International Journal of Information Retrieval Research
Volume 11 • Issue 1 • January-March 2021

A Novel Fuzzy Frequent Itemsets Mining Approach for the Detection of Breast Cancer

Ramesh Dhanaseelan F., St. Xavier's Catholic College of Engineering, India
Jeyasutha M., St. Xavier's Catholic College of Engineering, India

ABSTRACT

Breast cancer, a type of malignant tumor, affects women more than men. About one-third of women with breast cancer die of this disease. Hence, it is imperative to find a tool for the proper identification and early treatment of breast cancer. Unlike the conventional data mining algorithms, fuzzy logic-based approaches help in the mining of association rules from quantitative transactions. In this study, a novel fuzzy methodology, IFFP (improved fuzzy frequent pattern mining), based on a fuzzy association rule mining for biological knowledge extraction, is introduced to analyze the dataset in order to find the core factors that cause breast cancer. It is determined that the factor, mitoses, has low range of values on both malignant and benign, and hence it does not contribute to the detection of breast cancer. On the other hand, the high range of bare nuclei shows more chances for the presence of breast cancer. Experimental evaluations on real datasets show that the proposed method outperforms recently proposed state-of-the-art algorithms in terms of runtime and memory usage.

KEYWORDS

Breast Cancer, Crisp Set, Data Mining, Fuzzy Frequent Itemsets, Fuzzy Logic

INTRODUCTION

Breast cancer is one of the diseases that mostly affect women and is the second leading cause of cancer death (Jemal et al., 2003). It is the most dangerous disease because about one third of women with breast cancer die of the disease, although it is curable when detected early (Scheidhauer et al., 2004). Women over the age of 50 are mostly affected by this disease. Every year about 48,000 cases occur in the UK (Dr. Tim Kenny et al., 2012). Around one in nine women is affected by this disease at some stage in their life. It can easily be cured if diagnosed at an early stage. Therefore it is necessary for the proper identification and early treatment of the disease.

Mammography is one of the frequently used methods to detect breast cancer (Chou et al., 2004). The value of mammography can identify breast abnormalities with 85-90% accuracy (Elsayad, 2010). Fine needle aspiration cytology (FNAC) is also widely used in the diagnosis of breast cancer. Still, its average correct identification rate is only 90% (Elmore et al., 1994). Hence, it is necessary to develop better identification method to diagnose the breast cancer. Several researchers have employed

DOI: 10.4018/IJIRR.2021010102

statistical techniques and artificial intelligence techniques to predict breast cancer (Pendharkar et al., 1999; Kovalerchuck et al., 1997). The main objective of these identification techniques is to identify whether the person belongs to either a benign group that does not have breast cancer or a malignant group that has a strong evidence of having breast cancer. So, the diagnosis of breast cancer problems is more general and widely discussed classification problem (Anderson, 1984; Hand, 1981; Johnson and Wichern, 2001; Dillon and Goldstein, 1984; Gilal et al., 2019; Okikiola et al., 2019).

RELATED WORK

Previous studies refer to a number of techniques to diagnose breast cancer pattern. Neural network, a classification method based on which many algorithms (Chou et al., 2004; Karabatak and Ince, 2009; Seral et al., 2007; Marcano-Cedeno et al., 2011; Abbass, 2002; Tuba and Yildirim, 2004, Shukla et al., 2018) have been developed for diagnosing breast cancer. Artificial neural networks and multivariate adaptive regression splines approach (Chou et al., 2004), Association rules and Neural network approach (Karabatak and Ince, 2009), radial basis function neural network classification technique (Subashini et al., 2009), Genetic algorithm based approach (Pena-Reyes and Sipper, 1999) and support vector machines (SVM) (Polat and Gunes, 2007; Akay, 2009; Majid et al., 2014; Maglogiannis et al., 2009; Zheng et al., 2014) are some of the techniques used in breast cancer detection. A data separation/classification method called isotonic separation technique (Ryu et al., 2007) is one of the methods followed in predicting breast cancer. In Salama et al., (2012) different classifiers like multilayer perceptron neural network, combined neural network, probabilistic neural network, recurrent neural network and SVM were analyzed for classification accuracies of breast cancer detection. Lu et al., (2017) proposed an automated computer aided diagnosis framework which consists of ensemble under-sampling (EUS) for imbalanced data processing, the relief algorithm for feature selection, the subspace method for providing data diversity, and Adaboost for improving the performance of base classifiers. They extracted morphological, various texture, and Gabor features for magnetic resonance imaging (MRI).

Wang et al., (2018) proposed an SVM-based ensemble learning algorithm to reduce the diagnosis variance and increase diagnosis accuracy. Sivakumar et al., (2018) developed an algorithm for breast cancer diagnosis based on Supervised Learning in Quest (SLIQ) and Decision Tree algorithms. Peng at al., (2016) proposed an automated breast cancer diagnosis algorithm which organically integrates artificial immune with semi-supervised learning. Jafari-Marandi et al., (2018) presented a data and decision analytic method that employs both supervised and unsupervised learning powers of ANNs to optimize breast cancer diagnosis with regard to decision-making goals. Alwidian et al., (2013) developed a new technique based on a weighted method to select more useful association rules and a statistical measure for pruning rules for breast cancer disease.

Data mining plays the main role in the detection of knowledge from medical data repositories that could benefit medical diagnosis and for disease prevention (Nahar et al., 2013). Most of conventional data mining algorithms find the relation among transactions with binary values. However, transactions with quantitative values are commonly seen in real world applications. Fuzzy logic based approaches (Lin et al., 2015; Khezri et al., 2014; Keles and Yavuz, 2011; Nilashi et al., 2017; Li et al., 2018; Tran et al., 2018; Lee et al., 2019) take the major role in mining of association rules from quantitative transactions.

Several fuzzy mining methods have been proposed for finding interesting linguistic association rules from transaction data with quantitative values. These methods can be divided into two types: level-wise (Chan and Au, 1997; Hong and Chen, 1999; Hong et al., 2004; Hong et al., 2001; Kuok et al., 1998) algorithms and pattern-growth (Lin et al., 2015; Lin et al., 2009, 2010; Hong et al., 2010; and Papadimitriou and Mavroudi (2005) algorithms. Level-wise approach generates patterns containing 1 items, then 2 items, 3 items, etc. It repeatedly scans the database to count the support of each pattern. On the other hand, pattern-growth algorithms utilize a depth-first search instead of

Related Content