World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:17, No:04, 2023

An Efficient Machine Learning Model to Detect Metastatic Cancer in Pathology Scans Using Principal Component Analysis Algorithm, Genetic Algorithm, and Classification Algorithms

Authors : Bliss Singhal

Abstract : Machine learning (ML) is a branch of Artificial Intelligence (AI) where computers analyze data and find patterns in the data. The study focuses on the detection of metastatic cancer using ML. Metastatic cancer is the stage where cancer has spread to other parts of the body and is the cause of approximately 90% of cancer-related deaths. Normally, pathologists spend hours each day to manually classifying whether tumors are benign or malignant. This tedious task contributes to mislabeling metastasis being over 60% of the time and emphasizes the importance of being aware of human error and other inefficiencies. ML is a good candidate to improve the correct identification of metastatic cancer, saving thousands of lives and can also improve the speed and efficiency of the process, thereby taking fewer resources and time. So far, the deep learning methodology of AI has been used in research to detect cancer. This study is a novel approach to determining the potential of using preprocessing algorithms combined with classification algorithms in detecting metastatic cancer. The study used two preprocessing algorithms: principal component analysis (PCA) and the genetic algorithm, to reduce the dimensionality of the dataset and then used three classification algorithms: logistic regression, decision tree classifier, and k-nearest neighbors to detect metastatic cancer in the pathology scans. The highest accuracy of 71.14% was produced by the ML pipeline comprising of PCA, the genetic algorithm, and the k-nearest neighbor algorithm, suggesting that preprocessing and classification algorithms have great potential for detecting metastatic cancer.

Keywords : breast cancer, principal component analysis, genetic algorithm, k-nearest neighbors, decision tree classifier, logistic regression
Conference Title : ICCST 2023 : International Conference on Cancer Science and Therapy
Conference Location : Boston, United States
Conference Dates : April 17-18, 2023