

عنوان مقاله:

CUOB-ReliefF: Diagnosis of breast cancer by balancing datasets

محل انتشار:

اولین کنفرانس سیستم‌ها و فناوری‌های محاسباتی مراقبت از سلامت (سال: 1398)

تعداد صفحات اصل مقاله: 6

نویسندگان:

Zeinab Abbasi - Faculty of Engineering Arak University Arak, Iran

Mohsen Rahmani - Faculty of Engineering Arak University Arak, Iran

Hossein Ghaffarian - Faculty of Engineering Arak University Arak, Iran

خلاصه مقاله:

One of the challenges of artificial intelligence and data mining algorithms in the automatic diagnosis of diseases is imbalanced dataset problem. The lack of data balancing will reduce accuracy of the results, which is very dangerous in diseases like breast cancer. This paper presents an algorithm for balancing number of instances in breast cancer datasets. The proposed algorithm uses ReliefF for weighting and ranking of instances. ReliefF is a well-known algorithm for ranking features, but, here, we used it with some modifications to rank the instances. After ranking the instances, based on the weight obtained, a combination of undersampling and oversampling methods is used to balance the dataset. The obtained results from testing the proposed algorithm on two datasets show the effectiveness of this algorithm.

کلمات کلیدی:

Breast cancer; Imbalanced datasets, ReliefF, Undersampling, Oversampling

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/923429>

