Using Machine Learning via Deep Learning Algorithms to Diagnose the Lung Disease Based on Chest Imaging: A Survey

Shaymaa Taha Ahmed[1(✉)], Suhad Malallah Kadhem[2]
[1]University of Technology, Baghdad, Iraq
[2]University of Diyala, Diyala, Iraq
cs.19.25@grad.uotechnology.edu.iq, mrs.sh.ta.ah@gmail.com

Abstract—Chest imaging diagnostics is crucial in the medical area due to many serious lung diseases like cancers and nodules and particularly with the current pandemic of Covid-19. Machine learning approaches yield prominent results toward the task of diagnosis. Recently, deep learning methods are utilized and recommended by many studies in this domain. The research aims to critically examine the newest lung disease detection procedures using deep learning algorithms that use X-ray and CT scan datasets. Here, the most recent studies in this area (2015–2021) have been reviewed and summarized to provide an overview of the most appropriate methods that should be used or developed in future works, what limitations should be considered, and at what level these techniques help physicians in identifying the disease with better accuracy. The lack of various standard datasets, the huge training set, the high dimensionality of data, and the independence of features have been the main limitations based on the literature. However, different architectures of deep learning are used by many researchers but, Convolutional Neural Networks (CNN) are still state-of-art techniques in dealing with image datasets.

Keywords—lung disease, machine learning, deep learning, CT-images, CNN, Covid-19

1 Introduction

Human cells, which heavily rely on a constant and sufficient oxygen supply to operate efficiently, will suffer if oxygen inflow is interrupted. The lungs are where the body takes in oxygen and releases extra carbon dioxide, which can be harmful to the body [1], [2]. However, since the lungs are fragile organs that are continuously vulnerable to the outer environment, they are exposed to many diseases known as lung diseases. Any of these diseases include lung cancer, chronic obstructive pulmonary disease, Covid-19, asthma, chronic bronchitis, influenza, lung fibrosis, sarcoidosis, and tuberculosis [3].

The most concentrated lung disease these days is Covid-19, because of the pandemic. To date, at least 106 million cases and at least 2 million reported fatalities have been reported by the COVID-19 pandemic. Early diagnosis of infectious patients and accurate surveillance of critically ill patients are crucial for pandemic management and mortality mitigation due to the high transmittable rate, long incubation time, and comparatively high mortality rate [4].

Computer-aided detection/diagnosis (CAD) in radiography and medical data imaging has advanced significantly in recent years, owing in large part to the introduction of deep learning (DL) strategies. In computer vision tasks, various deep learning models, mostly convolutional neural networks (CNNs), have had considerable success. Breast cancer, lung cancer, glaucoma, and skin cancer are among the diseases for which early identification and diagnosis are available [5]. X-rays, computerized tomography scans (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) are the traditional technologies used to diagnose and identify lung cancer [6]. The analysis of cardio arrhythmia has also benefited from the use of algorithms such as the Discrete Wavelet Transform [7], [8]. Artificial intelligence strategies have been used in several research trials to classify lung diseases. Multilayer, stochastic, vector quantization, feedforward neural network, and generalized regression NN have been used to diagnose lung disorders such as nodules, pneumonia, and lung cancer using chest radiographs. However, they did not do as well as deep networks in terms of accuracy, calculation time, and achieved minimum square error [9].

Machine learning is a fascinating area of computer science and engineering study. Since it allows the extraction of meaningful patterns from examples, which is a part of human intelligence, it is considered a branch of artificial intelligence. The allure of having a machine that performs well-defined, repetitive tasks is obvious: computers can consistently perform a given mission [10]. Machines have recently demonstrated the potential to learn and even master tasks previously thought to be beyond their capabilities, demonstrating that machines can learn and master tasks previously thought to be beyond their capabilities. Learning algorithms may be useful in computer-assisted diagnosis and decision-making systems [11]. Perhaps more fascinating is the discovery that machines can, in certain situations, "sense" patterns that are invisible to humans. As a result of this observation, machine learning has sparked a lot of interest, particularly in terms of how it can be applied to medical images [12].

Deep learning-based algorithms have been used to increase image recognition performance in a variety of fields and have significant and positive results. Convolutional Neural Network (CNN) is one kind of deep learning and has been used beyond the medical field for biometric identification of palm vein attributes [13]. and was also used to classify texts into common datasets and for text categorization [14], [15]. Neural Networks were also used as a unique solution to the growing demand for facial expression recognition by using feature extraction using the Haar wavelet and Karhunen-Loève Transform (KL-T) [16]. Effective machine learning techniques were also used to build a strong model for stock market prediction [17], solve challenges associated with

identifying null values problems in datasets [18], and recognizing and classifying Arabic, handwritten text using Discrete Wavelet Transform and Support Vector Machine techniques [19]. Algorithms such as the Light Gradient Boosting Machine (LightGBM) have even been applied to insider attacker threats, which is one of the most challenging cybersecurity challenges [20]. Machine learning and Decision Tree Algorithms have even been used to discriminate between skin and non-skin in imagery analysis using an Improved ID3 Algorithm [21]. The diversity of uses of Neural Networks and machine learning techniques and algorithms indicate the appropriateness and robustness of applying them to medical challenges.

In the medical sciences, researchers were motivated by the success of deep learning networks to apply them to medical data images for tasks like lung disease identification, and the results revealed that deep networks can successfully extract useful features that distinguish various image classes [22]. The most commonly used deep learning framework is the convolutional neural network (CNN). CNN has been used to classify various medical photographs due to its ability to derive various type characteristics from pictures [23].

In this survey, we focus on the most recent medical imaging-based (CT and X-ray) deep learning works which have been conducted for lung disease classification. Due to Covid-19 and other traditional lung illnesses, it became so crucial to understands which modern techniques are best for developing such a system that can decrease the huge load on the medical staff and enhance conventional human diagnostics. This research aims to firstly, recap and categorizes several types of deep learning works in medical fields that are employed to enhance the performance of lung disease diagnostics systems in various applications. Secondly, describe the problems and directions for designing deep learning models. The remaining sections of the research are formed as follows: Deep Learning methodology describes in section 2; Lung Disease Classification System Developments is section 3 is presents, the detailed Discussion is illustrated in section 4, and Lastly, the research is concluded in section 5.

2 Deep learning methodology

Deep learning is a branch of algorithm-related machine learning influenced by the brain's work and design. The recognition, measurement, and classification of models in medical data images are supported by new advancements in machine learning, specifically deep learning. Due to the capacity of deep learning to learned features, these innovations were made possible solely from data, rather than hand-made features based on domain-related information [24]. Deep neural networks are used to solve and handle complicated problems such as image processing, natural language processing, and large-scale machine vision. It can also use Deep Learning to identify image features dynamically [25]. In many medical applications, deep learning is increasingly becoming state-of-the-art, headway to enhanced results [26]. Figure 1 described the entire process of lung disease detection using DL methodology.
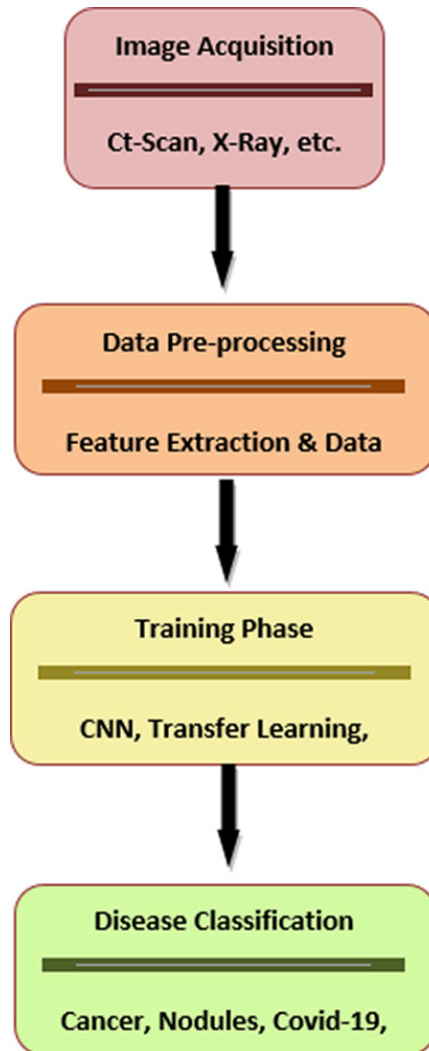
Image Acquisition

Ct-Scan, X-Ray, etc.

Data Pre-processing

Feature Extraction & Data

Training Phase

CNN, Transfer Learning,

Disease Classification

Cancer, Nodules, Covid-19,

Fig. 1. Lung disease detection using DL methodology block diagram

A CT scan is a type of medical imaging that uses system analysis to produce sub-images from images acquired across the person's body from many angles at various levels of detail. The sub-image can be seen separately or assembled to produce a patient's 3-dimensions image displaying the organs, body tissues, bones, and all anomalies present. Images from CT scans offer more accurate details than X-rays. In various works reported in the survey, for pneumonia, lung cancer, and COVID-19 recognition, CT scans had utilized to diagnose lung disease [4], [27].

Detection of lung disease usually deals with identifying a photo into healthy lungs or lungs contaminated with disease. The classifier for lung disease, also known as a model, is obtained through training. Deep learning can be used to build a classifier that can identify images into their related class labels [28]. As a result, the first step

is to collect photographs of lungs with the illness to be classified to use deep learning to identify lung disease. The second step would be to training the model before the diseases can be recognized by it. The last step is to categorize other images that are unknown to the model [14]. Classifying images is done by identifying objects in those images, techniques like edge detection (which is the first step of image segmentation) aim to recognize spots in an image where the image concentration variate dramatically. To limit the currency, note that is the region of interest, edge detection is needed [29]. In digital image processing, edge detection is one of the most popular techniques. Edges are focused on localized changes in the intensity of an image caused by the borders of object surfaces in a picture [30].

It's really important to have a huge training dataset in deep learning, which can help increase the accuracy of training. Using image augmentation is one way of growing the training dataset without collecting new images. Variations of the original images are generated by image increase. This is accomplished by performing various processing techniques, such as rotations, twists, transformations, zooms, and noise addition [10].

To minimize the number of sample images required from each person, the system utilizes the data augmentation process on the input image [31]. Augmentation of data also helps avoid overfitting by adding more complex data to the model. This knowledge diversity decreases variance and increases the model's generalization. However, increased preparation time, transformation processing costs, and extra memory costs are drawbacks of data augmentation [32].

CNN's are a family of machine learning with several convolutional layers that can be trained to pick features of an image, with the majority of them can be used in computer vision for image processing. During the training phase, Deep CNN can reduce the gap among its current model and classified components. During this method, the features of the different factors are also obtained in the various wrapped layers. When analyzing a new input picture, this approach would allow for the accurate identification and classification of structural components [33] [40].

In particular, the Convolutional neural network (CNN), the exceedingly popular DL method, is beneficial for finding objects in images. CNN's consist of neurons with learnable biases and weights, analogous to the human brain's neural networks. Several inputs are obtained by each neuron. After that, a weighted sum of the inputs is calculated. The cost function is then transferred to a kernel function, which generates an output [34]. In the diagnosis of chest diseases, different types of CNN architectures, including AlexNet, VGGNet, GoogLeNet, ResNet to DenseNet, have achieved tremendous success. Figure 2 illustrates the form of a standard CNN used in the chest X-ray image for disease diagnosis.
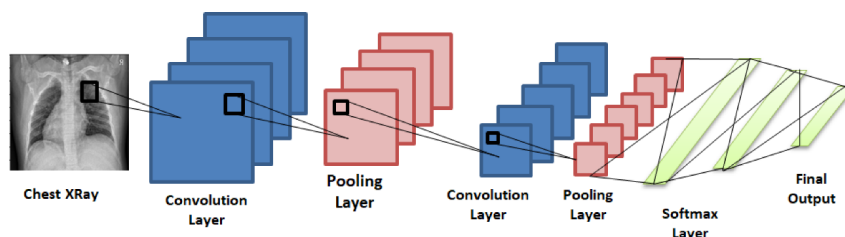

Fig. 2. A traditional CNN design for the diagnosis of medical diseases [34]

3   Literature review

In the literature, many works use deep learning techniques to identify lung diseases. This section will illustrate the most recent works in this field.

In 2018, A. Masood, et al. [35], using the new DL model and MBAN metastasis information, a computer-assisted pulmonary cancer judgment support system was proposed. The proposed methodology, DFCNet, is built on a deep full convolutional neural network (FCNN) that classifies each lung nodule into one of four cancer stages. Experimental results show the effectiveness of the proposed method for identifying and classifying lung nodules. These findings show the ability of the proposed technique to enable radiologists to effectively increase the accuracy of nodule detection. The suggested testing methodology is evaluated using several datasets, including the LIDC/IDRI dataset, RIDER, LungCT Diagnosis, and the Lung Nodule Analysis (LUNA) 2016 Set of data. The SPIE Test Database, which was not transcribed, was also a dataset that has been used primarily for testing to see whether the DFCNet could differentiate between a nodule (benign or malignant) and a non_ nodule without markings. CNN and DFCNet's average accuracies were 77.6% and 84.58%, respectively.

In 2019, S. Suresh, et al. [36], a strategy for automated tumor stage classification of pulmonary chest radiography is introduced using an edge learning deep CNN (DCNN). The photographs used in the study were collected from 1018 cases from the public archive of LIDCIDRI database. The method seeks to examine and obtain the self-learned salient characteristics composed of differently shaped nodules from the NROI. DCNN are learned with NROI observations and are also graded as not-cancerous, benignant, or cancerous models according to tumor patterns. The study incorporated a thorough analysis of the design of the LIDC datasets focused on the levels of nodule disease. On 3 LIDC datasets, the technique was tested using 10-fold cross-validation. The analysis showed a 97.8% Accuracy of classification.

In 2019, Q. Zhang, et al. [13], by conducting an end-to-end identification of 3D-nodule CT pads by utilizing CNNs, nodule categorization and feature extraction are removed in the CAD system. In particular, the CNNs are updated to 3-dimensional CNN models and examined on a common lung CT data scan dataset: The LIDC/IDRI. Experimental studies indicate that for lung nodule prediction, DenseNet121 and Xception produce better outcomes. Multiple most common DL methods are utilized in experiments to identify nodules, which can increase the performance and effectiveness of CAD systems. Primarily, end-to-end classification can be carried out from raw 3D nodule CT pads by automatically identifying discriminative optical features using CNN models. With the DenseNet121 model, the accuracy of the device was 87.77%.

In 2019, A. K. Jaiswal, et al. [37], a DL-based technique is adopted in Chest X-rays (CXRs) images for the detection and characterization of pneumonia. Via crucial improvements to the learning procedure and a novel post-processing phase that incorporates borders on cases from various frameworks, the work obtains robustness. The proposed model of identification provides a better output on the chest radiograph dataset that indicates possible causes of pneumonia. The use of a Mask-RCNN-based modeling approach to classifying lung transparency that is possible to represent

pneumonia. Mask-RCNN is a DL network developed specifically for object detection. The study used the RSNA pneumonia dataset 5, which is a subset of the NIH CXR14 dataset that transcribed 30,000 samples from a total of 112,000 X-rays. In this work, the accuracy obtained was 97%.

In 2020, X. Chen, et al. [38], a new (self-supervised) DL method for the automatic Covid_19 diagnosis is proposed, requiring just a few patterns for training. The algorithm could capture expressive visual features on huge and publicly attainable chest datasets by using contrastive learning, and follow the typical classification network and create a separated dictionary for continuous big-dimensional data inputs. Two publicly accessible COVID-19 CT datasets are used to test the proposed model: COVID-19 CT 3 and a database produced by the Association of Medical and Interventional Radiology (Italy) 4 and formatted by MedSeg. For this suggested model, the accuracy was 86.80.

In 2020, K. Hea, et al. [39], 3-dimensional CT data images, by in conjunction, conducting lung lobe partitioning and multi-task (M-2 UNet) classification, a learning synergistic system for automatic numerous successes of Covid_19 is proposed. The multi-task is designed to measure the seriousness of Covid_19 cases and concurrently splits the lung lobe as well. The M2 UNet comprises an encoder at the patch-stage, a lung lobe partitioning semi-network of partitioning, and a severity evaluation semi-network of classification. The tests were conducted on a genuine Covid_19 CT image database comprised of 666 chest data images collected from 242 Covid-19-confirmed patients. 98.50 was the accuracy achieved in this experiment on the mentioned dataset.

In 2020, X. Wu, et al. [41], A self-supervised active DL algorithm called COVID AL is presented to diagnose Covid 19 using CT scans and participant marks. For the diagnosis of Covid 19, COVID AL blends a 2-Dimensional U-Net lung field categorization with a modern hybrid adaptive learning strategy. With just 30% of the classified results, the COVID AL obtained over 95% accurateness of the DL technique using the entire dataset. The COVID-19 diagnostic dataset was collected from samples of the Chest CT data Investigation Consortium of China (CC/CCII). In terms of accuracy, the suggested COVID_AL obtains 86.06, which is claimed to be the best result among all the active learning strategies.

In 2020, M. Turkoglu. [42], a new Multi Kernels ELM-based DNN (MKs—ELM—DNN) approach is proposed for the diagnostic of Covid_19 using CT scan images. In the proposed method, deep characteristics are derived from CT scan data by utilizing a CNN. The pre-trained DenseNet201 CNN-based system, a transfer learning methodology, is being conducted for this task. An available dataset comprised of Covid_19 and Non-Covid_19 categories was examined in experiments conducted to check the efficacy of the proposed MKs--ELM--DNN model. For the DenseNet201 architecture, the transfer learning approach is used in the present analysis. This strategy is founded on collecting deep features using the pre-trained architectural design fully-connected layer. The accuracy score of 98.36% was achieved using the MKs—ELM—DNN framework, based on the results achieved.

In 2020, Z. Tao, et al. [43], an ensemble DL model is proposed for the new Covid_19 detections from Image data. 2933 CT data photographs from Covid_19 cases were obtained from past reports, reputable media reports, and online databases. To acquire

2500 high-quality images, the images were normalized. The hospital received 2500 CT data of lung cancer and 2500 of non-lung cancer. To configure input variables and pre-train 3 deep CNN models, transfer learning was used: AlexNet, ResNet, and GoogleNet. For extracting features on all images, these models have been used. They indicate that the whole recognition accuracy of the ensemble framework was higher than the part model. Soft-max was used as the identification model for the completely linked layer. For this work, the accuracy rate achieved was 99. 44%.

In 2020, G. Savitha, et al. [44], for extracting features, a sophisticated deep learning method is used that enhances the system's nodule localization or identification process. Additionally, by implementing a dependent random structure in the framework, the false positives slipping out of the process are significantly decreased. By using a set of convolution and de-convolution layers, the DNNs develop deep features and recognize the necessary item. Also, in the current context, a Conditional Random Field (CRF) approach is used to remake the findings of the built DL model more solid and precise. For the study of the CAD system, the LIDC-IDRI dataset is assessed. In the current test, 80 percent of the images (710) are used for preparation, 10 percent (89) for validation, and 10 percent (89) for testing amongst those total 888 images. Due to memory restrictions, the established CNN framework is learned for 100 iterations and the cycle time is determined as 10. The output in terms of the correctness of the suggested algorithm was 89.48%.

In 2020, D. Ezzat, et al. [45], GSA—DenseNet121—COVID—19, a new approach, a hybrid-based CNN model, is proposed as an enhancement algorithm. Dense-Net121 is the CNN model employed, and the generalized search engine is the optimization technique (GSA). The GSA is being used to choose the optimal cost function values for the DenseNet121 model. To support this design in accurately identifying Covid_19 using x-ray scans. The binary COVID-19 dataset was handled and divided into 3 packages as a result of the mismatch: training set, testing set, and Evaluation Package. Using various data augmentation methods, after increasing the amount of training data set in the first level, With the validation package, they were used in the second level. The results obtained showed that 98.38 percent of the test data could be identified accurately by the proposed method.

In 2020, S. Bharati, et al. [46], integrate VGG, data augmentation, and the spatial-transformer network (STN) with CNN's models to generate a good hybrid DL system. VGG-Data-STN with CNNs is the name of this modern technique is VDS-Net. The experiment employs the new VDS-Net technique to evaluate lung disease datasets and estimate lung illnesses. The dataset entry attribute (such as maturity level, X-ray photos, gender, and view status) is used to perform binary classification for this purpose, with the result being the identification of 'Yes' or 'No' identified diseases. In the case of the entire dataset, VDSNet has a validation accuracy of 73%, while regular gray, regular (R/G/B), hybrid CNN's, and VGG, as well as the updated module system, possess accuracy results of 67.8, 69.5, 69.5, and 63.8 percent, respectively. The new model now includes the NIH lung X-ray image dataset from the Kaggle repositories. It includes (112,120) X-ray images of the chest or lungs, as well as disease lists for

30,805 different patients. The chest X-ray image dataset is extremely versatile as well as large in size.

In 2021, M. Canayaz, et al. [47], a set of data was created that included three Covid_19 groups, healthy and pneumonia lung-X-ray photos, each with 364 photos. Extracted features from this data set are completed using DL approaches such as Alex-Net, VGG-19, Google-Net, and Res-Net. The best features were chosen using two meta-heuristic methods: binary particle and binary gray wolf. After integrating the features obtained in the feature selection of the enhancement data set, they were classified using SVM. In the experiment, two datasets were used, with the first dataset being the standard X-ray (by Joseph Paul Cohen). The public dataset provided by Kaggle was the secondary dataset we used for Covid_19. In the process of producing the enhancement data set, the opacity enhancement was performed separately per each image in the source data using the image contrast enhancement algorithm (ICEA). Both sets of knowledge. 70% of the data was used as training data in research tests, while 30% was used as evaluation data. The BPSO algorithm was used to pick and identify the function derived from the VGG19 model, yielding an overall accuracy of 99.38 percent.

In 2020, A. R. Akkar, et al. [48], proposed a method that is working based on Artificial Intelligence, which in the back end supported by ANN to predict skin carcinoma images, a moth flame optimization is utilized for the training of the artificial neural network. The comparative result on a variety of metrics such as average detection rate, average mean squared error, average effective processing time, and the average effective iteration number show that the proposed method MFO yields 85% ARD with minimum time cost compared to the different features.

In 2020, HAR Akkar, et al. [49], collected real data from Baghdad Educational Hospital, the medical city, the Radiology department. Some pre-processing operations were conducted on the data, to boost the images' quality for better algorithm performance. The authors applied classification techniques using backpropagation artificial neural network followed by three activation functions such as trainlm, trainbr, and traingd. The performance was the show with the trainlm activation function which is said to be 95.9% accuracy.

In 2021, Xiang Yu, et al. [50], proposed a framework using the DL method called CGNet to make binary classification on X-ray data images of the chest into pure and pneumonia. The framework contains three main parts, extraction of features, reconstruction of features using graph method, and classifier algorithms. The authors applied CNN first to extract features with higher importance for the two following steps. Then graph-based approach was used to reconstruct features and lastly, the identification was made using a shallow NN called GNet. The accuracy yielded by this model on the public pneumonia dataset was 98.72 % and the result of the test on the COVID-19 CT dataset is 99%.

In 2021, AK Das, et al. [51], proposed a framework for using chest X-ray images to scan Covid_19 cases The framework aims to categorize photos into three groups: covid_19 positive cases, all other pneumonia-infected cases, and mild or

uninfected cases. VGG 16, CNN, and Res-Net-50 are the algorithms used in the model. The framework was validated on Kaggle's normal Covid_19 radiography dataset, which consists of chest X-ray images. The result of VGG-16 was better compared to the two others, as it yields 97.67 % accuracy.

In 2021, E Hussain, et al. [52], proposed a model called CoroDet using CNN to detect Covid-19 cases automatically by using raw chest images of patients which could be X-ray or CT Scan. The model is aimed to classify images into two classes which are covid-19 positive or normal, three classes which are Covid-19 positive, normal cases and non covid_19 pneumonia, and four pure classes, Covid_19 cases, non Covid_19 bacterial pneumonia cases, and non Covid_19 viral pneumonia. The results of their approach are 95.1% accuracy for 2 classes, 94.2% for 3 classes, and 91.2% for 4 classes. The dataset was prepared by themselves.

In 2019, Israa S. Abed [53], proposed an automated system using a combination of PCA and backpropagation neural network with the capability of improving the detection process of carcinoma disease. The model works based on X-ray images, MRI, CT Scan images, etc. PCA was used for feature selection and BPNN was employed for classification purposes. The higher result obtained was 96.67% accuracy by the proposed method.

In 2016, Muayad Sadik Croock, et al. [54], proposed an edge detection model using MATLAB tool with the algorithms of CLAHE, Canny, Otsu's, and 8-Connectivity to extract features from bitewing X-ray images, perform image segmentation, and conduct image classification. The performance of the model was claimed to be satisfactory in terms of feature extraction and edge detection. The data of X-ray is used for the evaluation of the method and the result was said to be 80%.

In 2015, Matheel Emaduldeen Abdulmunim, et al. [55], proposed a method using the ID3 algorithm for breast cancer detection using wavelet and contourlet to support physicians to read mammogram images and help them to detect the issue at an earlier time. The method is going to classify images as normal and abnormal or benign. Based on the studies conducted, the result of this experiment shows that ID3 yields 81% and 95% accuracy in the classification process using wavelet and contourlet respectively.

The survey that we present in Table 1, is the most recent advancement of the Lung Disease diagnostic system (2018–2021), it gives us a brief overview of what has been done in this important medical area by focusing on the deep learning methodology. The titles of this Table's columns are organized as follows:

1. Author. This refers to the researcher who developed the techniques.
2. Year. This refers to when the research is done.
3. Country. This refers to the nationality of the authors.
4. Techniques. This refers to the DL modeling methods that are utilized in the research.
5. Image Type. This indicates the type of image data in the dataset.
6. Dataset. This refers to the dataset, whether it is a standard dataset or private.
7. Accuracy. It indicates the accuracy of the system.

Table 1. The most recent works of lung disease detection using
deep learning approaches

Author	Year	Country	Techniques	Image Type	Dataset	Accuracy
A. Masood, et al. [35]	2018	China, Pakistan, Macau, Hong Kong, and Australia	FCNN	CT Scan	LIDC-IDRI, RIDER, LungCT-Diagnosis, LUNA, and SPIE	84.58%
S. Suresh, et al. [36]	2019	India	DCNN	CT Scan	LIDC-IDRI	97.8%
Q. Zhang, et al. [13]	2019	USA	CNN	CT Scan	LIDC-IDRI	87.77%
A. K. Jaiswal, et al. [37]	2019	UK, Italy, Russia, India, Brazil, and Portugal	Mask-RCNN	X-ray	RSNA	97%
X. Chen, et al. [38]	2020	Australia, UAE, and the USA	Self-Supervised Contrastive Learning	CT Scan	Covid-19 CT 3 and Covid-19-CT Segmentation (MedSeg)	86.80%
M. Turkoglu [42]	2020	Turkey	MKs-ELM-DNN and CNN	CT Scan	Private (746)	98.36%
Z. Tao, et al. [43]	2020	China	CNN (Ensemble Deep Learning)	CT Scan	Collection of earlier publications, authoritative media accounts, and public databases	99.44%
G. Savitha, et al. [44]	2020	India	CNN	CT Scan	LIDC/IDRI	89.48%
D. Ezzat, et al. [45]	2020	Egypt	CNN and GSA	X-ray and CT Scan	Covid-19 X-ray (Dr. Joseph Paul) And The Kaggle X-ray	98.38%
S. Bharati, et al. [46]	2020	Bangladesh	CNN+ VGG+ STN (VDSNet)	X-ray	Kaggle repository	73%
X. Wu, et al. [41]	2020	China	COVID-AL – Active Deep Learning technique	CT Scan	CC-CCII	86.06%

(Continued)

Table 1. The most recent works of lung disease detection using
deep learning approaches *(Continued)*

Author	Year	Country	Techniques	Image Type	Dataset	Accuracy
M. Canayaz, et al. [47]	2021	Turkey	CNN, Particle Swarm, and SVM	X-ray	Covid-19 X-ray (Dr. Joseph Paul) And The Kaggle X-ray	99.38%
A. R. Akkar, et al. [48]	2020	Iraq	AI, ANN	CT Scan	(ASCS) 2017	85%
HAR Akkar, et al. [49]	2020	Iraq	BP-ANN	CT Scan	Real data, Baghdad Educational Hospital/ medical city/ Radiology department	95.9%
Xiang Yu, et al. [50]	2021	UK	CNN, Shallow NN	X-ray	public pneumonia dataset/ COVID-19 Chest X-ray	98.72%
AK Das, et al. [51]	2021	India	VGG-16, CNN, and ResNet-50	X-ray	Covid-19 standard dataset from kaggle	97.67%
E Hussain, et al. [52]	2021	Bangladesh	CNN	CT Scan and X-ray	Self-collected dataset	95.1%
Israa S. Abed [53]	2019	Iraq	PCA and BPNN	X-ray, CT scan, MRI	Hospital data	96.67%
Muayad Sadik Croock, et al. [54]	2016	Iraq	CLAHE, Canny, Otsu's, and 8-Connectivity	X-ray	Self-generated dataset from clinical data	80%
Matheel Emaduldeen Abdulmunim, et al. [55]	2015	Iraq	ID3 algorithm	X-ray and CT	Hospital data	81% and 95%

4   Discussion

More work on the identification of lung diseases using DL has been published as time goes on. It is so crucial that comprehensive surveys on the present situation of research and implementation are needed. This paper is therefore made to provide an exhaustive overview of the latest works that utilize DL in the field of lung disease detection.

Particularly published on Pneumonia, Lung Cancer, Nodules, and COVID-19, from the end of 2018 to the present date. Based on the survey of the works considered, a taxonomy of cutting-edge deep learning-assisted lung disease diagnosis was collected and present an overview of the core ideas and emphasis of the current work on lung disease identification using DL. Since CNN was the state-of-art in most works, it's good to mention that for an ideal model that training should consist of at least around 800 images per class unless it would affect the method evaluations. Therefore, the works with a limited amount of data would not be considered as a big contribution to this field. The usage of CNN from the evaluations of the distribution of works is substantial. There are three main drawbacks of lung disease identification using DL methodology were realized. The first is the imbalance of data, which can be solved in future works by incorporating more features or adding knowledge-based features in the model. The second issue was dealing with a huge amount of data, for this problem the usage of cloud computing would be better than locally training a huge amount of data which required more effort technically and manually. The last issue was limited available datasets; this is the biggest problem currently in this area. There are a few standard datasets of CT Scan images that can be authentic for scientific researches.

5   Conclusions

It is very important to build diagnostic systems for lung diseases that can help medical personnel make better decisions easier. In the future, these systems could be expanded to remote areas where it is difficult to receive immediate medical assistance. As professionals, physicians, radiologists, and patients should be introduced so that patients can at any stage take professional advice from them. Also, improvements in algorithms to produce activation maps can be made to the approach for localization of lung illnesses in Chest CT and X-ray data images. By training on better datasets, the recognition rate of the DL model can also be substantially enhanced.

In conclusion, it is extremely important to investigate how deep learning has been used in the detection of lung diseases to make sure that future works focus on the correct path, that way enhancing the productivity of disease recognition systems. Other researchers could use the described taxonomy for planning their work participation and tasks. The possible future orientation proposed could better enhance effectiveness and boost the volume of applications for the identification of lung disease aided by DL. From this survey, it can be concluded that until the current date, the CNN is the most used algorithm in this area. Besides, the Lung Image Database Consortium image collection (LIDC-IDRI) database is the most conducted standard dataset for chest imaging data.

6   References

[1] O. D. M.; H. S. A. (2021). ICMAICT 2020. An Efficient Prediction Model Based on Machine Learning Techniques for Prediction of the Stock Market, 1–11. https://doi.org/10.1088/1742-6596/1804/1/012008

[2] Abbas, A. R., & Farooq, A. O. (2019). Skin detection using improved ID3 algorithm. Iraqi Journal of Science, 60(2), 402–410. https://doi.org/10.24996/ijs.2019.60.2.20

[3] Abdulhussein, A. A., & Raheem, F. A. (2020). Hand Gesture Recognition of Static Letters American Sign Language (ASL) Using Deep Learning. Engineering and Technology Journal, 38(6A), 926–937. https://doi.org/10.30684/etj.v38i6a.533

[4] Abdulwahab, A., Attya, H., & Ali, Y. H. (2020). Documents Classification Based On Deep Learning. INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME, 9(02), 62–66. http://www.ijstr.org/final-print/feb2020/Documents-Classification-Based-On-Deep-Learning.pdf

[5] Akkar, H. A. R., & Hadad, S. Q. (2020). Diagnosis of Lung Cancer Disease Based on Back-Propagation Artificial Neural Network Algorithm. Engineering and Technology Journal, 38(3B), 184–196. https://doi.org/10.30684/etj.v38i3b.1666

[6] Akkar, H. A. R., & Salman, S. A. (2020). Detection of Biomedical Images by Using Bio-inspired Artificial Intelligent. Engineering and Technology Journal, 38(2A), 255–264. https://doi.org/10.30684/etj.v38i2a.319

[7] Al-dabag, M. L., ALRikabi, H. T. S., & Al-Nima, R. R. O. (2021). Anticipating Atrial Fibrillation Signal Using Efficient Algorithm. International Journal of Online and Biomedical Engineering, 17(2), 106–120. https://doi.org/10.3991/ijoe.v17i02.19183

[8] Al-rawi, S. S., Sadiq, A. T., & Alaluosi, W. M. (2016). Feature Extraction of Human Facail Expressions Using Haar Wavelet and Neural network. Iraqi Journal of Science, 57(2C), 1558–1565. https://iasj.net/iasj/download/41faa815da56c9e8

[9] Alsultani, H. S. M., Ahmed, S. T., Khadhim, B. J., & Kadhim, Q. K. K. (2018). The use of spatial relationships and object identification in image understanding. International Journal of Civil Engineering and Technology, 9(5), 487–496. https://iaeme.com/Home/article_id/IJCIET_09_05_053

[10] Bharati, S., Podder, P., & Mondal, M. R. H. (2020). Hybrid deep learning for detecting lung diseases from X-ray images. Informatics in Medicine Unlocked, 20, 100391. https://doi.org/10.1016/j.imu.2020.100391

[11] Bhattacharya, S., Reddy Maddikunta, P. K., Pham, Q. V., Gadekallu, T. R., Krishnan S, S. R., Chowdhary, C. L., Alazab, M., & Jalil Piran, M. (2021). Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey. Sustainable Cities and Society, 65(November 2020), 102589. https://doi.org/10.1016/j.scs.2020.102589

[12] Canayaz, M. (2020). MH-COVIDNet: Diagnosis of COVID-19 using deep neural networks and meta-heuristic-based feature selection on X-ray images. Biomedical Signal Processing and Control, 64(January), 1–12. https://doi.org/10.1016/j.bspc.2020.102257

[13] Chen, X., Yao, L., Zhou, T., Dong, J., & Zhang, Y. (2021). Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images. Pattern Recognition, 113(1), 1–22. https://doi.org/10.1016/j.patcog.2021.107826

[14] Dr. Ahmed, T., Sadiq, S. M. (2013). Hybrid Intelligent Techniques for Text Categorization. International Journal of Advanced Computer Science and Information Technology (IJACSIT), 2(2), 59–76. https://www.researchgate.net/publication/261429194_Hybrid_Intelligent_Technique_for_Text_Categorization

[15] Ezzat, D., Ella, A., & Aboul, H. (2020). Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information. Applied Soft Computing Journal Journal, 98(January), 1–14. https://doi.org/10.1016/j.asoc.2020.106742

[16] Savitha, G., Jidesh, P. (2019). A holistic deep learning approach for identification and classification of sub-solid lung nodules in computed tomographic scans. Computers & Electrical Engineering, 84(June), 1–23. https://doi.org/10.1016/j.compeleceng.2020.106626

[17] Abdullah, H.N., H. K. A. (2019). Deep CNN Based Skin Lesion Image Denoising and Segmentation using Active Contour Method. Engineering and Technology Journal, 37(11), 464–469. https://doi.org/10.30684/etj.37.11A.3

[18] Hamid, M. S., Manap, N. F. A., Hamzah, R. A., & Kadmin, A. F. (2020). Stereo matching algorithm based on deep learning: A survey. Journal of King Saud University—Computer and Information Sciences, 7, 1–11. https://doi.org/10.1016/j.jksuci.2020.08.011

[19] Hashmi, M. F., Katiyar, S., Keskar, A. G., Bokde, N. D., & Geem, Z. W. (2020). Efficient pneumonia detection in chest xray images using deep transfer learning. Diagnostics, 10(6), 1–23. https://doi.org/10.3390/diagnostics10060417

[20] Hassan, N. F., & Abdulrazzaq, H. I. (2018). Pose invariant palm vein identification system using convolutional neural network. Baghdad Science Journal, 15(4), 502–509. https://doi.org/10.21123/bsj.2018.15.4.0502

[21] He, K., Zhao, W., Xie, X., Ji, W., Liu, M., & Tang, Z. (2020). Synergistic learning of lung lobe segmentation and hierarchical multi-instance classification for automated severity assessment of COVID-19 in CT images. Pattern Recognition, 113(January), 1–13. https://www.sciencedirect.com/science/article/pii/S0031320321000157

[22] He, K., Zhao, W., Xie, X., Ji, W., Liu, M., Tang, Z., Shi, Y., Shi, F., Gao, Y., Liu, J., Zhang, J., & Shen, D. (2020). Synergistic learning of lung lobe segmentation and hierarchical multi-instance classification for automated severity assessment of COVID-19 in CT images. Pattern Recognition, 113(8), 1–11. https://doi.org/10.1016/j.patcog.2021.107828

[23] Hind Hameed Abid, D. M. (2016). Segmentation Brain Tumor and Diagnosing Using Watershed Algorithm. American Journal of Engineering Research (AJER), 5(11), 31–35. http://www.ajer.org/papers/v5(11)/F0511031035.pdf

[24] Hussain, E., Hasan, M., Rahman, M. A., Lee, I., Tamanna, T., & Parvez, M. Z. (2021). CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images. Chaos, Solitons and Fractals, 142, 110495. https://doi.org/10.1016/j.chaos.2020.110495

[25] Jaiswal, A. K., Tiwari, P., Kumar, S., Gupta, D., Khanna, A., & Rodrigues, J. J. P. C. (2019). Identifying pneumonia in chest X-rays: A deep learning approach. Measurement: Journal of the International Measurement Confederation, 145, 511–518. https://doi.org/10.1016/j.measurement.2019.05.076

[26] Abdalrada, Ahmad Shaker et al. (2019). "A Predictive Model for Liver Disease Progression Based on Logistic Regression Algorithm." *Periodicals of Engineering and Natural Sciences* 7(3): 1255–64. http://pen.ius.edu.ba/index.php/pen/article/view/667/395

[27] Joshi, R. C., Yadav, S., Pathak, V. K., Malhotra, H. S., Khokhar, H. V. S., Parihar, A., Kohli, N., Himanshu, D., Garg, R. K., Bhatt, M. L. B., Kumar, R., Singh, N. P., Sardana, V., Burget, R., Alippi, C., Travieso-Gonzalez, C. M., & Dutta, M. K. (2021). A deep learning-based COVID-19 automatic diagnostic framework using chest X-ray images. Biocybernetics and Biomedical Engineering, 41(1), 239–254. https://doi.org/10.1016/j.bbe.2021.01.002

[28] Kadhem, S., & Al-dargazli, M. (2012). Propose a Mixture Edge Detection Method for Infrared Image Segmentation. British Journal of Science, 6(2), 61–72. https://uotechnology.edu.iq/dep-cs/mypdf/research/2012/r8.pdf

[29] Kadhim, Q. K. (2016). Image Compression Using Discrete Cosine Transform Method. In International Journal of Computer Science and Mobile Computing (Vol. 5, Issue 9). https://www.researchgate.net/publication/308983925

[30] Kadhim, Q. K. (2017). Classification of Human Skin Diseases using Data Mining. International Journal of Advanced Engineering Research and Science, 4(1), 159–163. https://doi.org/10.22161/ijaers.4.1.25.

[31] Karim, A., Hassan, A., & Alawi, M. (2017). Proposed Handwriting Arabic Words classification Based On Discrete Wavelet Transform and Support Vector Machine. Iraqi Journal of Science, 58(2C), 5511–1168. https://doi.org/10.24996/ijs.2017.58.2c.19

[32] Khudhur, S. D., & Taqi, A. K. (2016). Edge Detection and Features Extraction for Dental X-Ray. Eng. & Tech. Journal, 34(September), 2420–2432. https://www.iasj.net/iasj/article/123659

[33] Kumar, A., Kalam, S., Kumar, C., & Sinha, D. (2020). TLCoV—An automated Covid-19 screening model using Transfer Learning from chest X-ray images. Chaos, Solitons & Fractals, 144(January), 1–12. http://dx.doi.org/10.1016/j.chaos.2021.110713

[34] Lara Hernandez, K. A., Rienmüller, T., Baumgartner, D., & Baumgartner, C. (2021). Deep learning in spatiotemporal cardiac imaging: A review of methodologies and clinical usability. Computers in Biology and Medicine, 130, 104200. https://doi.org/10.1016/j.compbiomed.2020.104200

[35] Masood, A., Sheng, B., Li, P., Hou, X., Wei, X., Qin, J., & Feng, D. (2018). Computer-Assisted Decision Support System in Pulmonary Cancer detection and stage classification on CT images. Journal of Biomedical Informatics, 79(June 2017), 117–128. https://doi.org/10.1016/j.jbi.2018.01.005

[36] Matheel Emaduldeen Abdulmunim Zainab Fawzy Abed. (2015). Classification Mammogram Images Using ID3 decision tree Algorithm Based on Contourlet Transform. Eng. & Tech. Journal, 33(3), 2015. https://www.iasj.net/iasj/article/105290

[37] Mohammed, M. A., Kadhem, S. M., & Ali, M. A. (2021). Insider Attacker Detection Based On Body Language and Technical Behavior Using Light Gradient Boosting Machine (LightGBM). Tech-Knowledge Journal, 1(1), 48–66. https://www.researchgate.net/publication/348937072

[38] Papoutsi, C., Drigas, A., & Skianis, C. (2018). Mobile applications to improve emotional intelligence in Autism—A review. International Journal of Interactive Mobile Technologies, 12(6), 47–61. https://doi.org/10.3991/ijim.v12i6.9073

[39] Prof, A., Ameerabdulla, A., Prof, A., & Ali, Y. (2016). A Survey : Medical Augmented Reality. International Journal of Emerging Trends & Technology in Computer Science, 5(2), 8–11. https://www.researchgate.net/publication/312029191_A_Survey_Medical_Augmented_Reality

[40] Rajesh, P., Murugan, A., Muruganantham, B., & Ganesh Kumar, S. (2020). Lung Cancer Diagnosis and Treatment Using AI and Mobile Applications. International Journal of Interactive Mobile Technologies, 14(17), 189–203. https://doi.org/10.3991/ijim.v14i17.16607

[41] Rao, S. (2021). IoT Enabled Wearable Device for COVID Safety and Emergencies. International Journal of Interactive Mobile Technologies, 15(3), 146–153. https://doi.org/10.3991/ijim.v15i03.17815

[42] Abed, S. I. (2021). Lung Cancer Detection from X-ray images by combined Backpropagation Neural Network and PCA. Engineering and Technology Journal, 37(5A), 166–171. https://doi.org/10.30684/etj.37.5a.3

[43] Sadiq, Ahmed T., & Shukr, N. H. (2013). Classification of Cardiac Arrhythmia using ID3 Classifier Based on Wavelet Transform. Iraqi J. of Sci., 54(4), 1167–1175. https://www.iasj.net/iasj/article/82695

[44] Sadiq, Ahmed Tariq, & Chawishly, S. A. (2012). Intelligent Methods to Solve Null Values Problem in Databases Intelligent. Journal of Advanced Computer Science and Technology Research, 2(December), 91–103. https://www.researchgate.net/profile/Ahmed-Sadiq/publication/267422791

[45] Sharma, N., Sharma, R., & Jindal, N. (2021). Machine Learning and Deep Learning Applications-A Vision. Global Transitions Proceedings, 2(1), 24–28. https://doi.org/10.1016/j.gltp.2021.01.004

[46] Sheikh, A., Anderson, M., Albala, S., Casadei, B., Franklin, B. D., Richards, M., Taylor, D., Tibble, H., & Mossialos, E. (2021). Health information technology and digital innovation for national learning health and care systems. The Lancet. Digital Health, 7500(21), 1–14. https://doi.org/10.1016/S2589-7500(21)00005-4

[47] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(1), 1–48. https://doi.org/10.1186/s40537-019-0197-0

[48] Suresh, S., & Mohan, S. (2019). NROI based feature learning for automated tumor stage classification of pulmonary lung nodules using deep convolutional neural networks. Journal of King Saud University—Computer and Information Sciences, 7(3), 1–12. https://doi.org/10.1016/j.jksuci.2019.11.013

[49] Suzuki, K., Otsuka, Y., Nomura, Y., Kumamaru, K. K., Kuwatsuru, R., & Aoki, S. (2020). Development and Validation of a Modified Three-Dimensional U-Net Deep-Learning Model for Automated Detection of Lung Nodules on Chest CT Images From the Lung Image Database Consortium and Japanese Datasets. Academic Radiology, 1, 1–7. https://doi.org/10.1016/j.acra.2020.07.030

[50] Turkoglu, M. (2020). COVID-19 Detection System Using Chest CT Images and Multiple Kernels-Extreme Learning Machine Based on Deep Neural Network. Irbm, 1, 1–8. https://doi.org/10.1016/j.irbm.2021.01.004

[51] Wu, X., Chen, C., Zhong, M., Wang, J., & Shi, J. (2021). COVID-AL: The diagnosis of COVID-19 with deep active learning. Medical Image Analysis, 68, 101913. https://doi.org/10.1016/j.media.2020.101913

[52] Xu, Y., Lam, H. K., & Jia, G. (2021). MANet: A two-stage deep learning method for classification of COVID-19 from Chest X-ray images. Neurocomputing, 443, 96–105. https://doi.org/10.1016/j.neucom.2021.03.034

[53] Yu, X., Wang, S. H., & Zhang, Y. D. (2021). CGNet: A graph-knowledge embedded convolutional neural network for detection of pneumonia. Information Processing and Management, 58(1), 102411. https://doi.org/10.1016/j.ipm.2020.102411

[54] Zhang, Q., Wang, H., Yoon, S. W., Won, D., & Srihari, K. (2019). Lung nodule diagnosis on 3D computed tomography images using deep convolutional neural networks. Procedia Manufacturing, 39, 363–370. https://doi.org/10.1016/j.promfg.2020.01.375

[55] Zhou, T., Lu, H., Yang, Z., Qiu, S., Huo, B., & Dong, Y. (2021). The ensemble deep learning model for novel COVID-19 on CT images. Applied Soft Computing, 98, 106885. https://doi.org/10.1016/j.asoc.2020.106885

7   Authors

Shaymaa Taha Ahmed M.Sc. (2015) in (India), currently a postgraduate student studying Ph.D. in Department of Computer Sciences, University of Technology, Baghdad, Iraq. Affiliation: the University of Diyala Dept.: computer science/College: basic of education Specialization: - Computer science/information system. Research Interests: Cloud Computing-Deep Learning-Machine learning-AI -Data mining Google Site: Google scholar: Google scholar: https://scholar.google.com/citations?user=GRl_liEAAAAJ&hl_ orcid.org/0000-0002-4986-2475/ ✉cs.19.25@grad.uotechnology.edu.iq & Shaymaa.taha.ahmed@basicedu.uodiyala.edu.iq & mrs.sh.ta.ah@gmail.com

Suhad Malallah Kadhem Ph.D. in Computer Sciences/ 2003/ Computer Science Department/the University of Technology. Scientific Specialization: Natural Language Processing. Scientific Title: Assistant Professor/ 20012. Scientific Research Interest: Natural Language Processing, Machine translation, Artificial intelligence, Information Security, Information Hiding, machine learning, deep learning, and data mining. Google scholar: https://scholar.google.com/citations?user=a6tG4E4AAAAJ&hl=ar Research-Gate: https://www.researchgate.net/profile/Suhad-Kadhem. Mail: 110102@uotechnology.edu.iq