Int J Syst Assur Eng Manag
https://doi.org/10.1007/s13198-024-02327-6

ORIGINAL ARTICLE

Noninvasive prediction of metastasis in esophageal cancer using ensemble-based feature selection

Syed Wajid Aalam[1] · Ab Basit Ahanger[1] · Assif Assad[2] · Muzafar A. Macha[3] · Muzafar Rasool Bhat[1]

Received: 14 November 2023 / Revised: 4 March 2024 / Accepted: 26 March 2024

Abstract Esophageal cancer (EC) is a significant health concern worldwide, and predicting its metastatic progression is essential for planning effective treatments. Histopathological intervention is the gold standard for diagnosing Esophageal Cancer Metastasis (ECM). However, we introduced a noninvasive, data-driven approach utilizing different machine learning (ML) algorithms on clinical data from TCGA to predict the risk of ECM. Among these algorithms, CatBoost stands out, achieving a 73% accuracy and a 75% area under the curve (AUC) using 5-fold cross-validation with a standard deviation of 4% among 5-folds. We visualized feature importance graphs and feature correlations to explain the decision-making of ML models. Our findings highlight associations between height, weight, age, alcohol consumption, the number of packs smoked, tumor location, and the risk of metastasis in EC patients. This approach offers a promising way to enhance EC metastasis prediction while minimizing invasive procedures.

Keywords Metastasis · TCGA · Machine learning · Feature importance · Esophageal cancer

1 Introduction

Esophageal cancer (EC) is a common malignancy of the upper gastrointestinal tract (Morita et al. 2002), ranking as the eighth most commonly diagnosed cancer worldwide and the sixth leading cause of global cancer-related mortality (Pakzad et al. 2016; Choksi et al. 2020). Esophageal cancer is very aggressive, demonstrating the infiltration of neighbouring organs, even during its initial stages (Shakeel et al. 2019). Due to the lack of early clinical symptoms, most patients are diagnosed with esophageal cancer at the middle and late stages, often with distant metastases (Huang and Yu 2018; Wang et al. 2023; Cincibuch et al. 2012). Distant metastasis significantly contributes to the mortality rate among patients diagnosed with EC (Langley and Fidler 2011). The tendency of esophageal cancer for metastasis is frequently observed in areas such as the lung, pleura, liver, stomach, peritoneum, kidney, adrenal gland, and bone (Verstegen et al. 2020; Mackay and Chen 2019; Hess et al. 2006; Mandard et al. 1981; Quint et al. 1995). Esophageal cancer (EC) continues as an exceptionally lethal malignancy, marked by a challenging prognosis, despite notable advancements in diagnosis and treatment over recent decades. At the point of initial diagnosis, roughly half of EC patients exhibit distant metastases, and more than one-third subsequently develop distant metastatic lesions following surgical

✉ Syed Wajid Aalam
wajid.aalam@iust.ac.in

✉ Muzafar A. Macha
muzafar.macha@iust.ac.in

✉ Muzafar Rasool Bhat
muzafar.rasool@islamicuniversity.edu.in

Ab Basit Ahanger
basit.ahanger@iust.ac.in

Assif Assad
assif.assad@islamicuniversity.edu.in

[1] Department of Computer Science, Islamic University of Science and Technology, 1-University Avenue, Awantipora, Pulwama, Jammu and Kashmir 192122, India

[2] Department of Computer Science and Engineering, Islamic University of Science and Technology, 1-University Avenue, Awantipora, Pulwama, Jammu and Kashmir 192122, India

[3] Watson Crick Centre for Molecular Medicine, Islamic University of Science and Technology, 1-University Avenue, Awantipora, Pulwama, Jammu and Kashmir 192122, India

Published online: 15 April 2024