Chapter 1

Introduction

1.1 Medical Imaging and chest radiographs

Medical imaging is the technique used to generate (two or three-dimensional) images of human physiological organs and structures without invasive procedures [1]. Medical imaging, one of the key developments in the 20th century for medical advancement, has increasingly developed our knowledge of human anatomy, physiology, and disease patterns [2]. Clinicians now have an objective-based medical diagnosis due to the interpretation of medical images, significantly improving patient care and treatment [3]. Two types of medical modalities exist: projection imaging and sectional (tomographic) imaging. In general, projection imaging techniques are inexpensive per examination and quick to acquire data. Image reconstruction is computationally simple since each investigation only requires one picture. However, projection imaging only creates two-dimensional graphics. On the other hand, tomographic imaging can rebuild volumetric three-dimensional images. However, the techniques are expensive and complex and require longer picture capture time. Image reconstruction is computationally challenging since each investigation requires multiple pictures. The algorithms required for reconstructing images and combining several measurements are computationally complex. The second basis of image classification types depends upon the radiation employed. The imaging techniques may depend upon ionizing and nonionizing radiation implementation. Nonionizing radiation, considered safe for patients, is magnetic resonance imaging (MRI), ultrasound, and magnetic particle imaging. In contrast, the popular ionizing radiation techniques that might lead to cell mutation are computed tomography (CT), positron emission tomography (PET), and conventional radiography (X-ray). Nevertheless, benefits like excellent spatial resolution, the contrast of bone structure, and imaging of metabolic processes by X-ray, CT, and PET surpass the risks [1].

The primary goal of radiology is to extract clinically relevant information from such images to diagnose and manage diseases. Radiologists use real-time imaging to direct them through blood vessels, arteries, and organs to the desired internal body structures during surgery [4]. Medical image analysis has emerged as a powerful diagnostic tool due to improved computer resources and the development of medical imaging techniques [1]. The purpose of image analysis is to create methods that offer radiologists pertinent information extracted from medical images. These methods provide repeatable, quantitative, and impartial evaluations of medical images. Experts who generally evaluate images qualitatively and subjectively can assist with medical image analysis for patient care.

There are three major categories for medical image analysis;

I. Image classification: Image classification gives a new image the appropriate class among categories. Image classification can help the radiologist detect the type of infection or lesion in the image [5].

II. Image registration: Image registration aligns two (or more) photographs to provide the anatomical correlation. CT and PET in medical imaging can show anatomical features and metabolic data, respectively. Image registration is needed to align both scans before overlaying them [6].

III. Image segmentation: Image segmentation is identifying various structures within an image. The segmentation of multiple organs, tissues, or pathologies is frequently of significant interest in medical imaging, such as detecting lesions, measuring the size of organs or lesions, and much more [7].

Medical image analysis has gone through phenomenal attainments over the last few years, primarily due to the incredible success of deep learning techniques that have made an outstanding performance in various imaging applications [8]. The process of conventional radiography includes projecting an item onto a detector in two dimensions. X-radiation, which is produced by the X-ray tube, penetrates objects. Depending on the varying densities and attenuation coefficients of materials (such as bones, tissues, and fluids), the intensity of X-rays is dispersed or attenuated [9].

Conventional radiography or X-ray has several advantages over other imaging methods, including a short examination time, good spatial resolution, a low cost per image, and a relative lack of artifacts (such as motion or reconstruction artifacts) [10]. Moreover, transportable radiography devices allow intensive care units to take X-ray images without transferring patients. Conventional radiography is one of the most effective medical imaging modalities due to its wide range of applications for various body parts and ailments.

Chest X-rays are primarily of three projection types: Posteroanterior (PA), lateral, and anteroposterior; how the radiation beam passes through the body determines how chest X-rays are typically characterized. The most common examinations of the thorax are PA and lateral. In the PA assessment, the patient stands straight and places his front chest against the detector. As a result, the radiation beam travels through the patient's back (posterior) and into the chest's anterior region. In the lateral examination, the patient is asked to stand with his left side up against the detector, his arms lifted, and radiation passed. AP examinations are done for patients who are bedridden or unable to stand generally. In contrast to PA, the patient lies with his posterior chest against the detector so that the radiation beam can pass through his chest from anterior to posterior. As a result of this location, interior features appear more pronounced in the X-ray image because the distance between the organs and the detector grows [11].

In a typical PA and lateral chest X-ray, the observer can typically see the trachea, clavicles, scapulae, ribs, diaphragm, heart, and vertebrae comprising the spine. However, being a projection image, chest X-ray images are incredibly challenging to interpret manually. This is primarily caused by overlapping anatomical structures and the inability to distinguish between visually identical diseases [12]. Nonetheless, deep learning algorithms powered by Artificial Intelligence (AI) can efficiently extract several image-based features that radiologists may be unable to observe manually in the original CXR. Regarding image feature extraction and classification, convolutional neural networks (CNNs) have proven their efficiency and are widely implemented by the research community [13].

1.2 Convolutional Neural Networks and AI-CAD

Artificial Intelligence-based Computer-Aided Diagnosis (AI-CAD) is a technology that uses artificial intelligence algorithms to assist radiologists in analyzing medical images to diagnose clinical conditions [14]. One application of AI-CAD is using trained image analysis algorithms to classify, detect, and quantify radiographic anomalies, improving the diagnostic process's accuracy and efficiency. Another application is the use of image analysis techniques such as image segmentation and registration algorithms to assist in accurately diagnosing clinical conditions. AI-CAD has the potential to transform the healthcare industry and improve image analysis, as digitized imaging data lends itself well to AI and machine learning (ML) analysis, making it easier to spot abnormalities and enhance disease management. Recent advancements in AI-CAD include implementing convolutional neural network (CNN) algorithms as deep learning methodologies [15].

A convolutional neural network (CNN) is a type of deep learning neural network that is primarily used for image recognition and processing. The human brain inspires its structure, consisting of multiple layers of interconnected nodes that hierarchically process information. CNN is designed to automatically learn spatial hierarchies of features using several neural nodes of convolution, pooling, and fully connected layers. The first layer of a CNN is usually a convolutional layer that applies filters to the input image and extracts specific features from the image. Next to convolution is the pooling layer, where the input is down-sampled to reduce the dimensionality of the feature maps. The subsequent layers of the CNN are typically fully connected layers that perform regression or classification tasks, and the output of the final layer signifies the predicted class [16]. CNNs have been used in various medical imaging and radiology applications, including image segmentation, registration, classification, and detection tasks. They have been shown to outperform traditional machine learning algorithms in many imaging tasks and have considerably advanced the state-of-the-art methods in AI-CAD [17].

CNNs are also commonly used for image segmentation tasks, where the goal is to identify and classify every pixel in an image into one of several predefined classes. Image segmentation is important for many applications in medical imaging. The basic idea behind using CNNs for image segmentation is to modify the architecture of the network to produce a pixel-wise classification rather than a single-class prediction. This is usually done by replacing the fully connected layers at the end of the network with a series of convolutional and up-sampling layers that produce a prediction map that matches the size of the input image [18].

One of the most popular architectures for image segmentation is the UNet, which consists of an encoder-decoder network. The encoder down-samples the input image and extracts features, while the decoder up-samples the feature maps and produces a pixel-wise classification. The encoder and decoders at corresponding layers are also connected by skip connections that directly transfer the image features from the encoder to the subsequent decoder [19]. Another popular architecture for image segmentation is the DeepLab, which uses atrous or dilated convolution to increase the receptive field of the network without increasing the number of parameters. This allows the network to make more informed decisions based on a broader context [20]. In conclusion, CNNs have proven to be a powerful instrument for image segmentation tasks, allowing for accurate and efficient pixel-wise classification of images. They have advanced the state-of-the-art in the imaging field and are expected to continue.

CNNs are predominantly used for image classification tasks, where the objective is to predict the class of an image from a set of predefined categories. CNNs have been shown to outperform traditional machine learning algorithms in many image classification tasks and have significantly advanced the state-of-the-art methods in medical imaging and AI-CAD [18]. The basic architecture of a CNN for image classification compromises a series of convolutional, pooling, and fully connected layers. The convolutional layers apply filters to the input image, each extracting a specific feature from the image. The pooling layers down-sample the convolutional layers' output to reduce the feature maps' dimensionality. Finally, the fully connected layers perform classification tasks based on the extracted features [21]. One of the most popular architectures for image classification is the VGG network, which consists of 16 or 19 convolutions and fully connected layers. The VGG network achieved superior performance on the ImageNet dataset in 2014, and its architecture has been followed as a basis for many subsequent CNNs [22]. Another popular architecture for image classification is the ResNet network, which uses residual connections to enable deeper networks without vanishing gradients. The ResNet network won the ImageNet classification challenge in 2015 and introduced the concept of residual learning, which has since been widely adopted in deep learning [23]. In conclusion, CNNs have revolutionized the field of image classification and have led to significant advances in medical imaging and computer-aided diagnosis. They have enabled accurate and efficient classification of medical images, with applications ranging from tumor detection [24], Cardiovascular disease diagnosis [25], Alzheimer's disease prediction [26], pneumonia detection [27], lung nodules detection [28], and much more.

1.3 Challenges in deep learning-based chest radiograph analysis

Current chest radiographs typically have high image size and resolution. A large image size is necessary to image the chest region fully, and radiologists need a high spatial resolution to recognize the minute details of different lung diseases. However, many convolutional neural networks for computer vision have an input size of roughly 224 to 299 pixels per square for image classification. The original image is frequently downscaled to the input size via bilinear interpolation to fix the inconsistency between the original image size and the input size. This downscaling lowers the spatial resolution and may impede key features' visibility [29]. However, developing and implementing a robust neural network for image feature selection may supersede the concern [30].

The availability of annotated datasets is one of the significant challenges for any CNN model training. The first two chest X-ray datasets, The "JSRT" dataset from Shiraishi et al. [31] and the "PLCO-Lung" dataset from Team PLCO Project et al. [32], were made accessible to the public in the year 2000. Notably, a small dataset comprising 247 images for lung nodule categorization was published by Shiraishi et al. The digitalized film images of the chest X-rays have a 2048 × 2048 pixel image size and a 12-bit grey level. The PLCO-Lung dataset provides extensive annotation for 13 pathologies and is relatively large (236,000 images from 70,000 patients). The location descriptions and total count for each pathology are included. The images come as TIFF files with a 2500 × 2100 pixel image size and a 16-bit grey level. After that, several chest X-ray deserts were presented for education and research purposes, such as Montgomery and Shenzhen datasets by Jaeger et al. [33], released in 2014, containing 132 and 662 tuberculosis-infected CXR images, respectively. RSNA chest X-ray pneumonia dataset released by RSNA (Radiology Society of North America) in 2020 compromises 26,684 pneumonia-infected CXRs. COVID-19 Radiography datasets released by Rahman et al. [34] in 2020 comprise 11,958 COVID-19-infected CXR images. However, large and diverse annotated datasets still require more robust and precise AI-CAD systems developments.

After the AI-CAD system development, its transformation into a clinical application is another challenge. Only a few software solutions are currently offered for automatic chest X-ray analysis, despite the fact that clinical applications are the main focus of the majority of research in the field and consist of the capabilities to detect diseases in CXR images; the current clinical environment and regulatory concerns are what lead to the majority of issues with clinical applications for lung disease diagnosis [35]. However, implementing a large and diverse dataset and high testing accuracy may increase the possibility of acceptance as clinical applications [36].

1.4 Background and Previous Works

Deep learning has shown great potential in improving the accuracy and efficiency of chest X-ray interpretation. These systems use deep neural networks to learn complex image features and have shown promise in detecting abnormalities in medical images, such as brain tumors, lung nodules, pneumonia, atherosclerosis, and breast lesions [37]. The first AI-CAD system for mammography was developed in the 1990s, and there have been significant advancements since then.

Previous developments in deep learning for chest X-rays include:

I. Convolutional Neural Networks (CNNs): CNNs are a type of deep learning algorithm that can automatically learn and extract features from images. CNNs have been used to detect pneumonia and lung nodules in chest X-rays [27, 28].

II. Transfer learning: Transfer learning involves using pre-trained CNNs on large datasets, such as ImageNet, to improve the performance of models trained on smaller medical imaging datasets. Transfer learning has been used to enhance the performance of CNNs in detecting pulmonary nodules and pneumothorax and predicting radiological diagnoses [38, 39].

III. Attention mechanisms: Attention mechanisms selectively target regions of an image relevant to a particular task. Attention mechanisms have been used to improve the performance of CNNs in detecting and localizing abnormalities in chest X-rays [40].

IV. Generative Adversarial Networks (GANs): GANs are a type of deep learning algorithm that can generate realistic images. GANs have been used to generate synthetic chest X-rays to augment small medical imaging datasets [41].

These developments in deep learning for chest X-rays have shown promising results and have the potential to improve the accuracy and efficiency of chest X-ray interpretation. Several researchers have developed and implemented numerous deep-learning methods for chest X-ray analysis, especially in pneumonia and COVID-19 detection. Deep learning approaches have shown significant advancement in chest X-ray interpretation in recent years.

1.4.1 Previous works in chest X-ray image segmentation

Chest X-ray contains a large area as a non-region of interest in the form of background. The unwanted region may interrupt or misguide the AI system for accurate diagnosis. That's why segmenting the region of interest, i.e., lungs from the chest X-ray, is a significant part of the detection system training and development. Previously, several researchers have introduced and implemented different methods to segment chest X-ray images, and many of them have reported significant outcomes. Table 1.1 represents the previous developments with methods utilized and results obtained by several authors during recent years. Hooda et al. [42] applied a novel deep CNN on the JSRT CXR dataset and achieved an accuracy of 98.92% with a Jaccard index of 95.88%. Ngo et al. [43] applied a combination of Distance Regularized Level Set and Deep Belief Network to segment the JSRT dataset and achieved an accuracy of 96.5%. Saidy et al. [44] also utilized the JSRT dataset for an encoder-decoder-based segmentation model development and achieved a dice coefficient of 96% on the test dataset. Mittal et al. [45] utilized the combination of JSRT and Montgomery CXR datasets for an encoder-decoder-based segmentation model and achieved an accuracy of 98.73% and a Jaccard index of 95.10%. Souza et al. [46] applied a combination of AlexNet and ResNet-based CNN segmentation models on the Montgomery dataset and achieved the accuracy, dice, and

Jaccard of 96.67%, 93.56%, and 88.07%, respectively. Reamarron et al. [47] applied the total variation-based active contour method for the segmentation and used a combination of the JSRT and Montgomery datasets. The model achieved a dice of 89%. Gaal et al. developed a novel segmentation method and applied it to the JSRT dataset. They got a dice coefficient of 97.5%. Munawar et al. [48] utilized three datasets, JSRT, Montgomery, and Shenzhen, for the training of the Generative Adversarial Network and achieved a dice coefficient of 97.4%. Zhang et al. [49] applied the Dual Encoder Fusion UNet model on a combination of Montgomery and Shenzhen datasets. They achieved an accuracy of 98.04% with dice and an AUC of 96.67% and 0.98, respectively. Teixeira et al. [50] applied the UNet model on the combination of five datasets, namely Cohen, JSRT, Montgomery, Shenzhen, and a private dataset. They achieved a dice coefficient of 98.2%.

Table 1.1. Previous developments in chest X-ray image segmentation.

Author & Year	Dataset (chest X-ray)	Technique	Accuracy	Dice	Jaccard	AUC
Hooda et al. (2018) [42]	JSRT	New deep CNN	98.92%	N.A.	95.88%	N.A.
Ngo et al. (2015) [43]	JSRT	DRLS (Distance Regularized Level Set) + DBN(Deep Belief Network)	96.5%	N.A.	N.A.	N.A.
Saidy et al. (2018) [44]	JSRT	Encoder-decoder neural network	N.A.	96%	N.A.	N.A.
Mittal et al. (2018) [45]	JSRT+Montgomery	Encoder-decoder neural network	98.73%	N.A.	95.10%	N.A.
Souza et al. (2019) [46]	Montgomery	AlexNet+ResNet based CNN	96.97%	93.56%	88.07%	N.A.
Reamarron et al. (2020) [47]	JSRT+Montgomery	TVAC(Total Variation-based Active Contour)	N.A.	89%	N.A.	N.A.
Gaal et al. (2020) [51]	JSRT	New deep CNN	N.A.	97.5%	N.A.	N.A.
Munawar et al. (2020) [48]	JSRT+Montgomery+ Shenzhen	GAN (Generative Adversarial Networks)	N.A.	97.4%	N.A.	N.A.
Zhang et al. (2021) [49]	Montgomery+ Shenzhen	DEFUNet(Dual Encoder Fusion UNet)	98.04%	96.67%	N.A.	0.98
Teixeira et al. (2021) [50]	Cohen v7labs+JSRT+Montgomery+ Shenzhen+Private	UNet	N.A.	98.2%	N.A.	N.A.

1.4.2 Previous works in chest X-ray image classification

Recently, COVID-19 detection using deep learning techniques has become a very popular area. Several researchers have proposed deep learning methods for the detection of disease in CXR images. However, the majority of them employed a limited dataset with a small number of COVID-19 samples. Consequently, their outputs may not be generalized, and accuracy cannot be a covenant on the larger dataset. Therefore, their system may not meet the regulatory requirements for a clinical setup. Table 1.2 shows the previous developments with methods applied and results for classification problems by different authors in recent years. Alom et al. [52] utilized the Kaggle dataset, which has 390 COVID-19 images and 234 normal images. They applied a novel NABLA-N network for the segmentation with an accuracy, dice, and Jaccard of 94.66%, 88.46%, and 86.50%, respectively. Afterward, they applied the Inception Recurrent Residual Neural Network model for the classification of segmented lung images into two classes. They achieved a classification accuracy of 87.26% and an AUC of 0.93. Oh et al. [53] utilized 180 COVID-19 and 322 other images taken from Kaggle and GitHub. They applied the DenseNet103 network for the segmentation and achieved the Jaccard of 95.5%. After the segmentation, they applied the ResNet-18 model to classify the segmented lung images into four classes and achieved an accuracy of 88.9%. Nayak et al. [54] performed binary classification into COVID-19 and normal class using 406 CXR images. Using the transfer learning method, the authors applied eight different pre-trained neural networks and obtained a maximum accuracy of 98.33% using the ResNet34 network. When it comes to the fusion of machine learning and deep learning. Choudhury et al. [55] applied eight different deep learning pre-trained CNNs for the classification of CXR images having three classes named COVID-19, viral pneumonia, and normal, with a total of 423, 1485, and 1579 images for each class, respectively. The authors showed an accuracy of 97.74% by CheXNet for three classes with the equivalent precision, sensitivity, and F1-score of 96.61% and specificity of 98.31%. Jain et al. [56] applied several pre-trained CNNs for the classification of CXR images into three classes: COVID-19, VP, and normal.

They utilized 490 COVID-19 images and got the maximum accuracy of 97.97% using the Xception model. Khan et al. [57] introduced a novel network, Coronet, inspired by Xception architecture. Using the Coronet model, the authors obtained an accuracy of 95% for three-class classification into COVID-19, VP, and normal. They also performed four-class classification into COVID-19, VP, BP, and normal with 89.6% accuracy. Hussain et al. [58] developed a novel deep neural network (DNN) named CoroDet. The authors used CXR images having four classes named COVID-19, viral pneumonia (VP), bacterial pneumonia (BP), and normal, with an image size of 500, 400, 400, and 800 for each class, respectively. They performed the classification experiment into two-class (COVID-19 *vs*. normal), three-class (COVID-19, VP, and normal), and four-class (COVID-19, VP, BP, and normal) with the maximum accuracy of 99.1%, 94.2%, and 91.2% for each experiment respectively. Wehbe et al. [59] utilized a private dataset having 4253 COVID-19 images and 14778 normal images. They applied an ensemble network for the classification of CXR images after the segmentation. They achieved an accuracy of 83% and an AUC of 0.9 for the two-class classification. Teixeira et al. [50] utilized the RYDLS-20-V2 dataset, having 503 COVID-19 and 2175 images from other classes. They applied the UNet model for the segmentation with a dice coefficient of 98.2%. Following segmentation, they applied Inception V3 for classification into three classes and achieved an accuracy of 88% and an AUC of 0.9. Keidar et al. [60] applied the ensemble method for the classification of segmented lung images into two classes. Their model performed with an accuracy of 90.3% and an AUC of 0.96. Nikolaou et al. [61] developed a novel CNN by modifying pre-trained EfficientNetB0. This network was applied for the binary (COVID-19 and normal) and three-class (COVID-19, pneumonia, and normal) classification, obtaining an accuracy of 95% for binary and 93% for three-class experiments. Yang et al. [62] applied the VGG16 network to classify into two and three classes. They utilized 3616 COVID-19 and 4845 other images and achieved an accuracy of 98% for two and 97% for three-class classification. AI-Timemy et al. [63] performed the five-class classification into COVID-19, VP, BP, TB, and normal class. They utilized 2,186, consisting of 435 COVID-19 images, for the experiment. The authors applied a combination of DL and ML methods and got 91.6% accuracy. Abdulah et al. [64] applied the Res-CR-Net model for the segmentation with dice and Jaccard of 98% each. Thereafter they classified a private dataset into two classes using an ensemble method and achieved an accuracy of 79% and an AUC of 0.85. Bhattacharyya et al. [65] used a GAN segmentation network with a VGG-19 and Random Forest classifier and achieved 96.6% accuracy for the three-class classification. Xu et al. [66] utilized 433 COVID-19 and 6359 other images. They applied ResUNet for the segmentation with a Jaccard of 92.50%. After that, they applied ResNet50 to classify segmented lung images into five classes. They achieved an accuracy of 96.32%. Fang et al. [67] applied a novel CLseg model for segmentation and achieved a dice of 94.09%. After the segmentation, they applied a novel SC2Net model for the two-class classification of the COVIDGR 1.0 dataset and achieved an accuracy of 84.23% and an AUC of 0.94. Khan et al. [68] applied the EfficientNetB network for the classification into four classes and achieved an accuracy of 96.13%. Our previous work used 3611 COVID-19 and 13833 other images to classify them into two, three, and five classes. We applied VGG16, NasnetMobile, and DenseNet201 models and achieved an accuracy of 99.84%, 96.63%, and 92.70%, with an AUC of 1.0, 0.97, and 0.92 for two, three, and five-class classifications, respectively. Hertel et al. [69] utilized 4013 COVID-19 with 12837 other class images. They applied a ResUnet segmentation network with a dice of 95%. Following segmentation, they applied an ensemble network to classify into two and three classes. They achieved an accuracy of 91% for the two-class and 84% for the three-class with an AUC of 0.95. Aslan et al. [70] applied an ANN-based segmentation method on the COVID-19 Radiography database (Kaggle) and a combination of DenseNet201 and SVM for the classification into three classes. They achieved an accuracy of 96.29% with an AUC of 0.99.

Table 1.2. Previous developments in chest X-ray image classification.

Author & Year	Segmentation method & Results (%)	Dataset & Number of chest X-ray images (COVID-19 + other)	Classification methods	Classification Accuracy	AUC
Alom et al. 2020) [52]	NABLA-N network Acc[1] - 94.66, Dice - 88.46, Jaccard - 86.50	Kaggle (390+234)	Inception Recurrent Residual Neural Network (IRRCNN)	87.26%	0.93
Oh et al. (2020) [53]	DenseNet103 Jaccard - 95.5	Kaggle+GitHub (180+322)	ResNet-18	4 class-88.9%	NA
Nayak et al. (2020) [54]	Not Implemented	GitHub (203+203)	ResNet-34	2 class-98.33%	0.98
Choudhury et al. (2020) [55]	Not Implemented	Covid-19 Radiography database (423+3064)	CheXNet	3 class-97.74%	NA
Jain et al. (2020) [56]	Not Implemented	Kaggle (490+5942)	Xception	3 class-97.97%	NA
Khan et al. (2020) [57]	Not Implemented	GitHub (284+967)	Coronet (novel CNN)	3 class-95%	NA
Hussain et al. (2020) [58]	Not Implemented	COVID-R dataset (500+1600)	CoroDet (novel CNN)	2 class-99.1% 3 class-94.2% 4class-91.2%	NA
Wehbe et al.(2021) [59]	N.A.	Private (4253+14778)	Ensemble CNN	2 class-83%	0.9
Teixeira et al.(2021) [50]	UNet Dice - 98.2	RYDLS-20-v2 (503+2175)	Inception V3	3 class-88%	0.9
Keidar et al. (2021) [60]	N.A.	Private (1289+2427)	Ensemble model	2 class-90.3%	0.96
Nikolaou et al. (2021) [61]	Not Implemented	Covid-19 Radiography database (3616+11537)	EfficientNetB0	2 class-95% 3 class-93%	NA
Yang et al. (2021) [62]	Not Implemented	Covid-19 Radiography database (3616+4845)	VGG16	2 class-98% 3 class-97%	NA
Timemy et al.(2021) [63]	Not Implemented	GitHub (435+1751)	ResNet-50 + ESD[2]	5 class- 91.6%	NA
Abdulah et al. (2021) [64]	Res-CR-Net Dice - 98, Jaccard - 98	Private (1435+3797)	Ensemble CNN	2 class-79%	0.85
Bhattacharyya et al. (2021) [65]	GAN network Acc - N.A.	GitHub (342+687)	VGG-19 + Random Forest	3 class-96.6%	NA
Xu et al. (2021) [66]	ResUNet Jaccard - 92.50	GitHub (433+6359)	ResNet50	5 class-96.32%	N.A.
Fang et al. (2022) [67]	CLSeg Dice - 94.09	COVIDGR 1.0 dataset (426+426)	SC2Net (novel CNN)	84.23%	0.94
Khan et al. (2022) [68]	Not Implemented	Covid-19 Radiography database (3616+17449)	EfficientNetB	4 class-96.13%	N.A.
Hertel et al. (2022) [69]	ResUnet Dice - 95	COVIDx5+MIDRC-RICORD-1C+BIMCV dataset (4013+12837)	Ensemble model	2 class-91% 3 class-84%	0.95
Aslan et al. (2022) [70]	ANN based segmentation Accuracy - N.A.	Covid-19 Radiography database (Kaggle) (219+2905)	DensenNet201+SVM	3 class-96.29%	0.99

[1]Acc – Accuracy, [2]ESD - Ensemble Subspace Discriminant.

1.5 Objective

The proposed work aims to develop a precise and robust deep learning-based system to detect diverse pneumonia types, including COVID-19, in the population. For COVID-19 detection in chest X-rays and CT, AI and deep learning methods have shown tremendous successful classification accuracies. However, most of them conducted classification without lung segmentation or executed only two or three-class (pneumonia types) classification. Some authors reported work on the segmentation-based classification model (since they remove unwanted regions of interest). However, high accuracy to qualify for a clinical setup in a multiclass framework was still missing. As a result, these systems cannot be adapted for clinical practice since they cannot meet the regulatory requirements of an error rate of <5%. Our hypothesis states that the diagnostic system can be adapted in clinical settings if a segmentation-based classification system could be designed with an error rate of < 5%, typically adopted for 510 (K) regulatory purposes. Additionally, the intent was to implement an explainable AI method based on the Grad-CAM heatmap to detect and manifest the lesion in the X-ray scans.

Several deep learning methods were implemented for segmentation and classification problems to develop a robust and accurate system for pneumonia detection. The strategy included implementing a transfer learning approach, modifications in standard networks, hybrid model implementation, and ensembling the networks. Additionally, many CNNs were trained to get the best results. Along with classification, different segmentation models were implemented to segment the chest X-ray and gain high accuracy and precision in the classification. Models were refined to get the best results, and model tuning was ensured by applying different optimizers, activation and loss functions, and learning rates. For the stability and robustness of the system, a considerable number of chest X-ray images taken from different big databases were employed for training. Also, extensive multiclass classification experiments were performed that were never done before. The purpose behind comprising images from comprehensive pneumonia classes was to make our system exceedingly profound in diagnosing all possible pneumonia types that persist in the population.

The UNet network has been shown to be powerful for lung region segmentation in X-ray scans. The model is ideal due to its ability to extract the grayscale features in supervised-based segmentation. The power of contextual and semantic features in low-lying layers and high-lying layers allows UNet-based architecture to extract the feature in the segmentation paradigm. The concatenation phase via skip connection allows for the recovery of the best features from the encoders. The upsampling in the decoder phase is equally powerful for the reconstruction of the image size while retaining the features. The UNet+ and UNet++ models have intermediate encoder stages between compression and expansion. The intermediate up-sampling units with varying depths in the UNet+ and UNet++ models have overcome the limitation of optimal depth in the UNet encoder-decoder network. These days, advanced hybrid models are in the pipeline that use UNet in combination with other networks such as SegNet-UNet, ResNet-UNet, VGG-UNet, and SegNet-UNet+. Some attention UNet models are also very popular in recent days that use UNet as a backbone for the segmentation of medical images. Based on their popularity, compatibility, reliability, and results, we selected three different versions of UNet: UNet, UNet+, and UNet++ networks for the segmentation of the chest X-ray images.

Delineating this thesis in Chapter 2, the implementation of seven transfer learning-based CNN networks has been presented. The networks VGG16, VGG19, DenseNet201, Xception, InceptionV3, NasnetMobile, and ResNet152 were implemented for chest X-ray image classification into binary, three, and five classes. Further, the networks were evaluated using different matrices, including ROC (receiver operating characteristics) and AUC (area under the curve). Chapter 3 deals with segmenting chest X-ray images using different versions of the UNet model. Further, the models were evaluated using accuracy, dice, and Jaccard. Finally, the best model was chosen to segment the pneumonia chest X-ray images dataset. In Chapter 4, the classification of the segmented CXR pneumonia dataset was revealed. The eight different CNNs were implemented for classification into binary, three, and five classes. Further, the performance of each model was evaluated. The visualization capabilities of the model were also investigated using the Gradient-weighted Class Activation Mapping (Grad CAM) to view the abnormalities in chest X-rays and discuss them from radiological perspectives. Chapter 5 deals with implementing three ensemble models to classify chest X-ray images. The performance of each model was evaluated using different evaluation metrics, AUC and ROC. The Grad CAM visualization technique was also employed for lesion detection in the images and was discussed from a radiological perspective.