



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 6, June 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



Lung Cancer Detection Using Machine Learning

Mr.R.KARTHIKEYAN.,¹ Mr.A.KARTHIKA²

Assistant Professor, Department of Master of Computer Application, Gnanamani College of Technology, Namakkal
Tamilnadu, India¹

PG Scholar, Department of Master of Computer Application, Gnanamani College of Technology, Namakkal,
Tamilnadu, India²

ABSTRACT: Lung cancer (LC) is one of the most serious cancers threatening human health. Histopathological examination is the gold standard for qualitative and clinical staging of lung tumors. However, the process for doctors to examine thousands of histopathological images is very cumbersome, especially for doctors with less experience. Therefore, objective pathological diagnosis results can effectively help doctors choose the most appropriate treatment mode, thereby improving the survival rate of patients. For the current problem of incomplete experimental subjects in the computer-aided diagnosis of lung cancer subtypes, this study included relatively rare lung adenocarcinoma (ASC) samples for the first time, and proposed a computer-aided diagnosis method based on histopathological images of ASC, lung squamous cell carcinoma (LUSC) and small cell lung carcinoma (SCLC). Firstly, the multidimensional features of 121 LC histopathological images were extracted, and then the relevant features (Relief) algorithm was used for feature selection. The support vector machines (SVMs) classifier was used to classify LC subtypes, and the receiver operating characteristic (ROC) curve and area under the curve (AUC) were used to make it more intuitive evaluate the generalization ability of the classifier. Finally, through a horizontal comparison with a variety of mainstream classification models, experiments show that the classification effect achieved by the Relief-SVM model is the best. The LUSC-ASC classification accuracy was 73.91%, the LUSC-SCLC classification accuracy was 83.91% and the ASC-SCLC classification accuracy was 73.67%. Our experimental results verify the potential of the auxiliary diagnosis model constructed by machine learning (ML) in the diagnosis of LC.

KEYWORDS: Lung cancer (LC), Area Under the Curve (AUC), Receiver Operating Characteristic (ROC) , Lung Squamous cell Carcinoma (LUSC), Adenosquamous Carcinoma (ASC), Small Cell Lung Carcinoma (SCLC).

I. INTRODUCTION

Images is one of the most important methods for evaluating LC subtypes . However, it is complicated and challenging for pathologists to review thousands of histopathological images, and it is even more difficult for doctors with less experience. Therefore, to relieve the pressure on doctors and improve the accuracy and efficiency of diagnosis, it is particularly important to study the computer-aided diagnosis model of LC . Li et al.: Research on Auxiliary Classification and Diagnosis of LC Subtypes NSCLC and the rest are SCLC . The main histological types of NSCLC are lung adenocarcinoma (ADC) and lung squalors cell carcinoma (LUSC). The other histological types of NSCLC are lung adenosquamous carcinoma (ASC), large-cell carcinoma . In particular, ASC is a relatively rare subtype of NSCLC that accounts for 0.3–5% of all NSCLCs . Due to the different histopathological types of LC, the treatment methods adopted are also different. When the lung tissue classification is determined, the appropriate treatment mode can be selected, such as the reasonable application of surgery, chemotherapy, radiotherapy, molecular targeted therapy and immune therapy. In addition, LC screening errors can be avoided, clinicians' multifarious work pressure can be slowed, patients' survival time can be maximized and the patient's quality of life improved. The LC imaging examination methods mainly include: (1) X-ray photography, which is one of the most basic lung imaging examination methods, but the resolution of the photography is low and there are blind spots in the examination . Computed tomography (CT), specifically, chest CT has advantages in detecting early peripheral LC and identifying the location of the lesion. Currently, it is one of the most commonly used imaging methods for preoperative diagnosis and staging of LC . However, one of the limitations of using CT examinations is that for patients undergoing repeated examinations, it is necessary to consider the impact of the radiation dose produced by the operation .



II. LITERATURE SURVEY

In this paper [1], Magnetic resonance imaging (MRI) has high sensitivity and specificity for vertebral and bone metastases, but it is not recommended for routine diagnosis of LC. Ultrasound is a non-invasive tool, which is usually superior to radiography in the examination of postoperative pulmonary complications (PPCs). Authors in [3], it has developed into a valuable method. Although the above imaging examination methods play an important role in detecting of LC, the results of each examination are only used as a reference for the diagnosis, staging, re-staging, efficacy monitoring and prognosis evaluation of LC, while histopathological examination is the gold standard for tumour qualitative and clinical staging. Authors in [4] due to the complex texture features of histopathological images, as far as the authors know, there is no computer-aided diagnosis method for ASC, LUSC and SCLC based on histopathological images. In this paper [5], ASC sample data for the first time and introduces a computer-aided model for automatic classification of LC subtypes based on histopathological images of LC. First, seven texture analysis methods are used to extract 265-dimensional features of LC histopathological images, and the relevant features (Relief) algorithm is used for feature selection.

III. EXISTING SYSTEM

There have been a number of lung cancer risk models developed and validated that one may consider to be a form of CADx tool (6-9). Typically based on logistic regression, such tools aim to provide an overall risk of the patient having cancer based on patient meta-data such as age, sex and smoking history and nodule characteristics such as nodule size, morphology and growth, if a previous CT was available.

Disadvantages

- Lung cancer diagnosed patient may suffer through Cough,
- Chest pain
- Shortness of breath
- Wheezing, Hemoptys.

IV. PROPOSED SYSTEM

The histopathological images are faced with a large number of rich geometric structures and complex textures caused by the diversity of structural morphology.

Advantages

- This study was the first to include relatively rare ASCs in lung histopathological images and applies them to the automatic classification of LC subtypes.
- This paper was the first to apply the Relief-SVM algorithm to the classification of LC histopathological images, which demonstrates the tremendous potential of ML algorithms to be used in the diagnosis of LC subtypes.

V. SYSTEM OVERVIEW

Data Collection - Research on Auxiliary Classification and Diagnosis of LC Subtypes. The unprocessed histological images of LUSC, ASC and SCLC. The processed histological images of LUSC, ASC and SCLC. To better display the image details, two pathological tissue images of each category were selected for display in this study, of which A was LUSC, B was ASC, and C was SCLC. 22 had LUSC, 27 had ASC and 45 had SCLC. All subjects had signed informed consent in this retrospective study. For improving the generalization performance of the model, this study took an average of 1-2 pathological tissue sections from each patient as the research object, and finally selected 121 histopathological images. According to the classification standards provided by the World Health Organization, LC patients were collected from the hospital pathology database, and the relevant pathological diagnosis results were confirmed by pathology. The detailed information is shown in Table 1. All tumor tissues included in the study were made into histopathological sections by haematoxylin and eosin (H&E) staining, which were collected through the microscope of the hospital pathology department and stored as JPG image files with four resolutions: 744 × 554, 2048 × 1536, 1024 × 768, 640 × 480 [6]. Patient demographics (n = 94). Histopathological examination is the gold standard for qualitative and clinical tumor staging. Histopathological images have been widely used by doctors for diagnosis and treatment, and are an important basis for predicting patient survivals. The histopathological images of LUSC, ASC and SCLC. According to reports, the following problems exist in histopathological images: The histopathological images are faced with a large number of rich geometric.

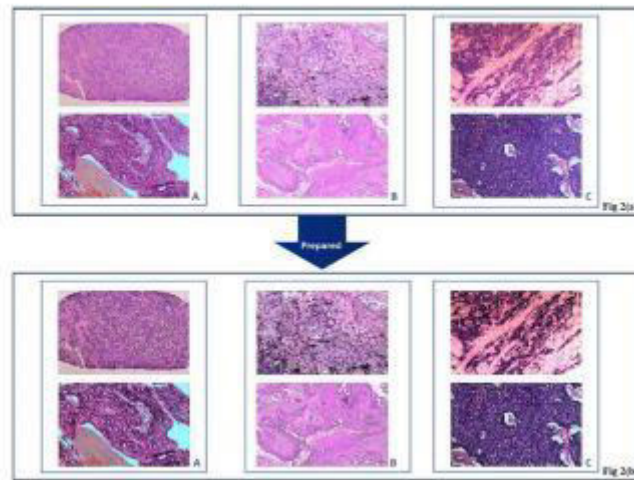


Fig.5.1 Histopathological Images

Feature Extraction Of Histopathological Images Of Lung Cancer - Because the rich features presented in histopathological images are an important basis for clinicians to carry out diagnosis, the effective extraction of image features is the key to improving the accuracy. Because the rich features presented in histopathological images are an important basis for clinicians to carry out diagnosis, the effective extraction of image features is the key to improving the accuracy extraction methods on LC histopathological image classification. 1) handcrafted texture extraction methods We extracted 265-dimensional features using seven handcrafted texture extraction algorithms, including the Hu invariant moments, GLGCM, wavelet transform, GLCM, LBP, GLDS.

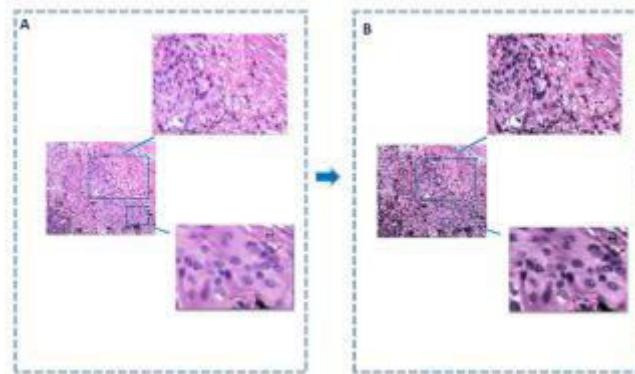


Fig.5.2 Handcrafted Texture Extraction Methods

VI. SYSTEM IMPLEMENTATION

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, uncovering key insights within data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase, requiring them to assist in the identification of the most relevant business questions and subsequently the data to answer them.



Fig.6.1. Model Optimization Process

- **Speech Recognition:** It is also known as automatic speech recognition (ASR), computer speech recognition, or speech-to-text, and it is a capability which uses natural language processing (NLP) to process human speech into a written format. Many mobile devices incorporate speech recognition into their systems to conduct voice search—e.g. Siri—or provide more accessibility around texting.
- **Customer Service:** Online chatbots are replacing human agents along the customer journey. They answer frequently asked questions (FAQs) around topics, like shipping, or provide personalized advice, cross-selling products or suggesting sizes for users, changing the way we think about customer engagement across websites and social media platforms. Examples include messaging bots on e-commerce sites with virtual agents, messaging apps, such as Slack and Facebook Messenger, and tasks usually done by virtual assistants and voice assistants.

$$\begin{aligned}
 \text{Rule: } X \Rightarrow Y & \begin{cases} \text{Support} = \frac{\text{freq}(X, Y)}{N} \\ \text{Confidence} = \frac{\text{freq}(X, Y)}{\text{freq}(X)} \\ \text{Lift} = \frac{\text{Support}}{\text{Supp}(X) \times \text{Supp}(Y)} \end{cases}
 \end{aligned}$$

- **Computer Vision:** This AI technology enables computers and systems to derive meaningful information from digital images, videos and other visual inputs, and based on those inputs, it can take action. This ability to provide recommendations distinguishes it from image recognition tasks. Powered by convolutional neural networks, computer vision has applications within photo tagging in social media, radiology imaging in healthcare, and self-driving cars within the automotive industry.
- **Recommendation Engines:** Using past consumption behavior data, AI algorithms can help to discover data trends that can be used to develop more effective cross-selling strategies. This is used to make relevant add-on recommendations to customers during the checkout process for online retailers.
- **Automated stock trading:** Designed to optimize stock portfolios, AI-driven high-frequency trading platforms make thousands or even millions of trades per day without human intervention.

HOW DOES MACHINE LEARNING WORK?

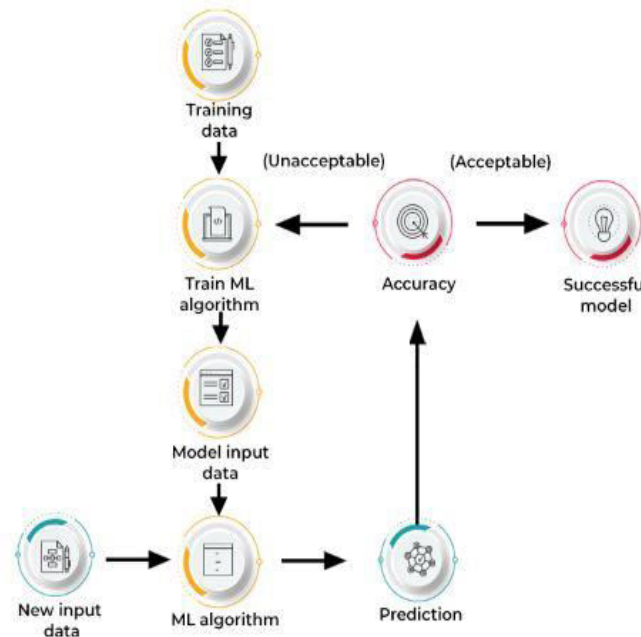


Fig.6.2. Machine Learning Work Process

VII. FUTURE ENHANCEMENT

To construct an efficient and accurate ML model for early LC, the model can be developed with the following parameters Data should be collected from large and highly qualified authorized center's. "e.g." www.cancerimagingarchive.net Data collected should be pre-processed by a powerful technique such that no important data is lost. Highly correlated Features with the output should be identified for best results. Using the Hybrid ML model, early prediction of LC can produce accurate results. Several ML tools and various platforms can be made available for researchers to provide good results. There are also many data analytical tool that can provide useful information for future data analysis.

VIII. CONCLUSION

Currently, ML Algorithms play a significant role in early LC prediction, and with the help of these techniques available data can be used to make predictions or decisions. Study work provided a proposed system followed by ML in predicting early LC which provides the researcher with better knowledge in ML Technique for early prediction of LC. Moreover, to make ML approaches easier in the field of LC, different types of the dataset used, various data preprocessing methods implemented, essential features that are been selected and extracted are been explained in detail. Also, the performance of different ML is evaluated. The parameters used in constructing an efficient and accurate ML model for early prediction of Lung cancer is a piece of additional information. This study will help the researchers to identify ML techniques that produces more accuracy and efficiency.

REFERENCES

1. R.Karthikeyan, & et all "Biometric for Mobile Security" in the international journal of Engineering Science & Computing, Volume7,Issue6, June 2017, ISSN(0):2361-3361,PP No.:13552-13555.
2. R.Karthikeyan, & et all "Data Mining on Parallel Database Systems" in the international journal of Engineering Science & Computing, Volume7,Issue7, July 2017, ISSN(0):2361-3361,PP No.:13922-13927.
3. R.Karthikeyan, & et all "Ant Colony System for Graph Coloring Problem" in the international journal of Engineering Science & Computing, Volume7,Issue7, July 2017, ISSN(0):2361-3361,PP No.:14120-14125.
4. R.Karthikeyan, & et all "Classification of Peer -To- Peer Architectures and Applications" in the international journal of Engineering Science & Computing, Volume7,Issue8, Aug 2017, ISSN(0):2361-3361,PP No.:14394-14397.



5. R.Karthikeyan, & et all "Mobile Banking Services" in the international journal of Engineering Science & Computing, Volume7,Issue7, July 2017, ISSN(0):2361-3361,PP No.:14357-14361.
6. R.Karthikeyan, & et all "Neural Networks for Shortest Path Computation and Routing in Computer Networks" in the international journal of Engineering and Techniques, Volume 3 Issue 4, Aug 2017, ISSN:2395-1303,PP No.:86-91.
7. R.Karthikeyan, & et all "An Sight into Virtual Techniques Private Networks & IP Tunneling" in the international journal of Engineering and Techniques, Volume 3 Issue 4, Aug 2017, ISSN:2395-1303,PP No.:129-133.
8. R.Karthikeyan, & et all "Routing Approaches in Mobile Ad-hoc Networks" in the International Journal of Research in Engineering Technology, Volume 2 Issue 5, Aug 2017, ISSN:2455-1341, Pg No.:1-7.
9. R.Karthikeyan, & et all "Big data Analytics Using Support Vector Machine Algorithm" in the International Journal of Innovative Research in Computer and Communication Engineering, Volume 6 Issue 9, Aug 2018, ISSN:2320 - 9798, Pg No.:7589 -7594.
10. R.Karthikeyan, & et all "Data Security of Network Communication Using Distributed Firewall in WSN " in the International Journal of Innovative Research in Computer and Communication Engineering, Volume 6 Issue 7, July 2018, ISSN:2320 - 9798, Pg No.:6733 - 6737.
11. R.Karthikeyan, & et all "An Internet of Things Using Automation Detection with Wireless Sensor Network" in the International Journal of Innovative Research in Computer and Communication Engineering, Volume 6 Issue 9, September 2018, ISSN:2320 - 9798, Pg No.:7595 - 7599.
12. R.Karthikeyan, & et all "Entrepreneurship and Modernization Mechanism in Internet of Things" in the International Journal of Innovative Research in Computer and Communication Engineering, Volume 7 Issue 2, Feb 2019, ISSN:2320 - 9798, Pg No.:887 - 892.
13. R.Karthikeyan & et all "Efficient Methodology and Applications of Dynamic Heterogeneous Grid Computing" in the International Journal of Innovative Research in Computer and Communication Engineering, Volume 7 Issue 2, Feb 2019, ISSN:2320 - 9798, Pg No.:1125 -1128.
14. R.Karthikeyan & et all"Entrepreneurship and Modernization Mechanism in Internet of Things" in the International Journal of Innovative Research in Computer and Communication Engineering, Volume 7 Issue 2, Feb 2019, ISSN:2320 - 9798, Pg No.:887– 892.
15. R.Karthikeyan & et all"Efficient Methodology for Emerging and Trending of Big Data Based Applications" in the International Journal of Innovative Research in Computer and Communication Engineering, Volume 7 Issue 2, Feb 2019, ISSN:2320 - 9798, Pg No.:1246– 1249.
16. R.Karthikeyan & et all "Importance of Green Computing In Digital World" in the International Journal of Innovative Research in Computer and Communication Engineering, Volume 8 Issue 2, Feb 2020, ISSN:2320 - 9798, Pg No.:14 – 19.
17. R.Karthikeyan & et all "Fifth Generation Wireless Technology" in the International Journal of Engineering and Technology, Volume 6 Issue 2, Feb 2020, ISSN:2395–1303.
18. R.Karthikeyan & et all "Incorporation of Edge Computing through Cloud Computing Technology" in the International Research I Journal of Engineering and Technology, Volume 7 Issue 9, Sep 2020 ,p. ISSN:2395–0056, e. ISSN:2395–0072.
19. R.Karthikeyan & et all "Zigbee Based Technology Appliance In Wireless Network" in the International Journal of Advance Research and Innovative Ideas in Education, e.ISSN:2395 - 4396, Volume:6 Issue: 5 , Sep. 2020. Pg.No: 453 – 458, Paper Id:12695.
20. R.Karthikeyan & et all "Automatic Electric Metering System Using GSM" in the International Journal of Innovative Research in Management, Engineering and Technology, ISSN: 2456 - 0448, Volume:6 Issue: 3 , Mar. 2021. Pg.No: 07 – 13.
21. R.Karthikeyan & et all "Enhanced the Digital Divide Sensors on 5D Digitization" in the International Journal of Innovative Research in Computer and Communication Engineering, e-ISSN: 2320 – 9801, p-ISSN: 2320 - 9798, Volume:9 Issue: 4 , Apr. 2021. Pg.No: 1976 – 1981.
22. R.Karthikeyan & et all "Crop Yield Prediction Based On Indian Agriculture Using Machine Learning" in the International Journal Of Engineering and Techniques, ISSN: 2395-1303, Volume:8 Issue: 4 , July. 2022. Pg.No: 11 – 22.
- 23.R.Karthikeyan & et all "College Bus Transport Management Web Application" in the International Journal Of Multidisciplinary Research In Science, Engineering and Technology, ISSN: 2582-7219, Volume: 6 Issue: 6, June. 2023. Pg.No: 1619 – 1625.
- 24.R.Karthikeyan & et all "Face Recognition Based Attendance System" in the International Journal of Innovative Research in Computer and Communication Engineering, ISSN: e 2320-9801, Volume: 11 Issue: 6, June. 2023. Pg.No: 8710 – 8717.



25. R.Karthikeyan & et all “Cloud data Deduplication System using per File parity and File Name Interpreter” in the International Journal of Advanced Research in Arts, Science, Engineering and Management, ISSN: 2395-7852, Volume: 10 Issue: 3, May. 2023.
26. R.Karthikeyan & et all “Secure Photo Sharing Social Networks Using Coverless Image Steganography Techniques” in the International Journal of Research in Science, Engineering and Technology, e - ISSN: 2319-8753, Volume: 12 Issue: 6, June. 2023. Pg.No: 8852 – 8861.
27. R.Karthikeyan & et all “Hotel Booking Mobile And Web Application” in the International Journal Of Multidisciplinary Research In Science, Engineering and Technology, ISSN: 2582-7219, Volume: 6 Issue: 6, June. 2023. Pg.No: 1823 – 1829.
28. R.Karthikeyan & et all “Campus Placements Prediction & Analysis Using Machine Learning” in the International Journal of Research in Science, Engineering and Technology, e - ISSN: 2319-8753, Volume: 12 Issue: 6, June. 2023. Pg.No: 8837 – 8841.
29. R.Karthikeyan & et all “E-Auction System Web Application” in the International Journal of Research in Science, Engineering and Technology, e - ISSN: 2319-8753, Volume: 12 Issue: 6, June. 2023. Pg.No: 8837 – 8841.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.423



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com



www.ijirset.com

Scan to save the contact details