Proceedings of the Second International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2020)
IEEE Xplore Part Number: CFP20K58-ART; ISBN: 978-1-7281-4167-1

Lung Cancer Prediction using Machine Learning: A Comprehensive Approach

Syed Saba Raoof
Department of Computer Science & Engineering
Vardhaman College of Engineering
Hyderabad, India

M A. Jabbar
Department of Computer Science & Engineering
Vardhaman College of Engineering)
Hyderabad, India
jabbar.meerja@gmail.com

Syed Aley Fathima
Department of Computer Science & Engineering)
Vardhaman College of Engineering
Hyderabad, India

***Abstract*— The prominent cause of cancer-related mortality throughout the globe is "Lung Cancer". Hence beforehand detection, prediction and diagnosis of lung cancer has become essential as it expedites and simplifies the consequent clinical board. To erect the progress and medication of cancerous conditions machine learning techniques have been utilized because of its accurate outcomes. Various types of machine learning algorithms(ML) like Naive Bayes, Support Vector Machine (SVM), Logistic regression, Artificial Neural Network (ANN), have been applied in the healthcare sector for analysis and prognosis of lung cancer. In this review, factors that cause lung cancer and application of ML algorithms are discussed up to date and also draws special attention to their relative strengths and weaknesses. This paper will help the researchers to quickly go through the related literature instead of referring to the many papers.**

Keywords— Lung Cancer, Machine Learning, SVM, ANN, NLP.

I. INTRODUCTION

Lung cancer is a harmful disease that causes a huge number of deaths globally. The primal encounter of lung cancer is necessary to decrease the mortality rate of patients. Thus it is a great challenge encountered by doctors and researchers to detect and diagnose lung cancer. Detection of lung cancer can be done by using medical images such as computed tomography, chest X-ray; MRI scans, etc., ML approaches recognize the main characteristics of complex lung cancer datasets. A CAD (Computer-Aided Diagnosis) was developed in the early 1980s to enhance the survival rate and efficiency that aid the doctors in interpreting medical images. Some of the machine learning algorithms that have a profound impact in health care are decision trees, linear regression, random forest, SVM, naive Bayes, K-nearest neighbors and so on. We have also discussed the deep learning methods techniques and algorithms that can be implemented for diagnosis, detection, and prediction of various cancers. The preeminent intent of this research work is to present a concise vision of present work on different cancers and mainly lung cancer prediction using deep learning and machine learning models.

Symptoms are categorized based on the location and size of the tumor [35]. During the early stages, it's difficult to analyze and detect as it will not any cause any pain and symptoms in some cases. Lung cancer diagnosed patient may suffer through Cough, Chest pain, Shortness of breath, Wheezing, Hemoptysis i.e. coughing up blood, Pancoast syndrome (shoulder pain), Hoarseness (paralysis of vocal cords), Weight loss, Weakness, and Fatigue. Types of lung cancer are pictorially shown in figure 1.

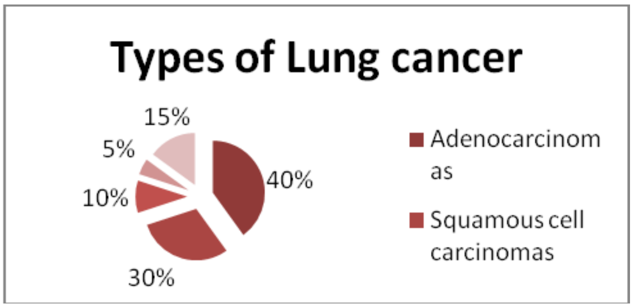

Figure 1: Types of lung cancer [1]

90% of it is induced due to smoking. Impregnation of tobacco smoke also causes lung cancer i.e. known as passive smoking. Another factor for lung cancer is heredity. vehicular pollution, industries, the intake of harmful gases such as Radon stands at the second position in causing the deaths from lung cancer. Factors causing lung cancer and mortality rate are shown in table 1.

Table I. Factors Causing Lung Cancer and Mortality Rate [2]

Causes	*Mortality in%*	*Figure*
Cigarette smoking	90%	70000 (USA)
Radon gases	12%	21000
Passive smoking	2-4%	-----

Physicians detect the presence and stage of cancer by provoking various tests such as X-ray, CT scan, bone scans, MRI scans, and PET scans. NSCLC is branched into four stages based on severity: Stage 1 Limited to lungs, Stage 2 Limited to the chest, and Stage 3 Limited to the chest however amidst to larger and major aggressive tumors and Stage 4 spread to different regions in the body. SCLC type medication is determined by the two-tiered model: Limited stage SCLC, Extensive stage (ES) SCLC. [1]

Globe pattern of lung cancer caused all over the world [3] are listed in Table 2.Lung cancer deaths are also rising in India. Statistics of lung cancer deaths are listed in table 3.

Table II. lung cancer cases and deaths over the globe [3]

Region	*Population*	*Cancer Cases*	*Deaths*
Asia	60%	66%	57.3%
Europe	9.0%	23.4%	20.3%
America	13.3%	21.0%	14.4%
Africa	17%	9-10%	7.3%

TABLE III LUNG CANCER CASES AND DEATHS IN INDIA [3]

Lung Cancer	*New Cases*	*Deaths*
Men	60%	66%
Women	9.0%	23.4%
Both cases	13.3%	21.0%

II. MACHINE LEARNING/ ARTIFICIAL INTELLIGENCE IN HEALTHCARE

ML/AI use complicated and sophisticated algorithms for the appraisal and analysis of human perception and cognition. Computer algorithms can approach cessations regardless of user input. The intelligence of Ai algorithms of acquiring the information, processing it and deciding/finalizing the accurate output discriminates the traditional technologies in Ai. These algorithms are implemented using ML techniques. AI in healthcare is branched into two categories based on the type of data I.e. analysis of structured data along with images, genes and biomarkers are done with ML techniques. The unstructured data such as prescription notes, medical magazines or journals are analyzed using natural language processing (NLP) methods. First, the input text acquired is converted to binary format using NLP methods then this binary data is analyzed by ML techniques to produce accurate output and decisions. [4]

Cancer, neurology, cardiology are the major parts of medicinal studies where AI is implemented. As this disease are superior in the mortality rate. Apart from these diseases, AI is even applied to other medicinal areas for prediction, analysis, and curing. Predominantly renowned ML algorithms extended in the healthcare sector are SVM, NN, random forest, logistic regression, discriminate analysis, decision trees, linear regression, nearest neighbor, naive bayes, etc.

The elite A.I. algorithms that can currently be initiated in healthcare are as follows:

- The algorithm detecting variation in a tumor.
- Classification of heart images.
- Heart attack predicting algorithm.
- More precise skincare cancer diagnosis with AI.
- AI system for ICU.
- Computers detecting breast cancer risk.
- AI. useful to diagnose breast cancer
- Smart algorithm predicting suicide risk
- Inpatients mortality can be predicted by AI

III. MACHINE LEARNING ALGORITHMS

Almost effective image-based ML systems are developed on SVM, linear regression, decision trees, KNN and so on. [5]

i. **SVM:** For prediction, regression and classification the most prominent method employed is SVM. It classifies the input data set by introducing a boundary called a hyperplane that separates the dataset into two parts. The favorable asset of SVM is, as SVM is a data-driven approach and feasible without a hypothetical scheme that produces an accurate classification. Particularly when the size of the sample is small. SVMs are broadly used for classification when the datasets are biomarkers, to predict and diagnose cancer, neurological and cardiology diseases.

ii. **ANN:** ANN is a computational method that is compromised profoundly with interconnected processing components called neurons, which sort out information as feedback to external stimuli. It is accomplished in two modes, i.e. learning and testing. Learning is used to classify new input. Attesting phase, it receives an input signal from the network and computes it to generate an output. In diversified fields of healthcare, ANN approaches are useful for example in diagnosis and indicating breast cancer, lung cancer, and other ontology prediction, diagnostic systems, drug analysis, etc.

iii. **Recurrent Neural Network (RNN):** This algorithm performs a similar function for all the elements in a sequence and the output relies on the information of previous outcomes. It uses a memory that stores the information of previous outcomes.

IV. LITERATURE REVIEW

Janee et al [6]; proposed “multi-stage lung cancer detection and prediction using multi-class SVM classifier” describes an effective algorithm using SVM for lung cancer detection, diagnosis and it's also capable to anticipate the possibility of lung cancer. An algorithm is developed using MATLAB and employing image processing techniques like image enhancement, detection and segmentation, detection and feature extraction is done by adopting the Gray Level Co-occurrence Method (GLCM) method. SVM is applied for classification purposes. Prediction is done through the binarization technique. UCI ML database is acquired, which includes 500 infected and 500 non-infected CT images. The proposed method detected 126 images as infected out of 130 and predicted 87 images as cancerous out of 100 formerly specified images. The experimental analysis shows 97% precision in the identification and 87% for prediction.

M.Gomathi et al [7]; defined "A Computer-Aided Diagnosis System for Detection of Lung Cancer Nodules using Extreme Learning Machine" In this paper CAD model is developed for the analysis of cancer in Ct images. The fundamental phase of CAD is to detect the region of interest in input CT images. The lung region extraction is preceded with lung region segmentation, detection of cancer nodules is accomplished with Fuzzy Possibility CMean (FPCM) clustering algorithm. A maximum Drawable Circle intensity value is used for the formulating the diagnostic rules. Then these rules are implemented to learn along with the assist of the ELM (Extreme Learning Machine).

Shingo kakeda et al [8]; proposed a commercial CAD model for detecting lung nodules on Chest films. This paper presents the CAD model, which comprises an image server and EpiSight/XR software. This method is carried out in four fundamental steps. The first complex automatic structure is reduced to produce different images. Nodule candidates are detected by employing the Multiple Gray Level thresholding methods. To differentiate the true nodule with false-positive nodule features are input and difference images are used to extract features. Formerly extracted features are utilized for the reduction of false-positive nodules; a rule-based analysis and ANN are employed for this purpose. For experimentation, the model developed was analyzed on a database that consists of 274 radiographs with 323 lung nodules. Out of 315, 235(75%) false-positive images were detected as a normal automatic structure and 155(49%) are detected as pulmonary vessels.

Metin N et al [9]; proposed "Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided diagnosis system" The proposed method is carried out in five steps, first regions are segmented by applying k-means clustering method. After segmenting the lung curves, by employing pre-processing algorithms suspicious regions are segmented from that of lung regions that produce a binary image which consists of holes due to the segmentation process, flood –the filled algorithm is used to fill this hole as nodule-candidate are considered as solid objects. This system may contain general regions and lung nodules comprising of blood vessels. To differentiate this nodule, Rule-based classifiers with 2D and 3D features are used. At last, the false positive objects are detected using Linear Discrimination Analysis (LDA). The proposed method was analyzed on a dataset that includes 1454CT images gathered from 34 diagnosed patients with 63 lung nodules.

Kazak Awai et al [10]; described a system for evaluating the CAD effect on radiologist's pulmonary nodules detection. The proposed method used image processing techniques for lung and intrapulmonary structure segmentation gray-level threshold, 3D labeling techniques and mathematical morphological techniques are used for lung segmentation. For the segmentation of intrapulmonary structures Top-hat transformation method is employed, on an input image to detect the smoothed image. Sieve filter is used for the identification of primary potential nodules, then features of these pulmonary nodules are extracted to differentiate true nodules from that of false-positive nodules, ANN is adapted to decide the probability of region of interest based on an image feature. Adaption of this system enhanced pulmonary nodule resident's detection of CT scans.

Cheran et al [11]; proposed "Computer-aided diagnosis for lung CT using artificial life models" This CAD model is developed by deploying various algorithms. First, the ribcage region is determined by employing a 3D region growing algorithm, and then the active contour technique is implemented to develop a specific area for the approaching ants which are redistributed to develop a specific and precise rebuilding of the vascular tree and pleura. To regenerate the bronchial and the vascular trees artificial life models are used. By utilizing active shape models, it is determined whether the previously constructed branches contain nodules and also to detect whether the nodules are connected to the pleura. By employing snakes and dot enhancement algorithm cleaner algorithm is produced to confine the nodules.

Lakshmanaprabu et al [12]; developed the Optimal Deep Neural Network (ODNN) and Linear Discriminate Analysis (LDA) based classification model for CT images. Lung nodule classification is done by applying LDR and optimization is done by applying the Modified Gravitational Search Algorithm to predict lung cancer. Standard CT database is used for experimental analysis which comprises 50-low dosage lung cancer CT images. This model is correlated with existing models such as KNN, NN, DNN SVM, and so on., and the experimental analysis shows best results for the developed model with 94.56% accuracy, sensitivity and specificity 96.2% and 94.2% respectively.

Worawate et al [13]; developed a method "Automatic Lung Cancer Prediction from Chest X-ray Images using Deep Learning Approach", Authors used DensNet-121 (121 layers Convolutional neural network) in conjunction with transfer learning for classifying using chest images. Model is trained on two datasets i.e. Chest X-ray 14 and JSRT to identify the nodules. The model obtained an accuracy of about 74.43±6.01%, sensitivity, and specificity of about 74.68 ± 15.33% and 74.96 ± 9.85% respectively.

Christoph et al [14]; proposed a prediction model based on tomography lung cancer images. A CNN is utilized for feature extraction by fine-tuning pre-trained ResNet18 and multimodal features CNN is trained by the Cox model for hazard prediction. Lung1 dataset is used for experimental analysis which can be accessed from "The Cancer Imaging Archive" (TCIA) that comprises 422 NSCLC (Non-Small Cell-Lung Carcinoma) images for 318 of 422 patients.

Jason et al [15]; developed a Deep Screener algorithm which is a form of different deep learning approaches. The Deep Screener is an end-to-end automated screening of lung cancer on low dose CT images. TCIA dataset is used for experimental analysis which comprises of 1449 low dose CT images. The model developed was correlated with the grt123 algorithm of Data Science Bowl 2017 for lung cancer analysis and the result was too close wining algorithm grt123. The proposed model predicted about 1359 out of 1449 CT scans about 82% accuracy, AUC about 0.885, AUPRC about 0.837.

In table 4: We have reviewed different lung cancer prediction & detection research papers and listed out the algorithms and Datasets and measures used in those papers.

V. DEEP LEARNING IN HEALTHCARE

This section intends to give a summary of the research direction in DL for health care. And we had discussed the DL concepts, algorithms, techniques, approaches and applications [16].

Deep learning algorithms specifically CNN, Fully Connected Convolutional Networks (FCN's), Deep Belief Networks (DBN's) had promptly evolved techniques and strategies to study and examine/analyze the imaging in medical area like MRI, X-Ray and computed tomography(CT) images, etc., [17,18,19] Deep learning approaches are used for image classification, lesion classification and detection, organ and lesion segmentation, enhancement and image generation and it can be also used for combining image data with reports.

In medical image analysis of CT images and X-rays, the prediction, detection, labeling, and classification has become the most prominent challenging application. For example in the application of lung cancer analysis, detection, large dataset alike LUNA16, Lung1 are utilized to train the model that integrates deep learning approaches like CNN, DBNs for image analysis and NLP, RNN's are applied for text analysis.

A. Algorithms

a. Convolutional Neural Network

CNN is influenced by the biological process, where the animal optical cortex system representation is used to display the connectivity pattern between the neuron elements. This algorithm is appropriate for a classifying multi-class problem and binary classification (for example — predicting whether a medical image has a malignant tumor or not). It comprises of a convolutional layer, activation, pooling, and output layer. The CNN changes the pel values of input CT image through weighing in convolutional layers and sub-sampling layers are sampled alternatively. The outcome of this algorithm is a recursive function with weighted input values. CNN in the medical area is employed auspiciously to aid disease diagnosis like breast cancer, lung cancer, cardiovascular diseases, neurological disorders, etc. The working of CNN is shown in figure 1.

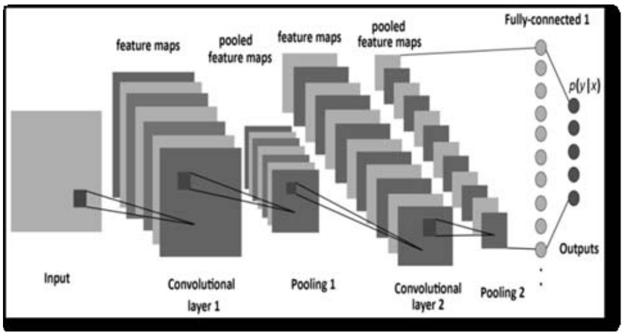

Fig. 2. Working on CNN

b. Fully Convolutional Networks (FCN's)

FCN's shown in figure 3 employs CNN to change the image pixel values into the pixel categories. The main variation among FCN's and CNN is that deconvolutional and upsampling layers in CNN are replaced with a fully connected layer in FCN. It generates a score map that is similar to the size of the input image for each class and it also categorizes every pixel value of the image. To enhance the prediction accuracy this algorithm merges the preceding layer results along with end layers deep convolution and upsampling layer results.

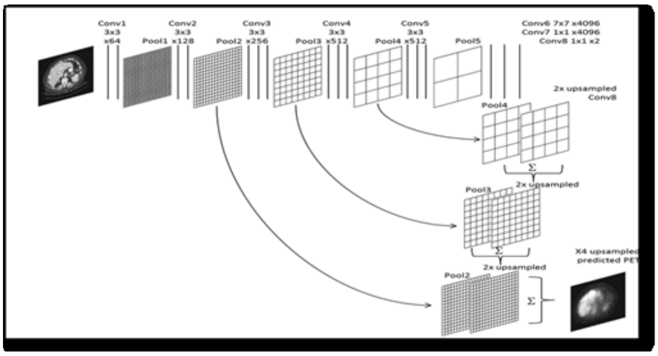

Fig. 3. Working with FCN's

c. Auto-Encoders (AE's)

Auto-Encoder shown in figure 4 is a standard of Neural Network that may be feasibly applicable for unsupervised data. It comprises three layers i.e. input, hidden, and output layer. Training of this model is accomplished in two stages i.e., encoding and decoding stage. Encoding is done in a hidden layer that describes a code that is used to define the input and decoding is done in the output layer.

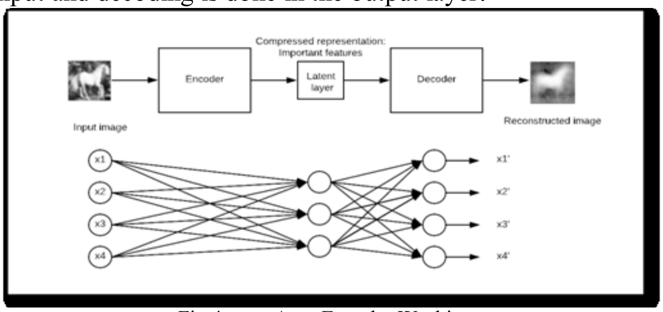

Fig.4. Auto Encoder Working

DBN is probability innovative system (shown in figure 5) that is developed by the pyramid of Restricted Boltzmann Machine (RBM) as a substitute for Autoencoder. It consists of two layers, one is visible and the other is a hidden layer. The RBM energy function is given as:

$$\mathbf{E\ (v,\ h) = \text{-}a,} \quad \mathbf{(1)}$$

Where a & b are the bias vector values for both the layers. To maximize the visible vector probability the RBM model is trained based on energy function as follows:

$$argmax_{w,a} P(v) = \frac{1}{2} \sum_{n} (-\mathbf{E}(\mathbf{v}, \mathbf{h})) \quad (2)$$

Where partition function is z. The RBM optimization is accomplished by applying a contrastive divergence algorithm.

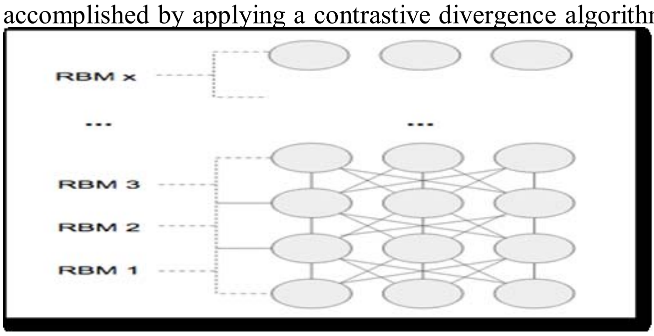

Fig. 5. DBN's Architecture

VI. DL In Cancer Diagnosis And Detection

This section provides a broad survey on different cancer prediction, diagnosis, and detection using deep learning approaches, techniques, algorithms and datasets used for experimental analysis. Summary of Lung cancer prediction using machine learning by various researchers is shown in Table 4 and in table 5.

VII. Conclusion

In this paper, we demonstrate a survey on lung cancer, its causes, symptoms, mortality rate due to cancer in India and throughout the Globe and deliberates the machine learning techniques, its applications in healthcare and cancer prognosis and detection. Most of the researchers developed the cancer prediction systems based on a supervised learning technique of ML and classification algorithms to produce an accurate outcome. Deep learning in health care and algorithms are emphasized. Prediction and diagnosis of the Lung cancer system can be embellished and extended further by employing deep learning techniques to enhance the accuracy of both identification and prediction of lung cancer. This paper will help the researchers to look insight into various ML techniques applied to lung cancer. In the future, we want to apply deep learning techniques to predict lung cancer.

References

[1] Lung Cancer Symptoms, Types, Causes, Treatment & Diagnosis, https://www.medicinenet.com/lung_cancer/article.html.

[2] ICMR-National Institute of Cancer Prevention & Research (ICMR-NICPR).

[3] The International Agency for Research on Cancer (IARC). Latest global cancer data.

[4] Top A.I Algorithms in Health Care-The Medical Futurists., 2019, https://medicalfuturist.com/top-ai-algorithms-healthcare/

[5] G. Lakshmanaprabu S.K., Sachi Nandan Mohanty, Shankar K., Arunkumar N., Gustavo Ramirez., 2018. Optimal Deep Learning Model for Classification of Lung Cancer on CT Images. Future Generation Computer Systems,

[6] Janee Alam, S., & Hossan, A. Multi-Stage Lung Cancer Detection and Prediction Using Multi-class SVM Classifier. 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)...

[7] M.Gomathi, Dr.P.Thangaraj. 2010. A Computer-Aided Diagnosis System for Detection of Lung Cancer Nodules using Extreme Learning Machine. International Journal of Engineering Science and Technology. Vol. 2(10), 2010, 5770-5779.

[8] Kakeda, S., Moriya, J., Sato, H., Aoki, T., Watanabe, H., Nakata, Doi, K. 2004. Improved Detection of Lung Nodules on Chest Radiographs Using a Commercial Computer-Aided Diagnosis System. American Journal of Roentgen logy, 182(2), 505–510. doi:10.2214/ajr.182.2.1820505.

[9] Gurcan, M. N., Sahiner, B., Petrick, N., Chan, H.-P., Kazerooni, E. A., Cascade, P. N., & Hadjiiski, L. 2002. Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided diagnosis system. Medical Physics, 29(11), 2552–2558. doi:10.1118/1.1515762.

[10] Awai, K., Murao, K., Ozawa, A., Komi, M., Hayakawa, H., Hori, S., & Nishimura, Y. 2004. Pulmonary Nodules at Chest CT: Effect of Computer-aided Diagnosis on Radiologists' Detection Performance. Radiology, 230(2), 347352. doi:10.1148/radiol.2302030049.

[11] Cheran, S. C., & Gargano, G. 2005. Computer-aided diagnosis for lung CT using artificial life models. Seventh International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC'05). doi:10.1109/synasc.2005.28.

[12] S.K., L., Mohanty, S. N., K., S., N., A., & Ramirez, G. 2018. The optimal deep learning model for classification of lung cancer on CT images. Future Generation Computer Systems. doi:10.1016/j.future.2018.10.009.

[13] A.Worawate. Thirach, A., Marukatat, S., & Wilaiprasitporn, T. 2018. Automatic Lung Cancer Prediction for Chest X-ray Images Using the Deep Learning Approach. Biomedical International Conference (BMEiCON-2018).

[14] Haarburger, C., Weitz, P., Rippel, O., & Merhof, D. 2019. Image-Based Survival Prediction for Lung Cancer Patients Using CNN. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). doi:10.1109/isbi.2019.8759499.

[15] Casuey,l, J., Guan, Y., Dong, W., Walker, K., Qualls, A, J., Prior, F., & Huang X. 2019. Lung Cancer Screening With Low-Dose CT Scans Using A Deep Learning Approach. ArXiv:1906.00240 [eess.IV]

[16] Hu, Z., Tang, J., Wang, Z., Zhang, K., Zhang, L., & Sun, Q. (2018). Deep learning for image-based cancer detection and diagnosis − A survey. Pattern Recognition, 83, 134–149. doi:10.1016/j.patcog.2018.05.014.

[17] Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and Structural Biotechnology Journal, 13, 8–17. doi:10.1016/j.csbj.2014.11.005.

[18] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88. doi:10.1016/j.media.2017.07.005.

[19] Alkadi, R., Taher, F., El-baz, A., & Werghi, N. (2018). A Deep Learning-Based Approach for the Detection and Localization of Prostate Cancer in T2 Magnetic Resonance Images. Journal of Digital Imaging. doi:10.1007/s10278-018-0160-1.

[20] X. Zhao., Y. Wu., G. Song., Z. Li., Y. Zhang., Y. Fan. 2017. A Deep Learning Model Integrating FCNNs and CRFs for Brain Tumor Segmentation. arXiv preprint arXiv:1702.04528.

[21] M. Havaei., A. Davy., D. Warde-Farley., A. Biard., A. Courville., Y. Bengio., C. Pal, P.-M. Jodoin, H. Larochelle. 2017. Brain Segmentation With Deep Neural Networks. Medical image analysis, 35 (2017) 18-31.

[22] D. Paredes., A. Saha & M.A. Mazurowski. Deep Learning for Segmentation of Brain Tumors: Can we train with images from different institutions? SPIE Medical Imaging, International Society for Optics and Photonics2017, pp. 101341P-101341P-101346.

[23] K.B. Ahmed., L.O. Hall., D.B. Goldgof., R. Liub., & R.A. Gatenby. 2017 Fine-Tuning Convolutional Deep Features for MRI Based Brain Tumor Classification. Proc. of SPIE Vol2017, pp. 101342E-101341.

[24] S. Suzuki., X. Zhang., N. Homma., K. Ichiji., N. Sugita., Y. Kawasumi., T. Ishibashi., & M. Yoshizawa. 2016. Mass Detection Using Deep Convolutional Neural Network for Mammographic Computer-Aided Diagnosis. Society of Instrument and Control Engineers of Japan (SICE). 55th Annual Conference of the IEEE2016, pp. 1382-1386.

[25] M.G. Ertosun., D.L. Rubin. 2015. Probabilistic Visual Search for Masses Within Mammography Images Using Deep Learning. Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on, IEEE2015, pp. 1310-1315.

[26] N. Dhungel., G. Carneiro., & A.P. Bradley. 2015. Deep Structured Learning for Mass Segmentation from Mammograms. Image Processing (ICIP), 2015 IEEE International Conference on, IEEE2015, pp. 2950-2954.

[27] H. Chen., Q. Dou., X. Wang., J. Qin & P.-A. Heng. 2016. Mitosis Detection in Breast Cancer Histology Images via Deep Cascaded Networks, AAAI2016, pp. 1160-1166.

[28] Y. Xu., T. Mo., Q. Feng., P. Zhong., M. Lai., I. Eric & C. Chang. 2014 Deep Learning of Feature Representation with Multiple Instance Learning for Medical Image Analysis. Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on, IEEE2014, pp. 1626-1630.

[29] L. Yu., H. Chen., Q. Dou., J. Qin & P.A. Heng. 2017. Integrating Online and Offline Three-Dimensional Deep Learning for Automated Polyp Detection in Colonoscopy Videos. IEEE Journal of biomedical and health informatics, 21 (2017) 65-75.

[30] E. Gibson., M.R. Robu., S. Thompson., P.E. Edwards., C. Schneider., K. Gurusamy., B. Davidson., D.J. Hawkes., D.C. Barratt & M.J. Clarkson. 2017. Deep Residual Networks for Automatic Segmentation of Laparoscopic Videos of the Liver. SPIE Medical Imaging, International Society for Optics and Photonics2017, pp. 101351M-101351M-101356.

[31] W. Li., F. Jia & Q. Hu. 2015. Automatic Segmentation of Liver Tumor in CT Images with Deep Convolutional Neural Networks. Journal of Computer and Communications, 3 (2015) 146.

[32] Z. Tian., L. Liu & B. Fei. 2017. Deep Convolutional Neural Network for Prostate MR Segmentation. SPIE Medical Imaging, International Society for Optics and Photonics2017, pp. 101351L-101351L-101356.

[33] L. Maa., R. Guoa., G. Zhanga., F. Tadea., D.M. Schustera., P. Niehc., V. Masterc., B. Fei. Automatic Segmentation of the Prostate on CT Images Using Deep Learning and Multi-Atlas Fusion. SPIE Medical Imaging, International Society for Optics and Photonics2017, pp. 101332O-101332O-101339. Internet: https://www.flickr.com/, [Mar. 28, 2018].

[34] E. Nasr-Esfahani., S. Samavi., N. Karimi., S.M.R. Soroushmehr., M.H. Jafari., K. Ward & K. Najarian. Melanoma Detection by Analysis of Clinical Images Using Convolutional Neural Network. Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the IEEE2016, pp. 1373-1376.

[35] S. Demyanov., R. Chakravorty., M. Abedini., A. Halpern & R. Garnavi. 201. Classification of Dermoscopy Patterns Using deep Convolutional Neural Networks. Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on, IEEE2016, pp. 364-368.

TABLE IV SUMMARY OF LUNG CANCER PREDICTION.

S. No	Paper title	Method used	Dataset	Pros/cons	Measure
1	A survey on object detection Multi-stage lung cancer detection and prediction using multi-class SVM	Multicase SVM	UCI MLDb	The detection rate is high but the prediction rate is low. The method is applied to the traditional classifier.	97%- identification 87%- prediction
2	A CAD system for the detection of lung cancer using an extreme learning machine	ELM	Data collected from a reputed hospital with 100 lung images	Compared with the conventional CAD system proposed method gives excelling detection	Able to detect false positive nodules correctly.
3	Improved Detection of Lung Nodules on Chest Radiographs Using a Commercial Computer-Aided Diagnosis System	ANN	Large Database with 274 radiographs and 323 lung nodules	The false-positive detections seem to be increased by these preexisting diseases.	73% detection sensitivity with 4.0 false-positive Detections per image.
4	Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided diagnosis system	Rule-based classification, LDA	Institutional Review Board with 1454CT images gathered from 34 diagnosed patients with 63 lung nodules.	Improved the sensitivity of cancer detection, reduce oversight Errors, and decrease inter- and intra-reader variations.	84% sensitivity
5	Pulmonary Nodules at Chest CT: Effect of Computer-aided Diagnosis of Radiologists' Detection Performance	Top-hat transformation method, sieve filter.	An institutional review board of Rinku General Medical Center with 133 CT	Improve both the residents' and the board-certified radiologists' performance In detecting	80%- true-positive rate

			images	pulmonary nodules.	
6	Computer-Aided Diagnosis for Lung CT Using Artificial Life Models	Artificial life method.	CT images composed of series 2D files in DICOM format	Artificial life models can be more precise virtual ants. If biological aspects are considered.	Improved accuracy
7	Optical Deep Learning Model for Classification of Lung Cancer on CT images.	ODNN & LDA	Standard CT database with 50 CT images	Manual labeling time is minimized. Accuracy & precision are enhanced.	Accuracy-94.56%, sensitivity-96.2% & specifity-94.2%
8	Automatic Lung Cancer Prediction From Chest X-Ray Images using Deep Learning approach.	DenseNet-121	Chest X-ray 14 & JSRT dataset	High accuracy is achieved and heatmap is implemented to identify the lung nodule region. Attention-Guided CNN (AG-CNN) can be used to identify the malignancy.	Accuracy-74.43±6.01, sensitivity-74.68±15.33% & specificity-74.96±9.85%.
9	Image-Based Survival Prediction for Lung Cancer Patients using CNNs	ResNet18 & CNN	Lung1 dataset-TCIA	CNN is much harder to interpret when compared with the Cox model.	CNN is much harder to interpret when compared with the Cox model. 0.623±0.039 prediction
10	Lung Cancer Screening with Low-Dose CT Scans using a Deep Learning Approach	Deep Screener algorithm	TCIA-1449 low-dose CT images	A false-positive rate is efficiently is reduced.	82% of accuracy

TABLE V SUMMARY OF VARIOUS ALGORITHMS USED IN LUNG CANCER PREDICTION

Cancer type	**Author**	**Input Images**	**Deep learning Model**	**Dataset**
Brain cancer	Zhao et al[20] Havaei et al [21] Paredes et al [22] Ahmed et al[[23]	MRI MRI MRI MRI	FCN CNN CNN CNN	BRATS BRATS TCIA Unpublished dataset
Breast cancer	Suzuki et al[24] Ertosun et al[25] Dungel et al[26] Chen et al[27]	PEM Mammography Mammography Histopathology	CNN CNN DBN FCN & CNN	DDSM DDSM DDSM & INbreast MITOSATYPIA-12
Colonial cancer	Xu et al[28] Yu et al[29]	Histopathology Colonoscopy	CNN FCN	Unpublished dataset Mayo clinic
Liver Cancer	Gibson et al[30] Li et al[31]	Laparoscopy CT scan	CNN CNN	Unpublished dataset Unpublished dataset
Lung Cancer	Christoph et al[14] Lakshmanaprabu et al [12] Worawate et al[13] Jason et al[15] Janee et al[6] Gomathi et al[7]	CT images CT images Chest X-Ray CT scans CT images CT scan	ResNet18 & CNN ODNN & LDA DenseNet-121 Deep Screener SVM ELM	Lung1 & TCIA Standard CT database Chest X-ray 14 & JSRT TCIA UCI MLDb Large dataset
Prostate cancer	Tian et al[32] Maa et al[33]	MRI MRI	FCN CNN	Unpublished dataset Unpublished dataset

Skin cancer	Nasr et al[34]	Clinical photography	CNN	MED-NODE
	Demyanov et al[35]	dermoscopy	CNN	ISIC