

Machine Learning models applied to the subtypes classification of Acute Myeloid Leukemia and Myelodysplastic Syndrome

Marcelo Mendes Brandão, Fábio Malta de Sá Patroni

UNIVERSIDADE ESTADUAL DE CAMPINAS

Abstract

Myeloid Malignancies are clonal diseases of hematopoietic or progenitor stem cells. Among the five main types are Acute Myeloid Leukemia (AML) and Myelodysplastic Syndrome (MDS). Leukemias are one of the most common cancers in Brazil. In 2018, they accounted for approximately 3% of new cancer cases. MDS are considered nowadays the most common class of acquired medullary failure syndromes in adults. They are also considered the most prevalent hematologic malignancies. There is a risk of transforming AML in approximately 1 out of 3 patients. Acute Myeloid Leukemia is a potentially fatal disease, common in children and adults, that can lead to death if left untreated. The MDS occurs predominantly in older male patients, with an average age of diagnosis of approximately 70 years. Here we are applying an automated method for classification of AML and MDS into their subtypes using a machine learning polytopic classifier. Two models will be created, one for each disease, which are trained with various clinical tests to predict accurate classification results. The models will be written in Python, using two popular frameworks: Scikit-Learn and Tensorflow. Input data for training and testing the models will come from three major public databases: the GDC Data Portal, the GDC Legacy Archive, and the NCBI GEO. The validation of the models will be performed with data from patients of the UNICAMP Blood Center. The classification process will be taken one step further in the research field. The approach proposed here can be used as a tool to help pathologists.

Funding: This work was supported by grants from CNPq (870370/1997-9).

Link to Video: