

ABSTRACT

Leukemia is a blood-forming cell cancer that affects the lymphatic and bone marrow systems. It is the most common type of cancer in children. Leukemia is a form of blood cancer that develops when the white blood cell (WBC) level in the bone marrow becomes abnormally high. Acute leukemia has a long remission period. Chronic leukemia expands slowly and steadily, whereas Acute leukemia spreads rapidly and swiftly. Microarray data on gene expression are critical sources of information for classifying different cancer types. Although many data mining algorithms have been developed, none have been successful due to the small sample size and high density of the microarrays used. A medical expert's opinion and various tests are required as part of the diagnostic process, which is time-consuming and expensive. An automated diagnostic system is essential for a reliable prediction. Detecting blood cancer through the combination of leukemia microarray and machine learning methods has become a significant field of research. Precision and efficacy require further improvements, such as automation and low computational requirements, to create reliable models, despite research efforts.

This study describes an approach for classifying blood cancers based on machine learning. The current study analyzed several microarray datasets associated with leukemia to develop these models. It has been discovered that ensemble learning can be used to produce an accurate and reliable diagnostic approach for leukemia. Voting classifiers and ensembles were used to determine the primary features of the investigation. A voting classifier is a computer-based model capable of generating several models and predicting an output based on the model with the highest probability of becoming the outcome. This research proposed a new combination voting classifier of machine learning techniques, called LDSVM, which combines logistic regression (LR), decision trees (DT), and support vector machines (SVM) to arrive at final predictions. The categorization in this study is accomplished by using a voting classifier. The target of the three mixture classifier models is to improve leukemia detection. Several machine learning components were evaluated to increase the accuracy, recall, and F1 score. An AutoML library that is both quick and lightweight A Fast and Lightweight AutoML Library (FLAML) was used to find accurate models. FLAML simplifies selecting learners and hyperparameters while incurring minimal computational overhead.

Along with extensive testing on the datasets, the importance of the proposed method was determined by good results using the proposed advanced techniques. The results indicate that the suggested model for LDSVM achieves an accuracy of 0.985, with a soft voting accuracy of 0.958, and an accuracy of 0.95 when utilizing a hard vote, respectively. After five-fold cross-validation, the proposed method achieves accuracy by employing the LDSVM's most efficient grid search. The LDSVM model outperformed the competing FLAML-based models on the microarray datasets. It can identify and categorize leukemia and other diseases in the medical profession as a machine learning tool that employs ensemble learning.

Keywords: *Voting Classifier, Ensemble Classifier; Leukemia; Machine Learning; Genes; Cancer; Classification, Microarray, DNA; A fast and lightweight AutoML library (FLAML), Logistic regression, Decision tree, and Support vector machine (LDSVM)*