International Journal of Electrical and Electronics Engineering (IJEEE)
ISSN(P): 2278-9944; ISSN(E): 2278-9952
Vol. 7, Issue 4, Jun – Jul 2018, 1-8

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

IMPROVE EFFICIENCY OF CANCER CLASSIFICATION BY COMBINING SELECTED FEATURE AND ADDITIONAL ELEMENTS

Duong Hung Bui[1], Manh Cuong Nguyen[2], Thi Hong Nguyen[3] & Xuan Tho Dang[4]

[1]Hanoi Trade Union University, Vietnam

[2,3,4]Faculty of Information Technology, Hanoi National University of Education, Vietnam

ABSTRACT

In fact, many problems with imbalanced data mean that the number of elements of a class is much larger than the number of elements of the remaining classes. This is the major reason for the declining performance of data classification. In addition, we found that in a number of imbalance datasets have many features redundant, unnecessary, not important to predict. Some reports have indicated if removing these features, it will increase the accuracy in imbalance data classification. Therefore, this paper studies, data balancing methods and reduces the number of attributes to improve the efficiency of data classification. Since then, we have developed a new method to reduce the number of feature and elements in the imbalance data classification. We experimented on some sets of biological data taken from the UCI like leukemia, colon-cancer and breast-p. These results show that our new method being more accurate classifiers with the G-mean measure compared with the method using original data. In addition, we use t-test evaluation indicated a method that the results have statistical significance with the p-value on the smaller datasets 0.05.

KEYWORDS: *Cancer Classification, Features Selection, Imbalanced Data, SMOTE*

Article History

Received: *14 Mar 2018* **|** ***Revised:*** *25 May 2018* **|** ***Accepted:*** *31 May 2018*