International Journal of Computational Science and Engineering Research
ISSN: 3107 -8605 (Online) , http://www.ijcser.com/
Regular Issue , Vol.1, Issue. 3 (July – September) , 2024 , Pages: 13 - 17
Received: 25 April 2024 ; *Accepted:* 04 May 2024; *Published:* 04 July 2024
https://doi.org/10.63328/IJCSER-V1RI3P4
Original Paper : Applied Research – Compare and Contrast Research Paper

Detection of Lung Cancer using SVM Algorithm and Comparing With K-Means Accuracy

C Manoj Kumar [1] , Sindaluri Devisree [2] , E Harshitha [3] , E Gayatri [4] , P Hema Mahuri [5]

[1-5] Aditya college of Engineering, Madanapalle, Andhra Pradesh, India

** Corresponding Author: C Manoj Kumar; manojkumar@gmail.com*

Abstract: Thousands of people are infected with lung cancer every year, and the patient's chances of survival are extremely low if the disease is not detected in its early stages. For the reasons already stated and to aid in overcoming this dreadful situation, early diagnosis with the use of AI methods is absolutely necessary. Using a convolutional neural network approach with AlexNet architecture, this research provided a computer-aided system for identifying lung cancer in a dataset obtained from Iraqi hospitals. The system helps with the identification of patient cases, whether they are normal, benign, or malignant. With a maximum accuracy of 93.548%, the suggested model is quite precise. Accuracy is 98.714 percent and specificity is 95.0 percent, two further performance criteria with very high results.

Keywords: Lung Cancer Detection, SVM, K- Means , AI.

1. Introduction

Among the world's most recognizable and deadly diseases, lung cancer is a major concern [1]. Most recent estimates from the "World Health Organization" (WHO) put the annual death toll from lung cancer at over 7.6 million people throughout the globe. On top of that, it is projected that the global death toll from this cancer type would reach about 17 million in 2030 [2, 3]. Data compiled by "The American cancer society" indicates that lung cancer ranks higher than any other type of head cancer in terms of mortality rates among American adults [4]. In 2013, there were an estimated 1660290 new cases of cancer overall (854790 males and 805500 women), with 228199 instances of lung cancer (118080 men and 110110 women). Whereas 580350 instances of cancer-related deaths are anticipated (306920 male and 273430 female), 159480 cases of lung cancer related deaths were recorded (87260 male and 732220 female).

In 2016, lung cancer [5] ranked second in Iraq among all cancers, according to the country's ministry of health. Both sexes account for 2,123 cases of lung cancer. Roughly 8.31% of all illnesses in the nation fall into this category. This fraction shows a minor improvement over the ratio of the previous year, which was around 8.1%. The fact that lung cancer accounts for around 13.27% of all cancer occurrences demonstrates that it is the most common kind among men. There is an increase as well as compared to the ratio from 2015, which was at 12.7%.The numbers [6, 7]. Although it is not the most common cancer in women, 638 women were diagnosed with lung cancer in 2016, accounting for 4.44 percent of all cancer diagnoses. This illness ranks fifth among female cancers. If we compare this year to last, we see a small increase of roughly 4.2%.[2, 7] With an ever-increasing incidence, cancer ranks as the fourth leading killer in Iraq and the eastern Mediterranean area. Smoking is the most prominent and leading cause of this increase.

Low levels of physical activity, poor nutrition, chronic exposure to carcinogens in industrial and agriculture, and pollution are further contributors [6]. There were 8,211 cancer-related deaths in Iraq in 2014, with about 4,525 men and 3,959 women losing their lives to the disease. The overall projected mortality from all cancers was 16.31%, with lung cancer accounting for the largest share of all cancers with 1,339 fatalities; 918 of these deaths were in males and 421 were in women. Lung cancer accounted for around 1257 of the total cancer deaths in 2016, or 16.61% of the total anticipated mortality, whereas overall cancer deaths decreased to 7568 cases [6, 7]. In light of the foregoing, it is imperative that a computer-aided diagnosis (CAD) system be put into place to aid physicians in the early detection of lung cancer by accurately identifying nodules. Various studies have utilized AI techniques for this purpose. For instance, in [8, 9], artificial neural networks were used to detect lung cancer; in [10-12], support vector machines were used; in [13], K nearest neighbour was applied; in [14, 15], genetic algorithms were used; in [16-18], fuzzy techniques were found to be effective; and in [19-21], convolutional neural networks

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 1, Issue 3 , 2024, https://doi.org/10.63328/IJCSER-V1RI3P4

were used. Artificial intelligence is not only utilized in the field of lung cancer detection, but it is also employed throughout all of biomedical engineering, including breast cancer diagnosis [22–24] , heart disease diagnosis [25–27], and diabetes diagnosis and classification [28]. Data must be fed into the algorithms aforementioned approaches.

By analysing data collected from hospitals in Iraq, a computer-aided approach may now help distinguish between benign and malignant lung cancers. The system utilizes the AlexNet architecture of convolutional neural networks.

2. Literature Survey

Schwartz LM, Woloshin S, and Welch HG [1] More effective cancer treatments and fewer cancer deaths are typically assumed to be the causes of an increase in cancer patients' 5-year survival rates. Nevertheless, improvements in cancer detection, which have led to an increase in the number of patients diagnosed with early-stage cancer—including those who would never have experienced any symptoms—may possibly explain the observed improvement in 5-year survival rates.

Authors: Asuntha and Srinivasan [2] There are around five million fatal occurrences of lung cancer annually, making it one of the leading causes of death globally for both men and women. Lung illnesses can be better diagnosed with the help of Computed Tomography (CT) scans. Finding malignant nodules in the lung using an input picture and then determining the severity of the malignancy is the primary goal of this study. This research employs innovative Deep learning techniques to pinpoint the exact location of malignant lung nodules.

In this study, top feature extraction methods including Zernike Moment, Local Binary Pattern, Histogram of oriented Gradients (HoG), and wavelet transform-based features are employed. The optimal feature is chosen using the Fuzzy Particle Swarm Optimization (FPSO) method after texture, geometric, volumetric, and intensity characteristics have been extracted. Lastly, Deep learning is used to categorize these properties. A new FPSOCNN algorithm simplifies CNN calculation. Another dataset, this one originating from Arthi Scan Hospital and updated in real-time, undergoes an extra appraisal. Experimental demonstrate that new outperforms competing methods. findings FPSOCNN Choa T-S, Hong R, Nie L, Zhang L, Yang Y, Wang M the third Current observational health records, which include frequently.

multimedia data, allow us to comprehend the progressions of chronic illnesses, which are successfully controllable even though they cannot be cured. Disease progression has been the subject of an ever-expanding corpus of research. But up until now, the following three chronic illness progression findings have received incredibly little attention: 1) there is a strong correlation between health statuses across time; 2) current multimedia and multimodal observations, including visual scans, digital measurements, and textual medical histories, can provide a comprehensive picture of each patient's future health status; and 3) the discriminative capabilities of various modalities differ greatly according to individual diseases. Based on these considerations, we offer an adaptive multimodal multi-task learning model that can co-regulate the modalities' temporal progression, discriminative capacities, and modality agreement.

Our suggested model is proven to be a linear system by theoretical means. We use a matrix factorization strategy to fix the missing data problem before we train our model. Thorough tests on a dataset containing actual cases of Alzheimer's disease confirm the validity of our suggested model. It is worth mentioning that our concept may also be used for other types of chronic disorders.

Ahmedin, Rebecca, Naishadham, Deepa, Jemal [4] According to the most recent statistics on cancer incidence, mortality, and survival, which are based on data from the National Cancer Institute, the CDC, and the North American Association of Central Cancer Registries, as well as mortality figures from the National Center for Health Statistics, the American Cancer Society makes an annual estimate of the number of new cancer cases and deaths anticipated in the US for the current year.

In 2013, there will likely be 1,660,290 new cases of cancer and 580,350 cancer-related deaths in the US. Cancer mortality rates fell 1.8% annually in males and 1.5% annually in women over the most current five-year data set (2005-2009), whereas delay-adjusted cancer incidence rates fell 0.6% annually in men and were steady in women. From 1991, when they were at 215.1 per 100,000 people, to 2009, when they were at 173.1 per 100,000 people, the overall cancer mortality rate fell 20%.

Lung, colorectal, breast, and prostate cancer mortality rates are all going down. The most significant yearly decreases in mortality rates throughout the last decade of data (2000 2009) were observed in chronic myeloid leukemia (8.4%), stomach cancer (3.1%), colorectum cancer (4.0%), and non Hodgkin lymphoma (4.0%). Jayalakshmi K, Subashini T, and Ganesan S [5] The fast growth in the utilization of medical digital photographs can be attributed to the fact that medical

organizations are constantly collecting thousands of these photos. There is a growing requirement for effective data management and precise access due to the proliferation of medical digital pictures. The challenges of manual categorization have prompted the proposal of an automated solution. This study aims to automatically categorize X-ray pictures into six different groups: chest, foot, spine, neck, head, and palm. The classification is done at the macro level, or global level, utilizing statistical moments and a support vector machine classifier. Sixty photos are chosen from the IRMA

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 1, Issue 3 , 2024, https://doi.org/10.63328/IJCSER-V1RI3P4

database for each class. The M3 filter is used for pre processing, and then Connected Component Labeling (CCL) is used to find the region of interest. Statistical moments are then used to extract features. Support Vector Machines (SVMs) were employed to classify the retrieved features, yielding a 92.58% accuracy rate.

3. Proposed System

Among the world's most recognizable and deadly diseases, lung cancer is a major concern (1). Most recent estimates from the "World Health Organization" (WHO) put the annual death toll from lung cancer at over 7.6 million people throughout the globe. Using a convolutional neural network approach with AlexNet architecture, this research provided a computer-aided system for identifying lung cancer in a dataset obtained from Iraqi hospitals. The system helps with the identification of patient cases, whether they are normal, benign, or malignant. With a maximum accuracy of 93.548%, the suggested model is quite precise. Sensitivity is 95.714 percent and specificity is 95.0 percent, two further performance criteria with very high results.

4. Related Work

To identify and categorize lung cancer CT images acquired from hospitals, the suggested method use SVM. When it comes to analyzing data with a grid pattern, like photographs (31), SVMs is a paradigm that uses Deep Learning in conjunction with computer vision. Visual challenges, such as those involving images and videos, are ideal for gaining insight into this method, thus it's helpful to picture a Neural Network Architecture. Additionally, support vector machines are a crucial tool for object recognition, face recognition, and autonomous vehicles.

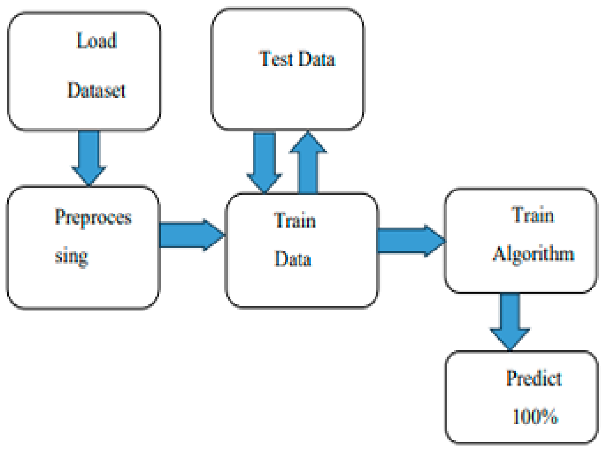

Figure. 1 System Architecture

Assigning learnable weights and biases to distinct objects inside an image and then training the algorithm to distinguish between them is the power of a Convolutional Neural Network, a type of Deep Learning method. Furthermore, as compared to other classification methods, our method requires far less pre-processing.

A convolutional neural networks (CNN) job is to simplify the pictures without sacrificing any of the elements that are crucial for making an accurate prediction (32). The convolutional (CONV), pooling (POOL), and classifier (FC) layers make up a typical convolutional neural network (CNN).

Implementation

Read Lung Cancer Dataset : The lung cancer dataset will be utilized. There are almost a hundred pictures in the collection. It includes pictures of both healthy and unhealthy lung disease. The purpose of these photographs is to train and test the algorithm.

Split Dataset : Following data loading, the dataset will be partitioned into two sections: one for algorithm training and another for algorithm testing.

SVM Algorithm : The data is trained using the SVM (support vector machine) technique. The algorithm will be trained and its accuracy will be provided.

Predict Lung Cancer : Here we provide the algorithm a picture as input. Whether it's normal or not, the algorithm will determine.

Accuracy Graph: Graphs will be used to display the algorithm accuracies.

5. Results and discusses

Here we have the accuracy calculations , Spiliting of the data and identifications of lung cancer images

Figure. 2 Data Splitting for training and testing

In above screen dataset loaded and now click on 'Read & Split Dataset to Train & Test' button to split dataset into train and test parts and application split 80% dataset for training and 20% dataset to test trained model.

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 1, Issue 3 , 2024, https://doi.org/10.63328/IJCSER-V1RI3P4

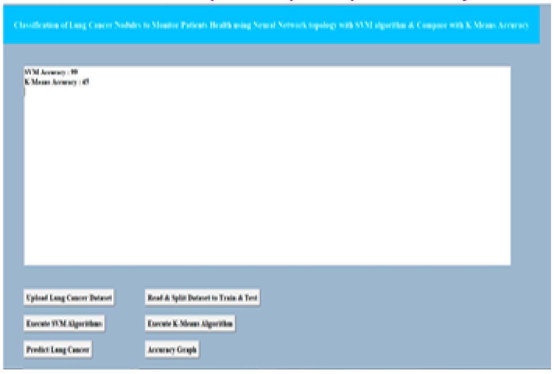

Figure.3 Shows both k means and SVM Accuracy

In above screen SVM accuracy is 98% and Execute K-Means Algorithm 48% from above figure we can say that SVM can predict more efficiently than k-mean.

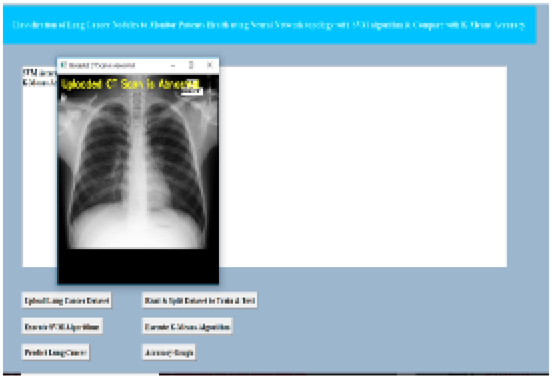

Figure.4 Identification of the Cancer

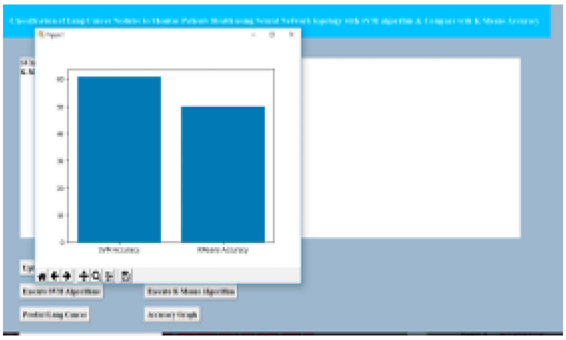

Figure.5 Comparing of both accuracies

The above figure shows the result shows the comparison between SVM and K-Means Accuracy .It proves that SVM can predict more accurately

6. Conclusion and Future Scope

The annual death toll from lung cancer is close to one million, making it one of the most lethal forms of the illness. The present state of medicine makes the diagnosis of lung nodules on chest CT scans an absolute need. Therefore, CAD systems are essential for the early diagnosis of lung cancer. Processing images is an essential task used in many different industries. Its primary usage is in X-ray lung imaging for the detection of tumors and other cancers. Finding cancerous areas in the lungs requires the application of image processing algorithms for tasks including noise reduction, damaged region feature extraction, identification, and comparison with data on the medical history of lung cancer. The use of machine learning and image processing in this work shows that lung cancer can be accurately classified and predicted. The first step is to gather pictures. The next step is to apply a geometric mean filter as a pre-processor to the photos. In the long run, this improves the picture quality. After that, the pictures are segmented using the k-means method. The region of interest may be more easily identified with this segmentation. The next step is to employ algorithms for classification that are based on machine learning. SVM is more accurate in predicting lung cancer. Systems that identify lung cancer using robust classification and prediction methods will benefit from this study's findings. For practical applications, this research provides new pictures generated by machine learning algorithms.

References

[1] Y . Welch HG, Schwartz LM, Woloshin S. Are increasing 5-year survival rates evidence of success against cancer? Jama. 2000;283(22):2975-8.

[2] Asuntha A, Srinivasan A. Deep learning for lung classification. Cancer detection Multimedia Applications. 2020:1-32. and Tools

[3] Nie L, Zhang L, Yang Y, Wang M, Hong R, Chua T-S, editors. 2015, Beyond doctors: Future health prediction from multimedia and multimodal observations. Proceedings of the 23rd ACM international conference on Multimedia;

[4] American Lung Association, URL www.lung.org. Siegel, Rebecca, Naishadham, Deepa, Jemal, Ahmedin. Cancer statistics, 2013. a cancer journal for clinicians. 2013;63(1):11-30.

[5] Republic of Iraq, 2016,Ministry of Health\Environment, Board. IC. Annual Report Iraqi Cancer Registry 2016.

[6] Republic of Iraq, Ministry of Health\Environment, 2015., Board. IC. Annual Report Iraqi Cancer Registry 2015.

[7] Nasser, Ibrahim M Abu-Naser, S. S. 2019, Lung Cancer Detection Using Artificial Neural Network. International Journal of Engineering Information Systems.;3(3):17-23.

[8] Taher, Fatma Sammouda, Rachid., editors. 2011, Lung cancer detection by using artificialneural network and fuzzy clustering methods. 2011 IEEE GCC Conference and Exhibition (GCC);: IEEE.

[9] Eskandarian P, Bagherzadeh J, editors. 2015, Computer-aided detection of Pulmonary Nodules based on SVM in thoracic CT images. 7th Conference on Information and Knowledge Technology (IKT); 2015: IEEE.

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 1, Issue 3 , 2024, https://doi.org/10.63328/IJCSER-V1RI3P4
[10] Ganesan S, Subashini T, Jayalakshmi K, editors. 2014,Classification of X-rays using statistical moments and SVM. 2014 International Conference on Communication and Signal Processing;: IEEE.
[11] Parveen SS, Kavitha CJIJoCA. 2014, Classification of lung cancer nodules using SVM Kernels.;95(25).
[12] Thamilselvan P, Sathiaseelan J. 2016, Detection and classification of lung cancer MRI imagesby using enhanced k nearest neighbor algorithm. Indian Journal of Science Technology.;9(43):1-7.
[13] Kurkure M, Thakare A, editors. 2016, Lung cancer detection using genetic approach. 2016 International Conference on Computing Communication Control and automation (ICCUBEA) IEEE.
[14] Tuncal, K., Sekeroglu, B., & Ozkan, C. (2020). Lung cancer incidence prediction using machine learning algorithms. Journal of Advances in Information Technology Vol, 11(2).
[15] Singh, G. A. P., & Gupta, P. K. (2019). Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans. Neural Computing and Applications, 31(10), 6863-6877.