International Journal of Sociotechnology and Knowledge Development
Volume 14 • Issue 1

Modified Dominance-Based Soft Set Approach for Feature Selection

Jothi G., Department of Computer Applications, Sona College of Arts and Science, Salem, Tamil Nadu, India

Hannah Inbarani H., Periyar University, Salem, India

Ahmad Taher Azar, College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia & Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt*

https://orcid.org/0000-0002-7869-6373

Khaled M. Fouad, Faculty of Information Technology and Computer Science, Nile University, Sheikh Zayed City, Giza, Egypt & Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt

Sahar Fawzy Sabbeh, College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia & Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt

ABSTRACT

Big data analysis applications in the field of medical image processing have recently increased rapidly. Feature reduction plays a significant role in eliminating irrelevant features and creating a successful research model for big data applications. Fuzzy clustering is used for the segmenting of the nucleus. Various features, including shape, texture, and color-based features, have been used to address the segmented nucleus. The modified dominance soft set feature selection algorithm (MDSSA) is intended in this paper to determine the most important features for the classification of leukaemia images. The results of the MDSSA are evaluated using the variance analysis called ANOVA. In the dataset extracted function, the MDSSA selected 17% of the features that were more promising than the existing reduction algorithms. The proposed approach also reduces the time needed for further analysis of big data. The experimental findings confirm that the performance of the proposed reduction approach is higher than other approaches.

KEYWORDS

Classification, Dominance Relation, Dominance Soft Set, Feature Reduction, Leukemia

1. INTRODUCTION

In recent years, feature selection in big data becomes tremendously important in several life sectors such as medicine, engineering and science (Aziz et al., 2013a,b, 2012;; Hassanien et al., 2019a,b, 2017, 2014a,b, 2015; Ahmed et al., 2020; Mallek et al., 2020; Asad et al., 2012, 2014a,b,c,d; Jothi et al. 2013; Sayed et al., 2019, 2020). One of the critical issues in analyzing biomedical data is the curse of dimensionality especially the data with few samples and presented in high dimensional feature space (Inbarani et al., 2020, 2018, 2014a,b,c,d, 2016, 2015a,b, 2012, 2013; Inbarani and Banu, 2012; Muliawaty et al., 2019). Most of the medical tissues are obviously 3D images, so that irrelevant features not only lead to insufficient classification accuracy, but also reduces the processing speed or computational time (Inbarani et al., 2014a). Feature reduction (also called as variable selection,

DOI: 10.4018/IJSKD.289036 *Corresponding Author

attribute selection or subset selection) is the process of choosing a prevailing subset of features that is most correlated to the decision classes (Inbarani et al., 2014b; Wahhab, 2015). The digital format of the medical images gives an opportunity for further analysis that may lead to a more accurate diagnosis and helps in the optimized patient management.

In media image processing, it is very difficult to cope as data volumes are increasingly growing. Big data analysis currently plays a crucial role in the management, organization and analysis of data. Big data solutions provide new outlines for the analysis of medical images in a manner close to big data. It is time to develop new methods/architectures based on big data technologies for the complete processing of biomedical image data. In biomedical image processing, images are recorded in various imaging modes such as CT Scan, MRI, X-Ray, Ultra Sound, and PET-CT. It geenrates a large volume of images and it is very difficult for the radiologist to manage these large volumes of images. This research is an attempt to apply digital image processing and machine learning techniques in the area of medical image analysis and recognition. It focuses on developing a method to segment and diagnose the Acute Lymphoblastic Leukemia (ALL) nucleus. Fuzzy Clustering is utilized to segment the nucleus. It is a novel population-based stochastic algorithm proposed by Civicoglu (Civicioglu, 2013). After segmenting the image, relevant and representative features are extracted from the segmented nucleus. During this process, different kinds of features are extracted, namely, shape, colour and texture features.

In texture features, Grey Level Co-occurrence Matrix (GLCM) is computed for the dimensions 0^0, 45^0, 90^0, and 135^0. Molodtsov in 1999 designed a kind of soft set theory as an effective tool to deal with inconsistencies (Molodtsov, 1999). A wide range of applications of soft sets have been developed in recent years such as big data analytics, image processing and decision making (Inbarani et al., 2018; Kumar et al., 2014, 2015a). In this research, a Modified Dominance Soft Set-Based Feature Reduction Algorithm (MDSSA) is proposed to select the most prominent features for the leukemia image classification. The existing algorithms namely, Supervised Quick Reduct (SQR) (Jensen & Shen, 2009), Supervised PSO based-Quick Reduct (SPSOQR) (Inbarani et al., 2014a), Supervised PSO based Relative Reduct (SPSORR) (Inbarani et al., 2014a) and Soft Set based Unsupervised Quick Reduct (SSUQR) (Jothi and Inbarani, 2012) are utilized to compare the effectiveness of the MDSSA algorithm. The main contributions of this work are as follows. First, Fuzzy Clustering Algorithm is designed to segment the leukemia blast cells. Second, texture based GLCM features, shape and colour based features are extracted from the segmented nucleus. Third, MDSSA is applied to reduce the dimension of the feature set for the leukemia image classification. Finally, different classification algorithms are evaluated using appropriate classification measures. This approach is capable of eliminates redundant features efficiently better than the other available algorithms, which was supported by experimental results. The proposed architecture is shown in Figure 1.

Figure 1. The proposed architecture