International Journal of Engineering Sciences & Emerging Technologies, May 2024.
ISSN: 22316604 Volume 11, Issue 3, pp: 619-624 ©IJESET

LUNG CANCER DETECTION: A REVIEW

Sanjeev Gupta[1], Bhaskar Saini[2], Amol Jain[3],
Hammad Hussain[4], Chandraveer Singh[5]

[1] Assistant Professor, CSE Department

Moradabad Institute of Technology, Moradabad, India
sandygupta2@gmail.com

[2]B.Tech (CSE), Moradabad Institute of Technology, Moradabad, India
Imbhaskarsaini@gmail.com.

[3]B.Tech (CSE), Moradabad Institute of Technology, Moradabad, India
amoljaiin@gmail.com.

[4] B.Tech (CSE), Moradabad Institute of Technology, Moradabad, India
hammadzaidi211@gmail.com

[5]B.Tech (CSE), Moradabad Institute of Technology, Moradabad, India
chandraveersingh091@gmail.com

ABSTRACT

Lung cancer is a disease in which cancer cells originate from the lung itself or from another organ. Lung Cancer is also called a lung tumor and is characterized by highly uncontrolled cell growth. Early Detection of lung cancer is hard as most of the symptoms appear in the final stage. Smoking increases the chances of developing a lung cancer but nonsmokers can develop lung cancer too. Lung cancer is the leading cause of death worldwide. It is essential to identify the presence of cancer in early stages to improve the treatment and treatment process.

KEYWORDS

cancer, machine learning, medical imaging, nature inspired algorithms

1. INTRODUCTION

1.A. Lung Anatomy

Lungs are a part of the human respiratory system and the role of the lungs is to oxygenate the blood and remove carbon dioxide and other gases that are waste for the human body. Lungs are a pair of spongy, roughly cone-shaped structures present on either side of the chest.
Air is inhaled through the mouth or nose. The windpipe (trachea) brings in the inhaled air to the left bronchi and right bronchi. The bronchi are further separated into small branch-like structures called bronchioles. The bronchioles forms clusters of air sacs called alveoli. In the lungs, there are 300Million alveoli. Each alveoli is surrounded by blood vessels that dissolve the oxygen and absorb carbon dioxide to exhale.

1.B. Stages in Lung Cancer

Doctors use a system called the TNM system for staging lung cancer. T is the tumor size and location, N is the node involvement and *M* is the metastasis.

Lung Cancer Staging can be classified into two categories. *Small-Cell and non-small cell lung Cancer Stages.* In Small Cell lung cancer doctors use the TNM system to classify cancer into one of these stages.

- Limited Stage: Cancer is only in one lung and the second lung is healthy.
- Extensive Stage: Cancer has spread to the chest and both lungs.

In Non-Small-Cell Lung Cancer Stages doctors use the TNM system or general staging method. This type of cancer is more common than small cell cancer.

Stages in Non-Small-Cell Lung Cancer:

- Occult stage: It is also called hidden cancer. In this stage, the mucus contains cancer cells. Cancer in this stage can't be detected by using a biopsy or any imaging technique.
- Stage 0: In this stage, the tumor formed is very small and the lungs are still non-infected.
- Stage I: In this stage, lung tissue shows the presence of cancer.
- Stage II: In this stage, lymph nodes close to the lungs might be infected with cancer.
- Stage III: In this stage, the chest and the lymph node might be infected with cancerous cells.
- Stage IV: In this stage, bones, liver, brain, or any other body part might be infected with cancer.

1.C. Medical Imaging Techniques

X-ray: X-ray is painless and quick imaging technique that produces the image of structure inside your body. The Imagine method that is used in X-Ray is called Ionizing Radiations. X-rays are mainly used to identify bone fractures, digestive tract problems, swallowed items, breast cancer, arthritis, osteoporosis, infections.

Computed tomography scan (CT scan): CT scan machine is a large doughnut-shaped machine and the patient is made to lie down flat on a bed that passes through the machine. CT scan takes a series of X-rays which is used to create a cross-sectional image of the insides of body which includes blood vessels, soft tissues and bones. The imaging method used in CT scan is ionizing radiations. CT scans are majorly used to diagnose bone fractures, vascular disease, tumors and cancers, heart disease, infections, injuries from trauma and guided biopsies.

Magnetic resonance imaging (*MRI):* In MRI patient is made lie on a bed that goes into the MRI machine. The magnets inside MRI machine create loud striking noises. Magnetic waves are used in the MRI machine to produce the image. MRI is majorly used to diagnose stroke, aneurysms, spinal cord disorders, Multiple Sclerosis (MS), tumors, blood vessel issues, joint or tendon injuries.

Ultrasound: The Ultrasound makes use of the sound waves to produce image of the inside of body. The sound waves that are used in ultrasound are of high frequency i.e. 20 kHz and higher to produce the image of organs and structures inside the body. The doctor applies gel to your skin when doing ultrasound then presses a ultrasound transducer also called a probe against it and moves it to capture images of the inside of your body. It is used in monitoring pregnancy, diagnose gallbladder disease, genital/prostate issues, breast lump, joint inflammation, blood flow problems and used in guided biopsies.

Positron emission tomography scan (PET scan): In PET scans radioactive drugs called FDG and a scanning machine is used to show the functioning of patient tissue and organs. In a PET scan, the patient is made to swallow or inject a radiotracer. Then the patient is made to lie on a bed that goes through the scanner. A scanner is a doughnut-shaped machine that reads the radiations given off by the radiotracer. PET scans are majorly used to diagnose cancer, Alzheimer's disease, heart disease, coronary artery disease, epilepsy, seizures.

2.LITERATURE REVIEW

In the literature survey, we focused on recent research done in lung cancer detection using various machine learning algorithms.

In paper [1] the dataset is taken from TCIA which contains CT scan images of normal lungs and cancerous lung patients. Images were resized, sharpened, and then a median filter is applied to enhance & reduce noise in the CT scan images. Feature Extraction is done is find Are, Perimeter, Eccentricity, Compactness, and circularity. In the final stage, the support vector machine (SVM) algorithm is applied as the classifier to classify the detected tumor into malignant or benign. The authors in paper [2] described a model in which segmentation is performed which extract lung regions from the image. Distinct areas are divided in multilevel thresholding from a grayscale image. Brightness regions are divided that represent one background and threshold values are calculated for each region. The thresholding process allows highlighting of specific pixels in the image. The thresholding process gives an output image of the binary type having black and white pixels. The classification is performed using the AlexNet CNN model. This paper claims an accuracy of 96%.

In research paper [3] images were firstly converted into grey scaled and then segmentation is done using the thresholding process. In the Feature Extraction Stage function, Open CV and GLCM are used to calculate the entropy, contrast, homogeneity, and area of the segmented lung region. In the classification stage, SVM is used as a classifier. This model achieved an accuracy of 83.33%.

In the paper [4] the CT scan images of four patients with normal lung disease and four patients with some anomalous lung disease were collected. 100 sample images were obtained from each patient and each image collected was of size 512x512 pixels. The preprocessing stage was completed by making use of the median filter which removes noise in the images and also some unwanted features. The original image was segmented manually by using Adobe Photoshop CS6 and ImageJ software. The colors of the image were firstly inverted by using ImageJ software. Then, this color inverted image was transformed into a binary image with the help of a plugin available in the software. The color of the image was black and white. When the color of the image was black and white the threshold was adjusted by 99.70%. After that, the image was transferred into adobe photoshop software to remove the background of the images. The background was separated so that the segmentation can only be performed within the inside of the lung. After manual segmentation was completed feature extraction was performed to obtain the centroid, area, and perimeter. k-nearest neighbors(kNN) algorithm was used to classify the input pattern. kNN is the method of a statistical base for data classification. This paper achieved an accuracy of 98.15%.

In paper [5] the authors used LIDC-IDRI and LUNA 16 medical image database which consists of "250,000" medical images. The images are of DICOM format and the CT scan image is made up of about 200 slices of images and each slice is a squared image of 512 pixels. DICOM stands for Digital Imaging and Communications in Medicine. It is the standard developed for the communication and management of medical imaging information. Standardization is performed by staggering which specifies a distance of 1 millimeter for each slice of the CT scan. To optimize the model, images are converted into a size of 50×50 pixels.

The model is developed using a U-Net convolutional neural network. This method is mainly used for Biomedical Image Segmentation. For classification, they tested 8 different algorithms which include 1 classical method, i.e. back-propagation techniques and gradient descent, and 7 most used swarm algorithms i.e. Firework Optimization Algorithm (FOA), Particle Swarm Optimization (PSO), Harmony Search Algorithm (HSA), Artificial Bee Algorithm (ABC), Firefly Algorithm (FA), Bacterial Foraging Optimization (BFO).

Swarm intelligence algorithms are nature-inspired algorithms. They are global optimization algorithms used to solve complex problems. The maximum accuracy achieved is 93.71% which was obtained by using the Particle Swarm optimization algorithm (PSO). The maximum sensitivity was 92.96% with Harmony Search Algorithm (HSA) and the maximum specificity of 98.52% with Gravitational Search Algorithm.

Authors of paper[6] proposed a simulation model by conducting a comparative study on the prediction of lungs cancer in an early stage of lung cancer. They used the LIDC-IDRI dataset of lung cancer CT scan images for prediction. On which they used Gaussian filtering to reduce noise in their image preprocessing part further for feature selection and segmentation they used the firefly algorithm and MKFCM algorithm to divide the data set into clusters. And then they used the HFPS feature selection algorithm for feature extraction from segmented images. Most of the parts of their model include machine learning and Deep Learning concepts. They used the HFPS Feature selection algorithm for reducing the computing time and better classification of the model. And to classify more accurately they used three deep learning methods Auto Encoder Classification, LSTM based Classification, and Recurrent neural network-based classification. The maximum accuracy achieved was 97.43% by using the HFPS Feature Selection algorithm along with the Recurrent Neural Network classifier.

In the paper [7] a dataset was used which had 181 samples of images in which 150 of images of lungs had Adenocarcinoma or ADCA with is a common type of cancer while the other 31 images of lungs had Malignant pleural mesothelioma or MPM which a type of rare cancer. The dataset which contained these images is called the Harvard 2 dataset. This paper made use of a variety of gene selection algorithms that are Relief-F, Gene T-statistic, Information Gain, and Chi-square statistic. Discretization of the attributes is performed before applying the gene selection algorithms. The major purpose of using the gene selection algorithms is to pick the most appropriate genes from the 12,533 available genes.

The 1000 most appropriate genes have to undergo a second level feature selection method to find the best 200 genes, 100 genes, and 50 genes by making use of various optimization techniques. In this paper, the author has used 5 optimization techniques that are Moth Flame Optimization Algorithm (MFO), Grasshopper Optimization Algorithm(GOA), Bacterial Foraging Optimization Algorithm(BFO), Artificial Fish Swarm Optimization Algorithm(AFSO), and Krill Herd Algorithm(KH).

The most appropriate gene values that are obtained after applying the second-level optimization techniques are then used as input classification algorithms. Naive Bayesian Classifier, Support Vector Machine, Decision trees, and k-nearest neighbors algorithms are used. The paper claims the accuracy of 99.10%.

In [8] Gabor filter is used for image enhancement. This image obtained after applying Gabor Filter is then segmented using a region-based segmentation technique. The segmentation technique is used to modify the representation of an image so that it becomes easier to analyze. After this, they applied morphological operations namely erosion and dilation followed by reconstruction of the enhanced image. After this binarization is performed and an active contour algorithm is applied to the binarized image later watershed transform is performed.

In [9] the authors made a system consisting of segmentation, feature extraction, feature selection, and classification subsystem. In segmentation, the Shape Fuzzy C-Means clustering algorithm is applied for partitioning the lung tissues from the input lung CT scan image. The region of interest (ROI) is located in the lung tissue and segmented from the lung tissues using the pixel-based segmentation method. The features such as run length, texture, and shape of the segmented lung tissue region are obtained, which describes the content of that specific segmented region.

Feature Selection (FS) involves selecting the best subset of features from the set of extracted features in feature extraction to improve the performance of the algorithm used for classification. This model for Feature Selection in this paper used two bio-inspired algorithms namely Paddy Field Algorithm (PFA) and Spider Monkey Optimization Algorithm (SMO). The ten-fold cross-validation technique is

used to train t SVM classifier with the optimal feature extracted using these algorithms. The model proposed in this paper achieved an accuracy of 93.74%.

In [10] images are extracted from the Cancer Imaging Archive database which stores the image in DICOM format. The first stage i.e. image pre-processing stage begins with image enhancement. In the image enhancement stage, they used the Gabor filter for image enhancement. After Image enhancement, they have used marker-controlled watershed image segmentation for image segmentation.

The Image features extraction is done by using algorithms to detect the region of interest or features of an image. The main objective is to obtain features such as area, perimeter and eccentricity of the enhanced and segmented image. For this purpose gray-level co-occurrence Matrix (GLCM) is used. For classification support vector machine is used. Support vector machine takes from the feature extraction stage and classifies the input CT scan image.

In [11] authors used a median filter for image pre-processing on gray-scaled CT scan images. A Gaussian filter is applied to the output images of the median filter. The work of the Gaussian filter is to remove speckle noise from the image and smoothes the image. In the segmentation stage, the watershed segmentation method is used to segment the image. In the feature extraction stage, functions in OpenCV are used to calculate eccentricity, area, perimeter, centroid, and diameter which are used as input to the classifier. In the final stage, the support vector machine is applied. This classifies the nodule as malignant or benign. This model achieved an accuracy of 92%.

In [12] they have used the original microarray cancer gene dataset as input. They have used the SMO algorithm for feature selection. For classification SVM is used as a classifier. The best gene subset calculated by spider monkey optimization is applied to the SVM classifier. The proposed method achieved an accuracy of 100%.

3. CONCLUSION

Every day a large number of people are dying due to lung cancer disease. Millions of people died till date. By each passing day the number of patients is increasing. So it becomes fruitful to detect it in early stages so that the death rate can be minimized. Also as the medical imaging is getting advanced the researchers are coming up with different approaches and models to detect it through computational intelligence.

Many of the research were using traditional ML algorithms while some of them are based on nature inspired phenomena like ACO, BFO, PSO, BFO etc for detection and classification. As the computational power is increasing day by day it is still open for researcher to propose optimum and valid solution to serve humanity.

REFERENCES

[1] Nawreen, Nusraat, Umma Hany, and Tahmina Islam. "Lung Cancer Detection and Classification using CT Scan Image Processing." In 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), pp. 1-6. IEEE, 2021.

[2] Agarwal, Aman, Kritik Patni, and D. Rajeswari. "Lung cancer detection and classification based on alexnet CNN." In 2021 6th International Conference on Communication and Electronics Systems (ICCES), pp. 1390-1397. IEEE, 2021.

[3] Firdaus, Qurina, Riyanto Sigit, Tri Harsono, and Anwar Anwar. "Lung Cancer Detection Based On CT-Scan Images With Detection Features Using Gray Level Co-Occurrence Matrix (GLCM) and Support Vector Machine (SVM) Methods." In 2020 International Electronics Symposium (IES), pp. 643-648. IEEE, 2020.

[4] Abdullah, Mohd Firdaus, Siti Noraini Sulaiman, Muhammad Khusairi Osman, Noor Khairiah A. Karim, Ibrahim Lutfi Shuaib, and Muhamad Danial Irfan Alhamdu. "Classification of Lung Cancer Stages from CT Scan Images Using Image Processing and k-Nearest Neighbours." In 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC), pp. 68-72. IEEE, 2020.

[5] de Pinho Pinheiro, Cesar Affonso, Nadia Nedjah, and Luiza de Macedo Mourelle. "Detection and classification of pulmonary nodules using deep learning and swarm intelligence." Multimedia Tools and Applications 79, no. 21 (2020): 15437-15465..

[6] Basha, B. Mohamed Faize, and M. Mohamed Surputheen. "The Lung Cancer Predictive Accuracy for Non-Smokers Using Classification and HFPS Algorithm." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12, no. 6 (2021): 3160-3171.

[7] Prabhakar, Sunil Kumar, Harikumar Rajaguru, and Dong-Ok Won. "A Holistic Performance Comparison for Lung Cancer Classification Using Swarm Intelligence Techniques." Journal of Healthcare Engineering 2021 (2021).

[8] Vasani, Khevna, And Ayushi Shah. "Lung Cancer Detection Using Ct Scan Images." (2021).

[9] Isaac, Anisha, H. Khanna Nehemiah, Snofy D. Dunston, VR Elgin Christo, and A. Kannan. "Feature selection using competitive coevolution of bio-inspired algorithms for the diagnosis of pulmonary emphysema." Biomedical Signal Processing and Control 72 (2022): 103340.

[10] Johora, Fatema Tuj, Mehdi Hassan Jony, Md Shakhawat Hossain, and Humayun Kabir Rana. "Lung Cancer Detection Using Marker Controlled Watershed with SVM." GUB Journal of Science and Engineering 5, no. 1 (2018): 24-30.

[11] Makaju, Suren, P. W. C. Prasad, Abeer Alsadoon, A. K. Singh, and A. Elchouemi. "Lung cancer detection using CT scan images." Procedia Computer Science 125 (2018): 107-114.

[12] Rani, R. Ranjani, and D. Ramyachitra. "Microarray cancer gene feature selection using spider monkey optimization algorithm and cancer classification using SVM." Procedia computer science 143 (2018): 108-116.

.