*Intelligent Automation & Soft Computing*
DOI:10.32604/iasc.2021.016242
***Article***

Tech Science Press

# Early Detection of Lung Carcinoma Using Machine Learning

**A. Sheryl Oliver[1], T. Jayasankar[2], K. R. Sekar[3,*], T. Kalavathi Devi[4], R. Shalini[5], S. Poojalaxmi[5] and N. G. Viswesh[5]**

[1]Department of CSE, St.Joseph College of Engineering, Chennai, 600119, Tamilnadu, India
[2]Department of Electronics and Communication Engineering, University College of Engineering, BIT Campus, Anna University, Tiruchirappalli, 620024, Tamilnadu, India
[3]School of Computing, SASTRA Deemed University, Thanjavur, 613401, Tamilnadu, India
[4]Department of Electronics and Instrumentation Engineering, Kongu Engineering College, Erode, 638060, Tamilnadu, India
[5]School of Computing, SASTRA Deemed University, Thanjavur, 613401, Tamilnadu, India
*Corresponding Author: K. R. Sekar. Email: krsekarsastra@gmail.com
Received: 28 December 2020; Accepted: 07 May 2021

**Abstract:** Lung cancer is a poorly understood disease. Smokers may develop lung cancer due to the inhalation of carcinogenic substances while smoking, but non-smokers may develop this disease as well. Lung cancer can spread to other parts of the body and this process is called metastasis. Because the lung cancer is difficult to identify in the initial stages. The objective of this work is to reduce the mortality rate of the disease by identifying it at an earlier stage based on the existing symptoms. Artificial intelligence plays active roles in tasks such as entropy extraction through preprocessing strategies, ordinal to cardinal value conversions, table normalizations for easy meta computations, and preparation of machine learning tools for iterative processes to achieve rational convergence. The machine learning methodologies incorporated in this work are the cross-validation classification tree, random forest cross-validation classification, and random tree, all of which are included in an ensemble algorithm. The ensemble algorithm classifies lung cancer with maximum precision rates. The outcome of the classification provides 94.3% accuracy, which is the highest precision rate in comparison with the conventional methodologies. Semantics preprocessing of a lung cancer training set is performed with least entropy, and then translation, aggregation, and navigation based methodologies are applied for identifying the disease at its initial stage.

**Keywords:** Carcinogenic; cross-validation; classification tree; random forest; cross-validation; classification; random tree; translation; aggregation

## 1 Introduction

Lung cancer is one of the most common types of cancer today. It is a malignant tumour capable of growing at rapid rates in an uncontrolled manner. The malignancy of the tumour can be determined with the help of the ground-glass capacity strategy as well as image cropping and feature extraction using gray-level co-occurrence matrices (GLCM). Classification procedures have been accomplished with the