ISBN: 978-1-948012-17-1

Ethics and Information Technology (ETIT)

DOI: <http://doi.org/10.26480/etit.02.2020.121.126>

CrossMark

INTERSTITIAL LUNG DISEASES CLASSIFICATION USING FEATURE EXTRACTION AND CONVOLUTION NEURAL NETWORK

Ayush Gupta, Chetan Patidar, Govind Patidar, Anusha Jain

Medi-caps University Indore, M.P.

*Corresponding Author Email: ayush6240@gmail.com

This is an open access article distributed under the Creative Commons Attribution License CC BY 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ARTICLE DETAILS

Article History:

Received 25 October 2020
Accepted 26 November 2020
Available online 03 December 2020

ABSTRACT

The main motive of this proposed work is to classify the interstitial Lung Diseases using convolution neural network. This proposed work technique used for tissue categorization of Interstitial Lung Diseases. The planned framework includes 5 complex layers and 2 fully connected layers. This work based on the different layers on convolution neural network. The final layer provides the output of various tissue patterns, such as health, ginger bread, opaque frosted glass (GMO), mesh structure, etc. We use HRCT data sets collected from various radiation centers for system exercise or assessment. In the future, we can use three-dimensional imageries scanned using CT, or we can integrate this system into a computer-aided diagnostic system (CAD) that will help the radiologist make a better diagnosis. In this analysis, a complex neural network for classifying lung patterns was proposed and evaluated. Seven ILD modes. To sleeper or assess CNNs, we used the first deep CNN designed for exact problems. Finally, we classified performance (91%), demonstrating the potential of CNN in lung type analysis. Here, we propose a deep, intricate neural network that divides CT plaques of dependent lungs into 7 categories, including 6 ILD models of truly healthy tissues. Even with further improved performance, the method can be easily trained under other lung conditions with structure.

KEYWORDS

ILD, GLCM, LBP, CNN, Feature Extraction, Classification.

1. INTRODUCTION

Lung Disease (ILD), also known as parenchyma disease, is a common lung disease, which is linked to the taste of the lungs. ILD can cause morbidity and mortality, and is considered a major health problem (Heitmann et al., 1997; Uppaluri et al., 1999; Krizhevsky et al., 2012). The most notable features of ILD include the swelling and degeneration of the interstitial and lungs. The electrode shows the epithelial area of the alveolar sac (Tulder et al., 2014). Often, interstitial and thickening of the lungs leads to impaired lung function (Bondfale et al., 2018). The lung parenchyma comprises the alveolar wall, Cancer is one of the most widespread diseases in the world and is a result of genetic, environmental and lifestyle factors. It is one of the most common diseases in India, with the highest annual mortality rate up to 0.3% (Bondfale et al., 2018). For radiologists, oncologists and pathologists, proper screening and screening programs are the primary place for cancer monitoring. Finding the first cancer plays an important role in cancer screening and can improve the long-term survival of people. Medical interventions such as X-rays, MRI (magnetic resonance imaging) and CT (computed tomography) are crucial for cancer detection and diagnostic techniques used by medical staff. As we all know, medical awareness has been widely used in cancer detection, surveillance and follow-up (Heitmann et al., 1997). Extraction of cancer cells from microscopic biopsy images is acceptable and may vary from doctor to physician depending on their knowledge and other factors, including the absence of accurate labeling. And accurate image classification as normal or cancerous. Method To mitigate the above issues, if a method that is biologically interpretable and based on important clinical features is used to identify cancer cells, cancer cells may be identified immediately by

'microscopic images and give better results.

In computer-based schemes with medical delivery, machine learning techniques are widely used in cancer detection and diagnosis. To use machine learning technology, feature removal is often the main step. For different medical imaging techniques and changed types of cancer, different imaging modalities have been studied (Uppaluri et al., 1999; Krizhevsky et al., 2012; Tulder et al., 2014; Li, 2014). For example, for breast cancer surveillance, bilateral image extraction, Laplacian Gaussian diffraction and Gaussian filtering were used as instruments for identifying areas of quality in mammograms. Previous research has focused on the design of fine feature descriptors and the study of therapeutic approaches with machine learning techniques. Methods based on feature extraction have many disadvantages. Weaknesses limit further improvements in operating CAD systems. To counteract the weaknesses and improvements of the CAD system environment, the importance of the study of representation has been stressed in recent years rather than highlighting engineering (Gao et al., 2015). Deep learning is a learning process that can learn to represent hierarchical forms of image data. One of the benefits of deep learning is that it can create high-resolution images straight from the original image. In addition, with the support of image sharing groups (GPUs), parallel learning technologies have improved the accuracy of many fields within a few years ago. For example, recent research shows that CNN can guarantee some success in cancer detection and detection (Simonyan et al., 2015).

2. RELATED WORK

In recent years machine learning has captured the imagination of society

Quick Response Code	This paper was presented at	Access this article online	
	International Conference on Contemporary Issues in Computing (ICIC-2020) - Virtual IETE Sector V, Salt Lake, Kolkata From 25th-26th July 2020	Website: www.intelcomp-design.com	DOI: 10.26480/etit.02.2020.121.126

and has capacity to provide better lives to us. Deep learning is playing an important role for disease diagnosis in health care sector. In India the large numbers of patients are diagnosing with cancer and large volume of data generated during the treatment of cancer, there is a specific interest in the healthcare to use the application of artificial intelligence to improve oncologic care. Artificial intelligence originally invented in the 1950s, it is the ability of a machine to perform tasks commonly related with intelligent human behavior (Al-Jumaily et al., 2015). Artificial intelligence (AI) is an area of research in the field of research and development technologies that can solve complex tasks in ways that require human perception of human beings (Wolf, 2018).

In this area of AI, machine learning (ML) has made significant progress in the past decade. Many scholars believe that ML can solve real problems. In the field of mechanism learning, deep learning is a form of computation in which algorithms provide programming by learning from the many examples of desired behavior without having to specify exactly what clear rules (Latif, 2019).

In this context, this article presents the first application of AI technology in cancer prevention. This technology may produce a paradigm definition at the oncological level, thus improving the detection and mitigation of these diseases (Schchidan et al., 2017).

3. ILD PATTERN CLASSIFICATION

Lung cancer is a major cause for deaths in India due to cancer and through periodic assessment of individual most of these deaths may be prevented and can reduced the chances of lung cancer.

Early detection of lung cancer is done by the various imaging techniques like chest x-ray, CT Scan and MRI. After diagnosis of these imaging reports medical practitioners can found the tumors are cancerous or not cancerous. Early detection can improve survival rate of human as compared to the tumors found at advance stage.

Many researchers have done their research for detection of cancer using machine learning but the hit ratio of early detection is not as much improved (Selvakumar et al., 2017). The danger of taking lung growth is at least 20 times higher in smokers than in non-smokers. Due to evolution and different treatments lung cancer is classified in 2 categories:

1. Non-small cell carcinoma (NSCC): Most of the people who are taking lung cancer have NSCC. Peoples having more than 65 years age and who smoke or who respire a lot of burn are most probable to get non-small cell carcinoma.
2. Small cell carcinoma (SCC): In small cell carcinoma cells of the lung start growing rapidly in an uncontrolled manner. In recent years, researchers have suggested using data-driven features that can adapt to specific problems. Greatest of these means rely on uncontrolled systems such as the outlined packages (Ricci et al., 2001), and the standard form model Boltzmann machines (RBM) are also used to study multivariate filtering and responses, such as operations. When the feature vector is changed to ROI or VOI, the feed will be delivered to the reader and trained to discern patterns. Various classification methods have been proposed, including quantitative analysis (Schchidan et al., 2017) and Bayesian classifier (Latif, 2019), which are the closest k-nearest neighbors associated with a neural network. Artificial (Roy et al., 2012),

4. CNN ARCHITECTURE

To reduce data complicated or concentration only on different phases of removal research, the method uses the same CNN system proposed in each step. The network can be accessed through a 32×32 pixel image wall (Li et al., 2014). The patch is connected to five consecutive intricate layers with 2×2 cores, the number of cores corresponding to the area of each card relative to the insert.

The number of cores we use for the L^{th} board is $k(L+1)^2$, which depends on the data complexity and is selected as 4. Finally, the release of the intricate layer is always mixed together. The world is therefore mapped to a series of three thin layers according to the weight of each card.

The thin-walled group (ReLU) serves as the electrical conductivity of the thin layer, while the convolution plate acts on the very thin ReLU, $a = 0.3$. Finally, use Dropout before losing 50% of the cable. To train the network, the optimized values and their hyper parameters are used by the Adam optimizer. Unless another 200 cycles pass at least 0.5% verification to improve network transmission, the training will end.

Figure 1: Lung tissue patterns in Interstitial Lung Diseases

The CNN architecture is shown in Figure 1. The format may be brief as $28 \times 28 \times 1-26 \times 26 \times 4-100M$, where M is sum of classes. The input is a grayscale picture area. The image block dimension is 28×28 pixels. Our CNN modifier covers only convolution layer consisting of 4 cores. The size of every core is 3×3 pixels. Other CNN architectures, our architecture do not use a grouping layer. Then a totally attached layer of 100 neurons immediately pursues the convolution layer. The last layer consists of the logmax and softmax regression, the regression generates the probability of each category, making

$$P(y = i | x, W1, \dots, WM, b1, \dots, bM) \quad (1)$$

$$\frac{e^{w_i x + b_i}}{\sum_{j=1}^m e^{w_j x + b_j}}$$

Where x is production of fully linked layer, W_i and b_i are weight or deviation of the i-th neuron in this layer, and M is the number of categories. The forecast category $\hat{y}$ is the category with highest probability, so

$$\hat{y} = \arg \max_i P(y = i | x, W1, \dots, WM, b1, \dots, bM).$$

In CNN convolutional layer or in the fully linked layer, corrected linear unit (ReLU) is used as a neuron. ReLU is specified as: $f(x) = \max(0, x)$, (3) where x is participation of neuron. Showed that using ReLU as neurons in CNN is superior to traditional S-type neurons (Latif et al., 2019).

Training is training CNN. For each super pixel, we produce a patch centered on super pixel. The patch is considered an entry to the system. The dimension of all patches is 28×28 pixels. The make for each color patch is the make for its center pixel. The training image patch is exploit to sequence system. On CNN, the cost function described as loss of cross entropy, which makes

$$L(X, Y) = -\frac{1}{n} \sum_{i=1}^n (\ln a(x(i)) + (1 - y(i)) \ln(1 - a(x(i)))) \quad (2)$$

Where $X = \{x(1), \dots, x(n)\}$ is the training picture block, or $Y = \{y(1), \dots, y(n)\}$ is the equivalent set of labels. The sum of teaching image patches is n. For each x(i), a(x(i)) is the CNN output, as defined in the equation. 1. CNN is trained by stochastic gradient descent (SGD) through the abandonment technique. The purpose of dropout is to avoid over suitable by hosting random noise into the training examples. Therefore, through training, the output of the neuron is masked with a probability of 0.5.

5. PROPOSED SYSTEM

The scanner is performed through different CT scanners through a slightly dissimilar pocket, hence the pre-processing requirement. The weight of the image is broken and the image is shared. Mesh, reinforcement, micro-nodules, the six most relevant ILD modules, is GGO bees and GGO blends and bolts. It also helps to organize health, which results in lesson

A-Pre-Processing: By adjusting the intensity of the image to highlight the target area, that is, the visual area that is diseased after the data collection is complete, the image value can be developed. The histogram equalization algorithm is used to improve image.

B-Feature Extraction: Function plays a vital role in distinguishing one virus from extra. We intend to use color characteristics such as the mean or standard deviation. We also idea to use surface features such as kurtosis, skewness, cluster highlights, and cluster shadows. To extract surface features, a grayscale match matrix will be used.

C. Classification: The classification assigns documents to precise

collections or classes. Classification is generally called managed education method. Classification is a two-step course: First, a classifier model is produced, which describes a predefined set of courses. This step is named learning phase where classification algorithm matures a classifier by "learning" records with specific class labels. In the 2nd step, use model produced in the first step for classification

D. Testing & validation: To evaluate and verify different types of convolution neural network models, the classification of rice disease will be evaluated. The test data is used to assess correctness of the training model by evaluating the model's performance on the test data. In the rice virus organization, disease is secret conferring to the characteristics removed from the image.

E-7 Classes:

- healthy,
- ground glass opacity
- GGO
- micro nodules,
- consolidation,
- reticulation,
- honeycombing
- combination of GGO/reticulation

Figure 2: Proposed Modules

6. RESULT ANALYSIS

Evaluation: The evaluation of diverse ILD patch classification methods is founded on the train verification test plan. The real teaching of method is done on teaching set, while support set is used to fine-tune the operation: classifying the pattern of lung tissue. After connecting "Browse Image", we can look or take copy as input. In preprocessing, we are adjusting the size of the input image. After resizing, we will improve the idea image. In the subdivision, we will apply threshold, reversal and covering. Later all these 3 steps, we obtain the segmented images of the lung area and the ROI (that is, the region of interest).

We use LBP (local binary mode, GLCM, statistical characteristics, wavelet characteristics) to extract the texture of the tissue. After clicking "estimate", you will get a difference analysis of lung tissue pattern

Figure 3: Input Image

Figure 4: cropped image

Figure 5: binary image

A- Image Segmentation

Segmentation divides the image into different regions, and each region contains individual pixels that have the same properties. In order to be relevant or valuable for image analysis or interpretation, these areas should be closely related to the matters or aspects of concentration presented. Meaningful partitioning is first step in converting low-level image processing from slides or colored images to one or more images for high-definition visualization in conditions of element and purpose the success of picture examination depends the confidence of the partition, but the separation of the true image is often difficult. The resolution of image segmentation is to combine pixels into areas of blurry images, that is, areas that correspond to individual aspects of the object or natural components of the object.

Figure 6: segmented image

Figure 7: CNN image

B- Local Binary Pattern (LBP)

Use local binary patterns to extract texture features. Local Binary Pattern

(LBP) is a method used to describe surface structure functions. By using LBP, the probability of texture pattern can be summarized as a histogram. (LBP) is a simple but effective training toolbar. Set a threshold for each pixel to select a screen image, and specify the result as a number. Because of its distinguish ability and computability, LBP script operators have become popular features in various applications. It can be considered as a visual analysis method combining traditional statistical models and structures. For example, the most important characteristics of LBP operators in practical applications may be their confidence in changes in currency gradients caused by changes in lighting. Another important feature is its simplicity, it can control images in real time

C- Statistical Feature

Technique analysis techniques determine the area of the image in terms of higher order time in its history of gray matter. The technique of strategic image analysis defines a column as the creation of elements that create a specific shape (such as the contour line). The properties and principles of the elements describing the elements determine the shape of the image. Model-based analysis technology generates a pixel-level model of the image based on the weighted average of the intensity of the adjacent pixels. The graphical representation of the image model is used as a text reader. The finite-element analysis technique uses the frequency domain of the north pixel to convert pixels into new shapes. The success of the genre rests on the kind of variation used to produce visual effects. Here, the analysis of micro channel images is analyzed using statistical analysis and transformation.

D-Gray-Level Co-Occurrence Matrix (GLCM)

Gray-Co Matrix (GLCM) image analysis of GLCM functions increases the size of an image by determines the occurrence of these figures with values. Specific or spatial relationships defined in the graph, create GLCM, and then retrieve Statistics in this matrix To determine the best form and remedy for CNN, CNN recommends that people first determine the wildlife of problematic (in this case) by looking at the organisation of ILD models. . Unlike high resolution solutions with specific instructions in colour images, the ILD image of the CT image is normalized. Although creativity is easily understood by humans, it is not always easy to formulate definitions, which is why the definitions found in the literature. Here, some (Greek) systems have been described as key to the visual reproduction of small systems compared to whole regions. Image development has a tendency to look like little corn in the whole picture area. In this way, the analysis of bank credit results has been successful in many research methods

Figure 8: Healthy class graph

Figure 9: micronodules ILD

Figure 10: reticulation ILD

Figure 11: reticulation / ground glass opacity class graph

The comparison results show that the method with the maximum AUC is implemented in the 7 categories. To test AUC difference statistics, the analysis is performed by class (single-to-many) and the size of all lessons is too given. The most difficult are mixtures, gingerbread, honey and hex / GMO. The difference in VGGNet is invalid. Considering the category are the benefits of the proposed method also considered important.

Figure 12: Ground Glass Opacity ILD

Figure 13: Consolaidation ILD

Figure 14: Honey Combing ILD

Figure 15: Accuracy graph

Map 20 Knowledge Map In deep learning, time is a hyperparameter defined before the model is trained. At one point, all of the signals are re-routed and retracted simultaneously through the neural network. For example: if the 2000 training data set is divided into 500 plates, the 4 series will complete 1 period.

Figure 16: Epochs graph

Figure 17: accuracy for different techniques

Table 1: The proposed method for the seven considered classes

Classes	AUCproposed (%)	AUCalexnet (%)	AUCalexnet (%)	AUCvgg (%)	AUCsorensen (%)
Consolidation	91.2	82.29	78.76	77.56	71.86
Healthy	97.1	90.51	86.08	81.75	80.00
Reticulation	97.18	92.4	87.03	86.44	79.86
Micro nodules	100	96.17	95.87	91.11	88.17
GGO	94.2	91.88	85.90	81.11	72.12
RGGO	97.18	92.44	84.09	79.39	76.58
Honey Combing	100	93.47	88.16	81.01	75.96
Average Accuracy (Proposed)	91 %				
Previous Accuracy Namrata Bondfale [1]	81%				

7. CONCLUSIONS

In this study, we developed a CNN framework to classify lung tissue patterns into normal, mesh, and honeycomb. A CNN network architecture is designed to capture the texture of lung tissue. The frame contains 5 convolution layers, followed by two totally associated layers. The developed system has a good precision of results in the HRCT scan data set, reaching 91. In this document, we propose a training method that can improve the precision and stability of CNN in the task of classifying patterns of lung tissue. CT images of the lungs are divided into 7 categories, including 6 different ILD patterns of healthy sand. This method can be easily trained in other textured lung modes, while further improving performance. For the same entry, appropriate to random initialization of the weights, the results fluctuate slightly. As result the proposed method showing higher accuracy as compare to existing technique

REFERENCES

- Al-Jumaily H.M.A. 2015. Scott Claxton Deep Learning of Facial Depth Maps for Obstructive Sleep Apnea Prediction Published in: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI) DOI: 10.1109/iCMLDE.2018.00036 Sydney, Australia,
- Aslam, H.A., Ramashri, T., Ahsan, M.I.A. 2013. A new approach to image segmentation for brain tumor detection using pillar K-means algorithm, *Int. J. Adv. Res. Comput. Commun. Eng.*, 2(3), 1429–1436.
- Bondfale N.D.S. 2018. Bhagwat Convolution Neural Network for categorization of Lung Tissue Patterns in Interstitial Lung Diseases Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018) IEEE Explore Compliant - Part Number: CFP18BAC-ART; ISBN:978-1-5386-1974-2
- Gao M. et al., 2015. Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolution neural networks, in *1st Workshop Deep Learn. Med. Image Anal.*, 41–48.
- Heitmann K. R. et al., 1997. Automatic detection of ground glass opacities on lung HRCT using multiple neural networks, *Eur. Radiol.*, 7(9), 1463–1472.
- Khalid, N.E.A., Ibrahim, S., Haniff, P.N.M.M. 2011. MRI brain abnormalities segmentation using K-nearest neighbors (k-NN) ', *Int. J. Comput. Sci. Eng. (IJCSE)*, 3(2), 980–990.
- Krizhevsky, I. Sutskever, and G. Hinton, 2012. ImageNet classification with deep convolution neural networks, *Adv. Neural Inf. Process. Syst.*, 9.
- Latif J. 2019. Chuangbai Xia Medical Imaging using Machine Learning and Deep Learning Algorithms: A Review 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) 10.1109/ICOMET.2019.8673502 Sukkur, Pakistan,
- Li Q. et al., 2014. Medical image classification with convolution neural network, in *Proc. 13th Int. Conf. Control Automat. Robot. Vis.*, 2014, 844–848.
- Li Q., Cai W., and Feng D. D., 2013. Lung image patch classification with automatic feature learning, in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 6079–6082.
- Maity, A., Pruitt, A.A., Judy, K.D.: 2008. Cancer of the central nervous system (Clinical Oncology, 2008, 4th edn.)
- Qurat-Ul-Ain, L.G., Kazmi, S.B., Jaffar, M.A., Mirza, A.M. 2015. Classification and segmentation of brain tumor using texture analysis, *Recent Adv. Artif. Intell. Knowl. Eng. Data Bases*, 10, 147–155
- Ricci, P.E., Dungan, D.H. 2001, Imaging of low and intermediate-grade gliomas, *Semradonc*, 11(2), 103–107
- Roy S, Bandyopadhyay S.K, 2012. Detection and Quantification of Brain Tumor from MRI of Brain and it's Symmetric Analysis, *IJICTR*, 2(6).
- Sachchidan and Singh. 2017. Object classification to analyze medical imaging data using deep learning Published in: 2017 International Conference on Innovations in Information, Embedded and Communication Systems

([ICIIECS](#))DOI: 10.1109/ICIIECS.2017.8276099
Location: Coimbatore, India

Conferece

Tulder G.V and Bruijne M.D, 2014. Learning features for tissue classification with the classification restricted Boltzmann machine, Med. Comput. Vis., Algorithms for Big Data, 47–58.

Selvakumar J., Lakshmi A., Arivoli T., 2017. Brain Tumor Segmentation and Its Area Calculation in Brain MR images using K-Mean Clustering and Fuzzy C-Mean Algorithm", IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012) March 30, 31.

Uppaluri R. 1999. Computer recognition of regional lung disease patterns, Am. J. Respir. Crit. Care Med., 160(2), 648–654.

Simonyan K. and Zisserman A., 2015. Very deep convolution networks for large-scale image recognition, in Int. Conf. Learn. Represent.

Wolf Y.B. 2018. Sivan Lieberman Date of Conference: 16-19 Chest pathology detection using deep learning with non-medical training DOI: [10.1109/ICIIECS.2018.8276011](#)

Singh K.K, Singh A, 2018. A Study Of Image Segmentation Algorithms For Different Types Of Images", IJCSI International Journal of Computer Science. 7(5).

Xu R. et al., 2011. Classification of diffuse lung disease patterns on high resolution computed tomography by a bag of words approach, Proc. MICCAI, 14(3), 183–90.