-fuzzy model for pattern classification based on granular computing is described in Chapter 4. In this model, the formulation of class-dependent granules in fuzzy environment is introduced. Fuzzy membership functions are used to represent the feature-wise belonging to different classes, thereby producing fuzzy granulation of the feature space. The fuzzy granules thus generated possess better class discriminatory information that is useful in pattern classification with overlapping classes. The neighborhood rough sets are used in the selection of a subset of granulated features that explore the local or contextual information from neighborhood granules. The model thus explores the mutual advantages of class-dependent fuzzy granulation and neighborhood rough sets. The superiority of this model over other similar methods is established with seven completely labeled data sets, including a synthetic remote sensing image, and two partially labeled real remote sensing images collected from satellites. Various performance measures, including a method of dispersion estimation, are used for comparative analysis. The dispersion score quantifies the nature of distribution of the classified patterns among different classes so that lower the dispersion, better the classifier. The rough-fuzzy granular space-based model is able to learn well even with a lower percentage of training set that makes the system faster. The model is seen to have the lowest dispersion measure (i.e., misclassified patterns are confined to minimum number of classes) compared to others, thereby reflecting well the overlapping characteristics of a class with others, and providing a strong clue for the class-wise performance improvement with available higher level information. The statistical significance of this model is also supported by the χ^2 test.

The selection of nonredundant and relevant features of real-valued data sets is a highly challenging problem. Chapter 5 deals with a feature selection method based on fuzzy-rough sets by maximizing the relevance and minimizing the redundancy of the selected features. By introducing the concept of fuzzy equivalence partition matrix, a new representation of Shannon's entropy for fuzzy approximation spaces is presented to measure the relevance and redundancy of features suitable for real-valued data sets. The fuzzy equivalence partition matrix is based on the theory of fuzzy-rough sets, where each row of the matrix represents a fuzzy equivalence partition that can be automatically derived from the given data set. The fuzzy equivalence partition matrix also offers an efficient way to calculate many more information measures, termed as *f-information measures*. Several f-information measures are shown to be effective for selecting nonredundant and relevant features of real-valued data sets. The experimental study also includes a comparison of the performance of different f-information measures for feature selection in fuzzy approximation spaces. Several quantitative indices are described on the basis of fuzzy-rough sets for evaluating the performance of different methods.

In pattern recognition, there are mainly two types of data: object and relational data. The former is the most common type of data and is in the form of the usual data set of feature vectors. On the other hand, the latter is less common and consists of the pairwise relations such as similarities or dissimilarities between each pair of implicit objects. Such a relation is usually stored in a relation matrix and no other knowledge is available about the objects being clustered. As the relational data is less common than object data, relational pattern recognition methods are not as well developed as their object counterparts, particularly in the area of robust clustering. However, relational methods are becoming a necessity as relational data becomes more and more common. For instance, information retrieval, data mining, web mining, and bioinformatics are all applications which could greatly benefit from pattern recognition methods that can deal with relational data. In this regard, the next chapter discusses a rough-fuzzy relational clustering algorithm, termed as *rough-fuzzy c-medoids algorithm*, and demonstrates its effectiveness in amino acid sequence analysis.

Although several experimental results on both artificial and real-life data sets, including speech and remotely sensed multispectral image data, are provided in Chapters 3, 4, and 5 to demonstrate the effectiveness of the respective rough-fuzzy methodologies, the next four chapters are concerned only with certain specific applications in bioinformatics and medical imaging. Problems considered include selection of a minimum set of basis strings with maximum information for amino acid sequence analysis (Chapter 6), grouping functionally similar genes from microarray gene expression data through clustering (Chapter 7), selection of relevant genes from high dimensional microarray gene expression data (Chapter 8), and segmentation of brain magnetic resonance (MR) images using clustering (Chapter 9).

In most pattern recognition algorithms, biological molecules such as amino acids cannot be used directly as inputs as they are nonnumerical variables. They, therefore, need encoding before being used as input. In this regard, bio-basis function maps a nonnumerical sequence space to a numerical feature space. It is designed using an amino acid mutation matrix. One of the important issues for the bio-basis function is how to select a minimum set of bio-basis strings with maximum information. In Chapter 6, the rough-fuzzy c-medoids algorithm is used to select most informative bio-basis strings. It comprises a judicious integration of the principles of rough sets, fuzzy sets, c-medoids algorithm, and amino acid mutation matrix. The concept of crisp lower bound and fuzzy boundary of a cluster, introduced in rough-fuzzy c-medoids, enables efficient selection of a minimum set of most informative bio-basis strings. Several indices are stated for evaluating quantitatively the quality of selected bio-basis strings.

Microarray technology is one of the important biotechnological means that allows recording the expression levels of thousands of genes during important biological processes and across collections of related samples. An important application of microarray data is to elucidate the patterns hidden in gene expression data for an enhanced understanding of functional genomics. However, the large number of genes and the complexity of biological networks greatly increase the challenges of comprehending and interpreting the resulting mass of data. A first step toward addressing this challenge is the use of clustering techniques. In this regard, different rough-fuzzy clustering algorithms are used in Chapter 7 to cluster functionally similar genes from microarray data sets. The effectiveness of these algorithms, along with a comparison with other related gene clustering algorithms, is demonstrated on a set of microarray gene expression data sets using some standard validity indices.

Several information measures such as entropy, mutual information, and f-information have been shown to be successful for selecting a set of relevant and nonredundant genes from high dimensional microarray data set. However, for continuous gene expression values, it is very difficult to find the true density functions and to perform the integrations required to compute different information measures. In this regard, the concept of fuzzy equivalence partition matrix, explained in Chapter 5, is used in Chapter 8 to approximate the true marginal and joint distributions of continuous gene expression values.

The performance of this methodology in selecting relevant and nonredundant continuous valued genes from microarray data is compared with that of existing ones using the class separability index and predictive accuracy of support vector machine.

Image segmentation is an indispensable process in the visualization of human tissues, particularly during clinical analysis of MR images. In Chapter 9, different rough-fuzzy clustering algorithms are used for the segmentation of brain MR images. One of the major issues of the rough-fuzzy clustering-algorithm-based brain MR image segmentation is how to select initial prototypes of different classes or categories. The concept of discriminant analysis, based on the maximization of class separability, is used to circumvent the initialization and local minima problems of the rough-fuzzy clustering algorithms. Some quantitative indices are described to extract local features of brain MR images, when applied on a set of synthetic and real brain MR images, for segmentation.

REFERENCES

1. P. A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach*. Prentice Hall, Englewood Cliffs, NJ, 1982.
2. R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification and Scene Analysis*. John Wiley & Sons, New York, 1999.
3. K. S. Fu. *Syntactic Pattern Recognition and Applications*. Academic Press, London, 1982.
4. S. K. Pal and D. D. Majumder. *Fuzzy Mathematical Approach to Pattern Recognition*. John Wiley (Halsted Press), New York, 1986.
5. S. K. Pal and P. Mitra. *Pattern Recognition Algorithms for Data Mining*. CRC Press, Boca Raton, FL, 2004.
6. S. K. Pal and S. Mitra. *Neuro-Fuzzy Pattern Recognition: Methods in Soft Computing*. New York: John Wiley & Sons, 1999.
7. S. K. Pal and A. Pal, editors. *Pattern Recognition: from Classical to Modern Approaches*. World Scientific, Singapore, 2001.
8. J. T. Tou and R. C. Gonzalez. *Pattern Recognition Principles*. Addison-Wesley, Reading, MA, 1974.
9. L. Kanal. Patterns in Pattern Recognition. *IEEE Transactions on Information Theory*, 20:697–722, 1974.
10. A. K. Jain, R. P. W. Duin, and J. Mao. Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:4–37, 2000.
11. A. Pal and S. K. Pal. Pattern Recognition: Evolution of Methodologies and Data Mining. In S. K. Pal and A. Pal, editors, *Pattern Recognition: from Classical to Modern Approaches*, pages 1–23. World Scientific, Singapore, 2001.
12. K. Cios, W. Pedrycz, and R. Swiniarski. *Data Mining Methods for Knowledge Discovery*. Kluwer Academic Publishers, Boston, MA, 1998.
13. U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors. *Advances in Knowledge Discovery and Data Mining*. MIT Press, Cambridge, MA, 1996.

14. T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, New York, 2001.
15. L. A. Zadeh. Fuzzy Sets. *Information and Control*, 8:338–353, 1965.
16. Z. Pawlak. *Rough Sets: Theoretical Aspects of Reasoning about Data*. Kluwer, Dordrecht, The Netherlands, 1991.
17. D. Dubois and H. Prade. Rough Fuzzy Sets and Fuzzy Rough Sets. *International Journal of General Systems*, 17:191–209, 1990.
18. R. Jensen and Q. Shen. Fuzzy-Rough Attribute Reduction with Application to Web Categorization. *Fuzzy Sets and Systems*, 141:469–485, 2004.
19. R. Jensen and Q. Shen. Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough-Based Approach. *IEEE Transactions on Knowledge and Data Engineering*, 16(12): 1457–1471, 2004.
20. S. K. Pal, S. Meher, and S. Dutta. Class-Dependent Rough-Fuzzy Granular Space, Dispersion Index and Classification. *Pattern Recognition*, under revision.
21. D. Sen and S. K. Pal. Generalized Rough Sets, Entropy and Image Ambiguity Measures. *IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics*, 39(1): 117–128, 2009.
22. Y. F. Wang. Mining Stock Price Using Fuzzy Rough Set System. *Expert Systems with Applications*, 24(1): 13–23, 2003.
23. P. Srinivasan, M. E. Ruiz, D. H. Kraft, and J. Chen. Vocabulary Mining for Information Retrieval: Rough Sets and Fuzzy Sets. *Information Processing and Management*, 37(1): 15–38, 1998.
24. P. Maji and S. K. Pal. Rough-Fuzzy C-Medoids Algorithm and Selection of Bio-Basis for Amino Acid Sequence Analysis. *IEEE Transactions on Knowledge and Data Engineering*, 19(6): 859–872, 2007.
25. P. Maji and S. K. Pal. Rough Set Based Generalized Fuzzy C-Means Algorithm and Quantitative Indices. *IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics*, 37(6): 1529–1540, 2007.
26. A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs, NJ, 1988.
27. A. K. Jain, M. N. Murty, and P. J. Flynn. Data Clustering: A Review. *ACM Computing Surveys*, 31(3): 264–323, 1999.
28. R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and I. Verkamo. Fast Discovery of Association Rules. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthuruswamy, editors, *Advances in Knowledge Discovery and Data Mining*, pages 307–328. MIT Press, Cambridge, MA, 1996.
29. P. Bradley, U. M. Fayyad, and C. Reina. Scaling Clustering Algorithms to Large Databases. In *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, New York*, pages 9–15, AAAI Press, Menlo Park, CA, 1998.
30. S. Guha, R. Rastogi, and K. Shim. CURE: An Efficient Clustering Algorithm for Large Databases. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 73–84, ACM Press, New York, 1998.
31. J. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR*, pages 226–231, AAAI Press, Menlo Park, CA, 1996.

32. P. Cheeseman, J. Kelly, M. Self, and J. Stutz. Autoclass: A Bayesian Classification System. In *Proceedings of the 5th International Conference on Machine Learning, Ann Arbor, MI*, Morgan Kaufmann, San Mateo, CA, 1988.
33. J. Gehrke, R. Ramakrishnan, and V. Ganti. RainForest: A Framework for Large Decision Tree Construction for Large Datasets. In *Proceedings of the 24th International Conference on Very Large Databases, San Francisco, CA*, pages 416–427, Morgan Kaufmann, San Mateo, CA, 1998.
34. J. Shafer, R. Agrawal, and M. Mehta. SPRINT: A Scalable Parallel Classifier for Data Mining. In *Proceedings of the 22nd International Conference on Very Large Databases*, San Francisco, pages 544–555. Morgan Kaufmann, San Mateo, CA, 1996.
35. D. Gusfield. *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. Cambridge University Press, Cambridge, 1997.
36. P. Maji and S. Paul. Rough Sets for Selection of Molecular Descriptors to Predict Biological Activity of Molecules. *IEEE Transactions on Systems Man and Cybernetics Part C-Applications and Reviews*, 40(6): 639–648, 2010.
37. J. Hale and S. Shenoi. Analyzing FD Inference in Relational Databases. *Data and Knowledge Engineering*, 18:167–183, 1996.
38. F. V. Jensen. *Bayesian Networks and Decision Diagrams*. Springer-Verlag, New York, 2001.
39. L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Wadsworth and Brooks/Cole, Monterey, CA, 1984.
40. D. W. Aha, D. Kibler, and M. K. Albert. Instance-Based Learning Algorithms. *Machine Learning*, 6:37–66, 1991.
41. S. K. Pal and S. C. K. Shiu. *Foundations of Soft Case Based Reasoning*. John Wiley & Sons, New York, 2004.
42. M. Craven and J. Shavlik. Using Neural Networks for Data Mining. *Future Generation Computer Systems*, 13:211–219, 1997.
43. H. Lu, R. Setiono, and H. Liu. Effective Data Mining Using Neural Networks. *IEEE Transactions on Knowledge and Data Engineering*, 8(6): 957–961, 1996.
44. J. F. Baldwin. Knowledge from Data Using Fuzzy Methods. *Pattern Recognition Letters*, 17:593–600, 1996.
45. W. Pedrycz. Fuzzy Set Technology in Knowledge Discovery. *Fuzzy Sets and Systems*, 98:279–290, 1998.
46. J. Komorowski, Z. Pawlak, L. Polkowski, and A. Skowron. A Rough Set Perspective on Data and Knowledge. In W. Klosgen and J. Zytkow, editors, *The Handbook of Data Mining and Knowledge Discovery*. Oxford University Press, Oxford, 1999.
47. T. Y. Lin and N. Cercone, editors. *Rough Sets and Data Mining: Analysis of Imprecise Data*. Kluwer Academic Publications, Boston, MA, 1997.
48. L. Polkowski and A. Skowron, editors. *Rough Sets in Knowledge Discovery*, volumes 1 and 2. Physica-Verlag, Heidelberg, 1998.
49. I. W. Flockhart and N. J. Radcliffe. A Genetic Algorithm-Based Approach to Data Mining. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR*, page 299, AAAI Press, Menlo Park, CA, 1996.
50. S. Mitra, S. K. Pal, and P. Mitra. Data Mining in Soft Computing Framework: A Survey. *IEEE Transactions on Neural Networks*, 13(1): 3–14, 2002.

51. R. L. Kennedy, Y. Lee, B. van Roy, C. D. Reed, and R. P. Lippman. *Solving Data Mining Problems Through Pattern Recognition*. Prentice Hall, NJ, 1998.
52. T. Mitchell. Machine Learning and Data Mining. *Communications of the ACM*, 42(11): 30–36, 1999.
53. W. Bian and X. Xue. Subgradient-Based Neural Networks for Nonsmooth Nonconvex Optimization Problems. *IEEE Transactions on Neural Networks*, 20(6): 1024–1038, 2009.
54. H. Chen and X. Yao. Regularized Negative Correlation Learning for Neural Network Ensembles. *IEEE Transactions on Neural Networks*, 20(12): 1962–1979, 2009.
55. S. Haykin. *Neural Networks: A Comprehensive Foundation*, 2nd edition. Prentice Hall, New Jersey, 1998.
56. R. Lippmann. An Introduction to Computing with Neural Nets. *IEEE ASSP Magazine*, 4(22), 1987.
57. S. L. Phung and A. Bouzerdoum. A Pyramidal Neural Network for Visual Pattern Recognition. *IEEE Transactions on Neural Networks*, 18(2): 329–343, 2007.
58. Y. Bengio, J. M. Buhmann, M. Embrechts, and J. M. Zurada. Introduction to the Special Issue on Neural Networks for Data Mining and Knowledge Discovery. *IEEE Transactions on Neural Networks*, 11:545–549, 2000.
59. A. A. Frolov, D. Husek, and P. Y. Polyakov. Recurrent-Neural-Network-Based Boolean Factor Analysis and Its Application to Word Clustering. *IEEE Transactions on Neural Networks*, 20(7): 1073–1086, 2009.
60. J. C. Bezdek and S. K. Pal, editors. *Fuzzy Models for Pattern Recognition: Methods that Search for Structures in Data*. IEEE Press, New York, 1992.
61. H. Bunke and A. Kandel, editors. *Neuro-Fuzzy Pattern Recognition*. World Scientific, Singapore, 2001.
62. A. Kandel. *Fuzzy Techniques in Pattern Recognition*. Wiley Interscience, New York, 1982.
63. S. K. Pal and A. Skowron, editors. *Rough-Fuzzy Hybridization: A New Trend in Decision Making*. Springer-Verlag, Singapore, 1999.
64. A. Skowron and R. Swiniarski. Rough Sets in Pattern Recognition. In S. K. Pal and A. Pal, editors, *Pattern Recognition: From Classical to Modern Approaches*, pages 385–428. World Scientific, Singapore, 2001.
65. E. Orlowska, editor. *Incomplete Information: Rough Set Analysis*. Physica-Verlag, Heidelberg, 2010.
66. L. Polkowski. *Rough Sets*. Physica-Verlag, Heidelberg, 2002.
67. S. Bandyopadhyay and S. K. Pal. *Classification and Learning Using Genetic Algorithms: Applications in Bioinformatics and Web Intelligence*. Springer-Verlag, Hiedelberg, Germany, 2007.
68. S. K. Pal and P. P. Wang, editors. *Genetic Algorithms for Pattern Recognition*. CRC Press, Boca Raton, FL, 1996.
69. D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, MA, 1989.
70. S. K. Pal. Soft Computing Pattern Recognition: Principles, Integrations and Data Mining. In T. Terano, T. Nishida, A. Namatame, S. Tsumoto, Y. Ohswa, and T. Washio, editors, *Advances in Artificial Intelligence* , *Lecture Notes in Artificial Intelligence*, volume 2253, pages 261–268. Springer-Verlag, Berlin, 2002.

71. L. A. Zadeh. Fuzzy logic, Neural Networks, and Soft Computing. *Communications of the ACM*, 37:77–84, 1994.
72. S. Mitra, R. K. De, and S. K. Pal. Knowledge-Based Fuzzy MLP for Classification and Rule Generation. *IEEE Transactions on Neural Networks*, 8:1338–1350, 1997.
73. S. Mitra and S. K. Pal. Fuzzy Multi-Layer Perceptron, Inferencing and Rule Generation. *IEEE Transactions on Neural Networks*, 6:51–63, 1995.
74. S. Mitra and S. K. Pal. Fuzzy Self Organization, Inferencing and Rule Generation. *IEEE Transactions on Systems Man and Cybernetics Part A-Systems and Humans*, 26:608–620, 1996.
75. S. K. Pal and A. Ghosh. Neuro-Fuzzy Computing for Image Processing and Pattern Recognition. *International Journal of System Science*, 27(12): 1179–1193, 1996.
76. K. Cpalka. A New Method for Design and Reduction of Neuro-Fuzzy Classification Systems. *IEEE Transactions on Neural Networks*, 20(4): 701–714, 2009.
77. A. Gajate, R. E. Haber, P. I. Vega, and J. R. Alique. A Transductive Neuro-Fuzzy Controller: Application to a Drilling Process. *IEEE Transactions on Neural Networks*, 21(7): 1158–1167, 2010.
78. L. B. Goncalves, M. M. B. R. Vellasco, M. A. C. Pacheco, and F. J. de Souza. Inverted Hierarchical Neuro-Fuzzy BSP System: A Novel Neuro-Fuzzy Model for Pattern Classification and Rule Extraction in Databases. *IEEE Transactions on Systems Man and Cybernetics Part C-Applications and Reviews*, 36(2): 236–248, 2006.
79. Z.-L. Sun, K.-F. Au, and T.-M. Choi. A Neuro-Fuzzy Inference System Through Integration of Fuzzy Logic and Extreme Learning Machines. *IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics*, 37(5): 1321–1331, 2007.
80. W.-C. Wong, S.-Y. Cho, and C. Quek. R-POPTVR: A Novel Reinforcement-Based POPTVR Fuzzy Neural Network for Pattern Classification. *IEEE Transactions on Neural Networks*, 20(11): 1740–1755, 2009.
81. J. Zhang. Modeling and Optimal Control of Batch Processes Using Recurrent Neuro-Fuzzy Networks. *IEEE Transactions on Fuzzy Systems*, 13(4): 417–427, 2005.
82. L. Zhou and A. Zenebe. Representation and Reasoning Under Uncertainty in Deception Detection: A Neuro-Fuzzy Approach. *IEEE Transactions on Fuzzy Systems*, 16(2): 442–454, 2008.
83. S. K. Pal and A. Skowron, editors. Special Issue on Rough Sets, Pattern Recognition and Data Mining. *Pattern Recognition Letters*, 24(6), 2003.
84. J.-H. Chiang and S.-H. Ho. A Combination of Rough-Based Feature Selection and RBF Neural Network for Classification Using Gene Expression Data. *IEEE Transactions on NanoBioscience*, 7(1): 91–99, 2008.
85. J. Jiang, D. Yang, and H. Wei. Image Segmentation Based on Rough Set Theory and Neural Networks. In *Proceedings of the 5th International Conference on Visual Information Engineering*, pages 361–365. IET, UK, 2008.
86. S. K. Pal, L. Polkowski, and A. Skowron, editors. *Rough-Neuro Computing: Techniques for Computing with Words*. Springer, Heidelberg, 2003.
87. S. K. Pal. Soft Data Mining, Computational Theory of Perceptions, and Rough-Fuzzy Approach. *Information Sciences*, 163(1–3): 5–12, 2004.
88. S. K. Pal. Computational Theory of Perception (CTP), Rough-Fuzzy Uncertainty Analysis and Mining in Bioinformatics and Web Intelligence: A Unified Framework. *LNCS Transactions on Rough Sets*, 5946:106–129, 2009.

89. L. A. Zadeh. A New Direction in AI: Toward a Computational Theory of Perceptions. *AI Magazine*, 22:73–84, 2001.
90. S. Bornholdt and D. Graudenz. General Asymmetric Neural Networks and Structure Design by Genetic Algorithms. *Neural Networks*, 5:327–334, 1992.
91. V. Maniezzo. Genetic Evolution of the Topology and Weight Distribution of Neural Networks. *IEEE Transactions on Neural Networks*, 5:39–53, 1994.
92. S. K. Pal and D. Bhandari. Selection of Optimal Set of Weights in a Layered Network Using Genetic Algorithms. *Information Sciences*, 80:213–234, 1994.
93. S. Saha and J. P. Christensen. Genetic Design of Sparse Feedforward Neural Networks. *Information Sciences*, 79:191–200, 1994.
94. D. Whitley, T. Starkweather, and C. Bogart. Genetic Algorithms and Neural Networks: Optimizing Connections and Connectivity. *Parallel Computing*, 14:347–361, 1990.
95. A. T. Bjorvand and J. Komorowski. Practical Applications of Genetic Algorithms for Efficient Reduct Computation. In *Proceedings of the 15th IMACS World Congress on Scientific Computation, Modeling and Applied Mathematics*, Berlin, volume 4, pages 601–606, 1997.
96. D. Slezak. Approximate Reducts in Decision Tables. In *Proceedings of the 6th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Granada, pages 1159–1164, 1996.
97. J. Wroblewski. Finding Minimal Reducts Using Genetic Algorithms. In *Proceedings of the 2nd Annual Joint Conference on Information Sciences*, North Carolina, pages 186–189, 1995.
98. G. Ascia, V. Catania, and D. Panno. An Integrated Fuzzy-GA Approach for Buffer Management. *IEEE Transactions on Fuzzy Systems*, 14(4): 528–541, 2006.
99. J. Casillas, B. Carse, and L. Bull. Fuzzy-XCS: A Michigan Genetic Fuzzy System. *IEEE Transactions on Fuzzy Systems*, 15(4): 536–550, 2007.
100. C.-H. Chen, V. S. Tseng, and T.-P. Hong. Cluster-Based Evaluation in Fuzzy-Genetic Data Mining. *IEEE Transactions on Fuzzy Systems*, 16(1): 249–262, 2008.
101. V. Giordano, D. Naso, and B. Turchiano. Combining Genetic Algorithms and Lyapunov-Based Adaptation for Online Design of Fuzzy Controllers. *IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics*, 36(5): 1118–1127, 2006.
102. C.-S. Lee, S.-M. Guo, and C.-Y. Hsu. Genetic-Based Fuzzy Image Filter and Its Application to Image Processing. *IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics*, 35(4): 694–711, 2005.
103. A. Mukhopadhyay, U. Maulik, and S. Bandyopadhyay. Multiobjective Genetic Algorithm-Based Fuzzy Clustering of Categorical Attributes. *IEEE Transactions on Evolutionary Computation*, 13(5): 991–1005, 2009.
104. H. Kiem and D. Phuc. Using Rough Genetic and Kohonen's Neural Network for Conceptual Cluster Discovery in Data Mining. In *Proceedings of the 7th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, pages 448–452, Yamaguchi, Japan, 1999.
105. K. K. Ang and C. Quek. Stock Trading Using RSPOP: A Novel Rough Set-Based Neuro-Fuzzy Approach. *IEEE Transactions on Neural Networks*, 17(5): 1301–1315, 2006.

106. M. Banerjee, S. Mitra, and S. K. Pal. Rough-Fuzzy MLP: Knowledge Encoding and Classification. *IEEE Transactions on Neural Networks*, 9(6): 1203–1216, 1998.
107. R. Nowicki. On Combining Neuro-Fuzzy Architectures with the Rough Set Theory to Solve Classification Problems with Incomplete Data. *IEEE Transactions on Knowledge and Data Engineering*, 20(9): 1239–1253, 2008.
108. R. Nowicki. Rough Neuro-Fuzzy Structures for Classification with Missing Data. *IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics*, 39(6): 1334–1347, 2009.
109. S. K. Pal, S. Mitra, and P. Mitra. Rough-Fuzzy MLP: Modular Evolution, Rule Generation, and Evaluation. *IEEE Transactions on Knowledge and Data Engineering*, 15(1): 14–25, 2003.
110. G. Leng, T. M. McGinnity, and G. Prasad. Design for Self-Organizing Fuzzy Neural Networks Based on Genetic Algorithms. *IEEE Transactions on Fuzzy Systems*, 14(6): 755–766, 2006.
111. G.-C. Liao and T.-P. Tsao. Application of a Fuzzy Neural Network Combined with a Chaos Genetic Algorithm and Simulated Annealing to Short-Term Load Forecasting. *IEEE Transactions on Evolutionary Computation*, 10(3): 330–340, 2006.
112. A. Quteishat, C. P. Lim, and K. S. Tan. A Modified Fuzzy Min-Max Neural Network with a Genetic-Algorithm-Based Rule Extractor for Pattern Classification. *IEEE Transactions on Systems Man and Cybernetics Part A-Systems and Humans*, 40(3): 641–650, 2010.
113. M. Russo. FuGeNeSys: A Fuzzy Genetic Neural System for Fuzzy Modeling. *IEEE Transactions on Fuzzy Systems*, 6(3): 373–388, 1998.
114. T. L. Seng, M. Bin Khalid, and R. Yusof. Tuning of a Neuro-Fuzzy Controller by Genetic Algorithm. *IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics*, 29(2): 226–236, 1999.
115. W.-Y. Wang and Y.-H. Li. Evolutionary Learning of BMF Fuzzy-Neural Networks Using a Reduced-Form Genetic Algorithm. *IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics*, 33(6): 966–976, 2003.