# Classification Prediction of Lung Cancer Based on Machine Learning Method

Dantong Li, Weifang Hospital of Traditional Chinese Medicine, China

Guixin Li, Weifang Hospital of Traditional Chinese Medicine, China*

Shuang Li, Weifang Hospital of Traditional Chinese Medicine, China

Ashley Bang, The St. Nicholas School, Vietnam

## ABSTRACT

The K-nearest neighbor interpolation method was used to fill in missing data of five indicators of coronary heart disease, diabetes, total cholesterol, triglycerides, and albumin;, and the SMOTE algorithm was used to balance the number of variable indicators. The Relief-F algorithm was used to remove 18 variable indicators and retain 42 variable indicators. LASSO and ridge regression algorithms were used to remove eight variable indicators and retain 52 variable indicators; The prediction accuracy, recall, and AUC values of the linear kernel support vector machine model filtered using Relief-F and LASSO features are high, and the prediction results are optimal; The test result of random forest screened by Relief-F and LASSO features is better than that of the support vector machine model. It is concluded that the random forest model screened by Relief-F features is better as a prediction of lung cancer typing. The research results provide theoretical data support for predicting lung cancer classification using machine learning methods.

## KEYWORDS

Lung Cancer Typing, Machine Learning, Random Forest, Support Vector Machine

## INTRODUCTION

With rapid economic and social development, human lifestyle and eating habits have changed significantly. In the face of long-term unhealthy living conditions, ionizing radiation, poor environment, and other adverse factors, the incidence rate of cancer in China is increasing year by year, and the types of cancer are also increasing. According to a survey from the International Agency for Research on Cancer, the number of cancer deaths worldwide is growing exponentially. The number of deaths due to different cancers in 2019 was 1.8 million for lung cancer, 870,000 for colorectal cancer, 780,000 for gastric cancer, 780,000 for liver cancer, and 630,000 for breast cancer (Wang & Yuan, 2019) (see Figure 1). Among them, lung cancer has the highest incidence rate and is particularly prominent among men. Lung cancer, as the most common fatal disease worldwide, is influenced by multiple factors. Smoking has been identified as the main risk factor for lung cancer, and smokers are more than 10 times more likely to develop lung cancer than nonsmokers. Harmful substances

**Figure 1. Statistics of cancer death cases**
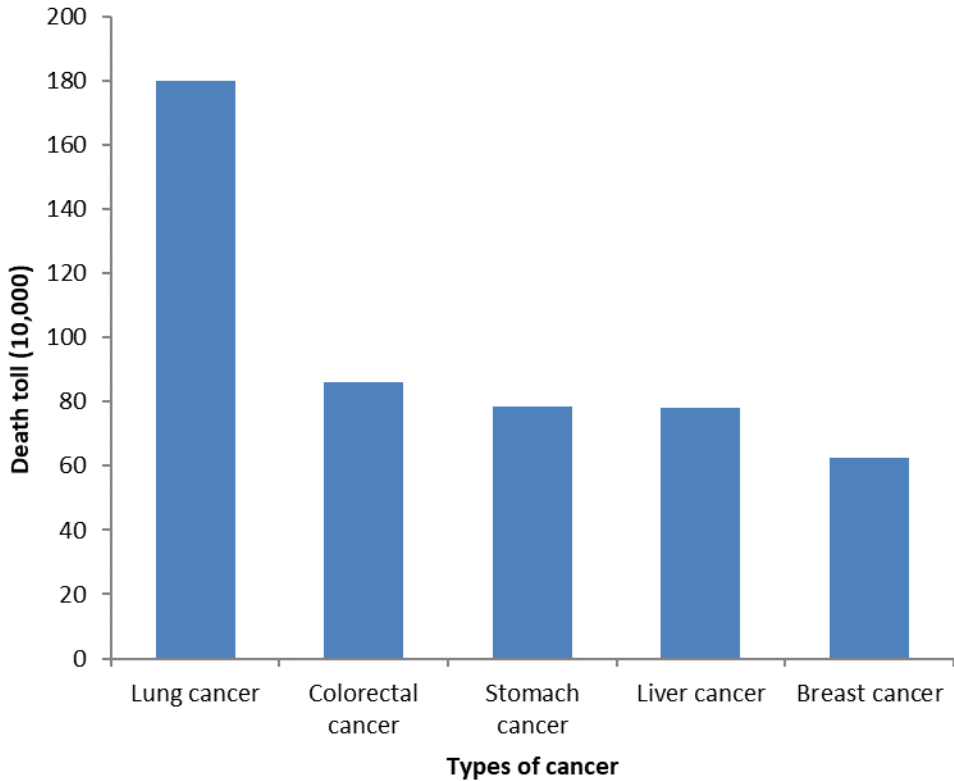


such as PM2.5, sulfur dioxide, and carbon monoxide in the air are also increasing the risk of lung cancer. At the same time, professions such as mining, welding, and painting are increasing the risk of lung cancer due to long-term exposure to harmful substances. According to the latest cancer burden data, there are over 4 million confirmed lung cancer patients worldwide each year, with nearly half of them dying from cancer.

Lung cancer poses a huge threat to human survival and health. The confirmed cases of lung cancer are mainly adenocarcinoma, small cell lung cancer, and squamous cell carcinoma of the lung. The treatment methods for different types of lung cancer vary greatly (Abdullah et al., 2021). At the same time, it is necessary to pay attention to the patient's psychological state and prescribe appropriate drugs before treatment. Tumor markers and imaging diagnosis of lung cancer are widely used in clinical practice, but some markers, such as carcinoembryonic antigen, are not specific enough to cause errors in clinical diagnosis. Imaging diagnosis (such as chest X-ray, CT, magnetic resonance imaging, etc.) has certain value for diagnosis; however, small pulmonary nodules or lymph node metastases may be missed due to poor imaging. The main treatment methods for lung cancer include surgery, radiotherapy, chemotherapy, and targeted therapy. For early-stage lung cancer patients, surgical treatment can be used and is currently the most effective treatment method. Radiotherapy, which kills cancer cells to alleviate symptoms, is mainly aimed at patients whose cancer cannot be surgically removed or who have residual cancer cells after surgery. Chemotherapy mainly targets patients with advanced lung cancer, killing cancer cells through intravenous injection or oral medication. Targeted therapy is the targeted killing of lung cancer cells by identifying their molecular targets. With the development of lung cancer screening technology, most lung cancer is easily detected in the early stage. At the

10 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/classification-prediction-of-lung-cancer-based-on-machine-learning-method/333631

## Related Content