Informatics in Medicine Unlocked 20 (2020) 100399

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: http://www.elsevier.com/locate/imu

Treatment outcome classification of pediatric Acute Lymphoblastic Leukemia patients with clinical and medical data using machine learning: A case study at MAHAK hospital

Amirarash Kashef [a], Toktam Khatibi [a,*], Azim Mehrvar [b,c]

[a] School of Industrial and Systems Engineering, Tarbiat Modares University (TMU), Tehran, 14117-13114, Iran
[b] Mahak Hematology Oncology Research Center (Mahak-HORC), Mahak Hospital, Tehran, Iran
[c] AJA Cancer Epidemiology Research and Treatment Center (AJA-CERTC), AJA University of Medical Sciences, Tehran, Iran

ARTICLE INFO

Keywords:
Acute Lymphoblastic Leukemia (ALL)
children blood cancer
machine learning
treatment-related complications
MAHAK hospital
SVM
XGBoost

ABSTRACT

Introduction: Acute Lymphoblastic Leukemia (ALL) is the most common cancer among children. With the advancements of science and technology, the mortality rate of ALL is highly reduced. The aim of this study is treatment outcome classification of ALL patients aged less than 18 years with clinical and medical data using machine learning. For this purpose, ALL pediatric patients younger than 18 years treated at MAHAK multi-super specialty hospital from 2012 to 2018 are analyzed. Furthermore, MAHAK hospital is a reference center for treatment of childhood malignancies in Iran.

Data: In this study, data is collected manually from the paper-based records of 241 patients. Features included are patient demographic characteristics, medical information and treatment-related complications.

Method: Two scenarios are designed for data analytical purposes in this study. The first one considers all pediatric ALL patients but the second scenario excludes the patients with unknown cause of death from the study. As a whole, common classification algorithms are employed and tuned properly and compared to find the model showing superior performance.

Results: Our experimental results show that the XGBoost algorithm outperforms the compared classifiers with an accuracy of 88.5% (95% CI: 82.3–94.0) in the first designed scenario. On the other hand, the superior model in the second scenario is SVM with an accuracy of 94.90% (95% CI: 88.49–98.32) accuracy.

Conclusion: Despite several previous works that have analyzed gene expression data for ALL patients, the experimental results in this study show that clinical and medical data has reasonable importance in this area of research, too. Results show a significant improvement in the treatment outcome prediction utilizing the SVM algorithm. Moreover, our findings illustrate that the frequency of fever for a patient is the most predictive factor of the ALL treatment outcome.

1. Introduction

Blood is an important human body component that performs numerous vital functions, such as passing minerals, oxygen and carbon dioxide to the whole body to maintain metabolism. Blood has four essential components: Red Blood Cells (RBC), White Blood Cells (WBC), platelets (PLT) and Hemoglobin (HG) [1]. Leukemia is a type of blood or bone marrow cancer characterized by an irregular dramatic increase in the number of immature white blood cells named “blasts”. The term “leukemia” covers a wide spectrum of blood diseases [2]. Leukemia is classified into acute leukemia with a rapid progressive ability, and chronic leukemia that progresses slowly and has several obscure complications [1,3–5]. Acute leukemia infects the blood and bone marrow. Children and adults can develop numerous abnormal white blood cells in their body. Still, very recent developments have occurred to discover accurate preventive methods for acute leukemia disease [3]. Several risk factors have been identified for this dangerous and life-threatening disease. For example, the environmental factors such as exposure to benzene and ionizing radiation are highly associated with the development of childhood acute leukemia. Maternal history of fetal loss can also contribute to raise the risk of this fatal disease [6].

Acute type leukemia is classified into two classes based on a

* Corresponding author. Tel.: +982182883913.
E-mail addresses: amirarash.kashef@modares.ac.ir (A. Kashef), toktam.khatibi@modares.ac.ir (T. Khatibi), DRAZIMMEHRVAR@yahoo.com (A. Mehrvar).

https://doi.org/10.1016/j.imu.2020.100399
Received 16 April 2020; Received in revised form 14 July 2020; Accepted 15 July 2020
Available online 19 July 2020