

# Hybrid System based on Rough Sets and Genetic Algorithms for Medical Data Classifications

*Hanaa Ismail Elshazly, Scientific Research Group in Egypt (SRGE), Faculty of Computers and Information, Cairo University, Giza, Egypt*

*Ahmad Taher Azar, Faculty of Computers and Information, Benha University, Benha, Egypt*

*Aboul Ella Hassanien, Scientific Research Group in Egypt (SRGE), Faculty of Computers and Information, Cairo University, Giza, Egypt*

*Abeer Mohamed Elkorany, Faculty of Computers and Information, Cairo University, Giza, Egypt*

---

## ABSTRACT

*Computational intelligence provides the biomedical domain by a significant support. The application of machine learning techniques in medical applications have been evolved from the physician needs. Screening, medical images, pattern classification, prognosis are some examples of health care support systems. Typically medical data has its own characteristics such as huge size and features, continuous and real attributes that refer to patients' investigations. Therefore, discretization and feature selection process are considered a key issue in improving the extracted knowledge from patients' investigations records. In this paper, a hybrid system that integrates Rough Set (RS) and Genetic Algorithm (GA) is presented for the efficient classification of medical data sets of different sizes and dimensionalities. Genetic Algorithm is applied with the aim of reducing the dimension of medical datasets and RS decision rules were used for efficient classification. Furthermore, the proposed system applies the Entropy Gain Information (EI) for discretization process. Four biomedical data sets are tested by the proposed system (EI-GA-RS), and the highest score was obtained through three different datasets. Other different hybrid techniques shared the proposed technique the highest accuracy but the proposed system preserves its place as one of the highest results systems four three different sets. EI as discretization technique also is a common part for the best results in the mentioned datasets while RS as an evaluator realized the best results in three different data sets.*

*Keywords: Biomedical Classification, Data Mining, Discretization, Entropy Gain Information (EI), Feature Selection (FS), Genetic Algorithm (GA), Machine Learning (ML), Rough Set (RS)*

---

## 1. INTRODUCTION

The increasing advent of information technology provides historical data a great value and allows its efficient reusing (Wu et al., 2005). Limited human capability makes the high size of stored information becomes less useful. Human inspection and interpretation of the data is not feasible since no one can understand and benefit from the massive databases. The huge size of the historical data is the damning factor that limits the promising expected benefit from this stored data. Typically, these resources may include irrelevant features and noise. Limited resources and learning speed are important factors that must be considered in generalization performance. Increasing computational complexity may lead to intractable behavior. Knowledge discovery is the process of extracted hidden pattern and clusters from massive stored data. It concerns with the elicitation of buried knowledge from massive databases (Liao et al., 2011; Han & Kamber, 2011). It includes ensemble of sub processes that can extract the required knowledge with an efficient manner and benefit significantly from this flood of information. Data preprocessing, feature selection, data mining and evaluation are the main sub processes of KD process. Data preprocessing is the first step in KD process. It concerns with the data preparation, cleaning, discretization and removal of outliers as well as the managing of inconsistency. Some preprocessing techniques are related to understanding the distribution of the data like central tendency and dispersion measure of the data. These descriptive data summaries can be presented graphically like bar charts or histograms which is very helpful for visual inspection of the data (Kotsiantis, 2007). Others like discretization are related to data preparation to be better understanding and managing. Discretization is a process which adopts the concept of hierarchy generation. It tends to improve the data mining process by reducing the number of managing values. Feature selection is the process of specifying a minimal feature subset including the most relevant

features that best contribute in classification process based on evaluation criteria (Ladha & Deepa, 2011; Chen et al., 2011; Lavrac, 1999). Zhao et al. (2008) recommended the application of a feature selection process for pattern recognition, machine learning and data mining. This recommendation is motivated by the considerable support provided to subsequent steps of knowledge extraction. Feature selection process reduces the dimensionality of feature space and computational cost which decrease drastically the storage resource and running time. In order to increase the speed of training and improve the predictive accuracy, we get rid of noise. The retained data is considered the most relevant data that make better understanding of extracted knowledge and facilitate the visualization function (Escalano et al., 2009; Tsai, 2009). Concerning the evaluation step, Rough Set (RS) theory has many applications including the medical domain (Hassanien et al., 2009). Rough sets theory provides a novel approach to knowledge description and approximation of sets. It was introduced by Pawlak during the early 1980s (Pawlak, 1982) and is based on an approximation space-based approach to classify sets of objects. RS outperforms many other techniques by some properties. Firstly, it doesn't need any external parameters which provide an advantage over many other techniques. Secondly, it can ascertain the completeness of the data for the classification task especially for limited or expensive information sources (Pawlak, 1991; Polkowski, 2003; Yu & Liu, 2004). Rough set methods can also be used to classify unknown data based on already gained knowledge. It can be utilized to determine whether sufficient data for a task is available to extract a minimal sufficient set of features for classification. Reduct is an important concept in rough set theory and data reduction is a main application of rough set theory in pattern recognition and data mining. In this context, this paper proposes an integrated model that decomposes applying genetic algorithms for feature selection and rough set during classification process for prediction problem in the

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

[www.igi-global.com/article/hybrid-system-based-on-rough-sets-and-genetic-algorithms-for-medical-data-classifications/101768?camid=4v1](http://www.igi-global.com/article/hybrid-system-based-on-rough-sets-and-genetic-algorithms-for-medical-data-classifications/101768?camid=4v1)

## Related Content

---

### Nature That Breeds Solutions

Raymond Chiong, Ferrante Neri and R. I. McKay (2012). *International Journal of Signs and Semiotic Systems* (pp. 23-44).

[www.igi-global.com/article/nature-that-breeds-solutions/101250?camid=4v1a](http://www.igi-global.com/article/nature-that-breeds-solutions/101250?camid=4v1a)

### A Transactions Pattern for Structuring Unstructured Corporate Information in Enterprise Applications

Simon Polovina and Richard Hill (2009). *International Journal of Intelligent Information Technologies* (pp. 33-47).

[www.igi-global.com/article/transactions-pattern-structuring-unstructured-corporate/2450?camid=4v1a](http://www.igi-global.com/article/transactions-pattern-structuring-unstructured-corporate/2450?camid=4v1a)

### Conditional Hazard Estimating Neural Networks

Antonio Eleuteri, Azzam Taktak, Bertil Damato, Angela Douglas and Sarah Coupland (2009). *Encyclopedia of Artificial Intelligence* (pp. 390-395).

[www.igi-global.com/chapter/conditional-hazard-estimating-neural-networks/10277?camid=4v1a](http://www.igi-global.com/chapter/conditional-hazard-estimating-neural-networks/10277?camid=4v1a)

### Hand Gesture Recognition Using Multivariate Fuzzy Decision Tree and User Adaptation

Moon-Jin Jeon, Sang Wan Lee and Zeungnam Bien (2011). *International Journal of Fuzzy System Applications* (pp. 15-31).

[www.igi-global.com/article/hand-gesture-recognition-using-multivariate/55994?camid=4v1a](http://www.igi-global.com/article/hand-gesture-recognition-using-multivariate/55994?camid=4v1a)