

Chapter 3

Early–Stage Ovarian Cancer Diagnosis Using Fuzzy Rough Sets with SVM Classification

Nora Shoaip

Mansoura University, Egypt

Alaa M. Riad

Mansoura University, Egypt

Mohammed Elmogy

Mansoura University, Egypt

Hosam Zaghloul

Mansoura University, Egypt

Farid A. Badria

Mansoura University, Egypt

ABSTRACT

Ovarian cancer is one of the most dangerous cancers among women which have a high rank of the cancers causing death. Ovarian cancer diagnoses are very difficult especially in early-stage because most symptoms associated with ovarian cancer such as Difficulty eating or feeling full quickly, Pelvic or abdominal pain, and Bloating are common and found in Women who do not have ovarian cancer. The CA-125 test is used as a tumor marker, high levels could be a sign of ovarian cancer, but sometimes it is not true because not all women with ovarian cancer have high CA-125 levels, particularly about 20% of ovarian cancers are found at an early stage. In this paper, we try to find the most important rules helping in Early-stage ovarian cancer Diagnosis by evaluating the significance of data between ovarian cancer and the amino acids. Therefore, we propose a Fuzzy Rough feature selection with Support Vector Machine (SVM) classification model. In the pre-processing stage, we use Fuzzy Rough set theory for feature selection. In post-processing stage, we use SVM classification which is a powerful method to get good classification performance. Finally, we compare the output results of the proposed system with other classification technique to guarantee the highest classification performance.

DOI: 10.4018/978-1-5225-2229-4.ch003

INTRODUCTION

Ovarian cancer is formed in tissues of woman ovary (National Cancer Institute, 2014). Abnormal cells of this cancer can be found in one or both ovaries that have the ability to spread to the pelvis and abdomen parts then it spread in all body (Mayoclinic, 2014). Annually ovarian cancer is diagnosed in nearly a quarter of a million women in the entire world and is responsible for Hundreds of thousands of deaths each year. A small percentage of infected women with ovarian cancer, about 45%, can live for only five years compared to 89% of other women infected with breast cancer (World Ovarian Cancer, 2014). Ovarian cancer can be increased with custom groups of women. For example, women aged above 50 years, women who did not give birth or have difficulty in pregnancy or women who have relatives infected with cancer, such as breast, ovarian cancer, colon, uterine, or cervical cancer (Centers for Disease Control and Prevention CDC, 2014).

Cancer antigen-125 (CA-125) is a protein found with high rates of ovarian cancer cells more than other normal cells. CA-125 is created on the surface of cells then moving to the blood stream (Johns Hopkins University, 2014). Physicians used CA-125 as a tumor marker for ovarian cancer by measuring the levels of the CA-125 protein in a woman's blood. However, it is not an adequate early detection tool. High levels of CA-125 could be used as a sign of ovarian cancer, but it is not accurate or effective role because not all women with ovarian cancer have high CA-125 levels. Ovarian cancer is difficult to be diagnosed in early stages until it spreads to different body parts. More than 60% of women discover this cancer in stage III or stage IV cancer (Wikipedia, 2014a).

Medical diagnosis contains a high degree of difficulty that faces two main problems. The first problem of the medical diagnosis is a classification process. It must analyze many factors in difficult circumstances, such as diagnosis disparity and limited observation. The second problem is the uncertainty of the processed data that affects the diagnosis process.

Among the machine learning techniques that can deal efficiently with different degrees of difficulty of data problems, such as incomplete, uncertain, and inconsistent data, Rough set theory (Pawlak, 1982) is used to analysis vague and uncertain data. In practice, Rough set classifies the discrete attributes with high accuracy. It cannot be done well with real-valued, continuous attributes. Therefore, it leads to the creation of hybrid systems to integrate the Rough set theory with other machine learning technique, such as a Fuzzy set. These methods are complementary to each other, and the combination of them can provide improved solutions for dealing with continuous attributes.

The fuzzy-rough set theory is a successful hybrid model between the Rough set and the fuzzy set that proposed by Dubois and Prade to be one of the most common and efficient feature selection algorithms (Dubois & Prade, 1990, Hassanien Aboul Ella et.al. 2010). The fuzzy-rough set (Jensen & Shen, 2004; Jensen, 2005; Shen & Jensen, 2007; Chen et al., 2008) is a generalization of the lower and upper approximation of the rough set to have greater flexibility in handling uncertainty. As compared to the Rough sets, fuzzy-rough handles the uncertainty present in real data type in a better way without making any transformation such as discretization.

On the other hand, there are several classification methods widely used in medical data based on symptoms and health conditions (Lee & Chang, 2010; Chen et al., 2011; Yoo et al., 2012; Abdul Khaleel et al., 2013). Naïve Bayes classifier is considered as simple and rapid classification algorithm, but it is based on strong feature independence assumptions that sometimes considered as a significant disadvantage. Decision tree classifiers provide good visualization of data that allows users to understand the structure of all data easy, but decision tree may be too complicated when a data set contains many attributes.

16 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/early-stage-ovarian-cancer-diagnosis-using-fuzzy-rough-sets-with-svm-classification/180939

Related Content

Coronary Heart Disease Prognosis Using Machine-Learning Techniques on Patients With Type 2 Diabetes Mellitus

Angela Pimentel, Hugo Gamboa, Isa Maria Almeida, Pedro Matos, Rogério T. Ribeiro and João Raposo (2017). *Ubiquitous Machine Learning and Its Applications* (pp. 89-112).

www.irma-international.org/chapter/coronary-heart-disease-prognosis-using-machine-learning-techniques-on-patients-with-type-2-diabetes-mellitus/179090

Collaborative Bayesian Image Annotation and Retrieval

Rui Zhang and Ling Guan (2011). *Machine Learning Techniques for Adaptive Multimedia Retrieval: Technologies Applications and Perspectives* (pp. 146-169).

www.irma-international.org/chapter/collaborative-bayesian-image-annotation-retrieval/49107

How Games Improve Language in People With Language Dysfunctions

Robert Wahlstedt (2017). *Ubiquitous Machine Learning and Its Applications* (pp. 195-216).

www.irma-international.org/chapter/how-games-improve-language-in-people-with-language-dysfunctions/179095

Measurement of Cognitive Functional Sizes of Software

Sanjay Misra (2009). *International Journal of Software Science and Computational Intelligence* (pp. 91-100).

www.irma-international.org/article/measurement-cognitive-functional-sizes-software/2795

An Ensemble of Hybrid Search-Based Algorithms for Software Effort Prediction

Wasiur Rhmann (2021). *International Journal of Software Science and Computational Intelligence* (pp. 28-37).

www.irma-international.org/article/an-ensemble-of-hybrid-search-based-algorithms-for-software-effort-prediction/280515