Master's Research Project Report

A Survey of Convolutional Neural Networks on COVID-19 Diagnosis

YU Fangyu

Supervisor: FUJINAMI Tsutomu

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
(Knowledge Science)

March, 2023

Abstract

The Covid-19 pandemic has caused a worldwide health crisis, leading to significant disruptions and challenges for public health professionals. With the rise of the pandemic, the use of artificial intelligence (AI) technologies, such as convolutional neural networks (CNNs), has increased as a promising tool for aiding diagnosis and treatment. CNNs are a type of deep learning algorithm that has been effectively used in various imaging tasks, including image classification and object detection.

In this paper, we present a comprehensive survey of the current literature on the application of CNNs in the diagnosis of Covid-19. We begin by introducing medicine-related concepts such as X-rays and lung infection diseases, with a specific focus on Covid-19. This provides the necessary background information for understanding the significance of using CNNs for Covid-19 diagnosis.

We then systematically present a summary of the CNN architecture and mechanism, including its technical aspects such as convolutional and pooling layers, activation functions, and loss functions. This provides a comprehensive understanding of the workings of CNNs, which is crucial for understanding their application in the diagnosis of Covid-19.

We also reproduce a Covid-19 chest X-ray image classification experiment using an open-source model and dataset. The experiment provides practical insights into the performance of CNNs in the diagnosis of Covid-19 and highlights their potential as an effective tool. With the help of saliency maps, we also explore the implications of CNNs-based Covid-19 diagnosis, specifically the areas that the model considers important. This analysis provides valuable insights into the workings of CNNs and highlights the need for further research in this area.

We conclude the paper by discussing the current research trends in using CNNs for Covid-19 diagnosis and the contribution of we to this survey. Based on the previous, we provide suggestions for future directions in this field.

The survey presented in this paper demonstrates the potential of CNNs as an effective aid to Covid-19 diagnosis. The advantages of CNNs include their ability to classify images and detect objects accurately, as their speed and flexibility. However, the accuracy of CNNs depends on the availability of high-quality training data, and further research is needed to improve the accuracy and robustness of the models.

In conclusion, we provide a comprehensive overview of the current literature on the application of CNNs in the diagnosis of Covid-19. The survey highlights the potential of CNNs as an effective tool in aiding Covid-19 diagnosis while also pointing out the need for further research to improve

their accuracy and robustness. This paper serves as a valuable resource for researchers, practitioners, and students interested in the application of AI technologies in the diagnosis of Covid-19.

Keywords: covid-19, convolutional neural network, saliency map

Acknowledgments

I am deeply grateful to my supervisor, Professor Tsutomu Fujinami, for his unwavering guidance, support, and mentorship throughout my graduate studies. His expertise in the field and his dedication to his students have been invaluable assets to my academic and personal growth. His insights, feedback, and constructive criticism have been instrumental in shaping my thesis and my understanding of the subject. I am truly honored to have had the opportunity to work under his guidance.

I would also like to express my sincere gratitude to Dr. Dinh Duy Tai for his invaluable contributions to my thesis. His expertise, knowledge, and experience have been essential in shaping my research and understanding of the subject. His comments, suggestions, and feedback have been extremely helpful in developing my thesis. I am truly grateful for his time and effort.

I would also like to express my heartfelt appreciation to my husband, Dr. Yu Yang, for his unwavering support, love, and encouragement throughout the journey of completing my Master's thesis. His patience, understanding, and constant motivation have been my rock during the challenges and stress of graduate school. His love and support have been essential to my life, and I couldn't have done it without him.

I would also like to acknowledge the support of the department and the university for providing me with the resources and facilities necessary to conduct my research. I am deeply grateful for the opportunity to have studied and worked in this academic environment.

Finally, I would like to thank all the people who have contributed in any way, directly or indirectly, to the completion of my master's thesis. Your support and encouragement have been invaluable, and I am truly grateful for all that you have done for me.

This thesis is the culmination of a journey that would not have been possible without these individuals' guidance, support, and contributions. I am deeply grateful for all they have done for me and will always be grateful for their mentorship.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Research Background

Nowadays, medical imaging has become an indispensable part of disease diagnosis and treatment. However, due to the impact of coronavirus disease 2019 (COVID-19), it is necessary to improve medical image screening and clinical management. As we all know, the symptoms of COVID-19 patients are mainly fever, sore throat, and cough. These symptoms can also be seen in patients with influenza and pneumonia. This creates challenges for physicians to diagnose COVID-19 patients early. Studies prove that disease classification from chest X-ray images has become an inexpensive and reliable option to aid medical diagnosis. However, a radiologist will need 5 to 15 minutes to analyze a suspected patient's chest X-ray images and make a clinical diagnosis, and sometimes the scan includes more than 300 images. Such efficiency cannot cope with the situation of significant population infections. In this background, convolutional neural networks (CNNs) can be a great helpful tool for physicians in medical image analysis [2].

With the increasing computing power and data volume, as well as the continuous development of medical imaging technology, the use of artificial intelligence (AI) to assist medical image analysis has played an essential role in clinical diagnosis and treatment. John McCarthy coined the term AI at a conference held in 1956. However, the possibility of modern computers achieving human intelligence was raised earlier by Alan Turing [3]. AI is a broad branch of computer science that involves building intelligent machines able to perform tasks that normally require human intelligence. A lot of artificial intelligence is already being used in the medical field.

In recent years, deep learning (DL), especially CNNs, has rapidly developed into a research hotspot in medical image analysis [4]. Conventional

medical image analysis uses edge detection, texture analysis, and other methods to extract image features manually. In contrast, deep learning is a data-driven approach that can automatically learn the relevant model and data features from large-scale data sets for specific problems [5]. Because there is a big difference between medical image and natural image analysis, it is vital to independently analyze the application of CNN in medical images.

1.2 Research Motivation

Research from medicine filed [6, 7] has already pointed out the increasing workload of radiology. On the other hand, The statistical analysis result [8] from an anonymized, self-administered questionnaire carried out among radiology residents enrolled in the Saudi Board of Radiology, Saudi Arabia, reveals approximately 60% of participants consider AI technology will first contribute to breast-related radiological subspecialties.

Inspired by the previous study, in this survey, we aim to investigate the applications of AI technology in medical image analysis, especially chest X-ray images. In addition to introducing medical and CNN knowledge, we test the AI model's performance by reproducing a chest X-ray classification experiment. We adopt and slightly modify a CNN-based classification model and collect a new dataset containing 33,920 chest X-ray images from the open data source. We further enhance our model by tuning the hyperparameters to provide better generalization capabilities during the model training phase. As the initial performance after the training, 90% of COVID-19 cases were successfully identified using our re-trained CNN model based on chest X-ray images. We also implemented a 5-fold cross-validation to reduce variability. Normal, COVID-19, and Non-COVID infections were classified with similar accuracy to the initial performance.

1.3 Thesis Outline

The rest structures of the paper are organized as follows. Chapter 2 shows the information on medical imaging and lung infection. It also presents previous medical image analysis and classification studies via deep learning. The methodology details will be introduced in Chapter 3. In Chapter 4, we will explain our datasets and the model architecture used in this study. Then, we will show the experiment results. Last but not least, we will summarize the paper and future work for medical image analysis in Chapter 5.

Chapter 2

Background and Literature Review

2.1 Medical Imaging

Medical imaging technologies have played an important role in healthcare and have become indispensable to disease diagnosis and treatment. It referred to server types of machines and devices used to view human body parts or organs. Most medical imaging procedures are non-invasive and painless ways to detect foreign objects early or ensure everything in the body is normal. Because medical imaging offered faster and more reliable information, it drastically improved patient outcomes and helped physicians achieve better results. Each type of technology provides different information about the area of the body related to possible injury or disease. Different types of medical imaging procedures include X-rays, Computerized tomography (CT scan), Magnetic resonance imaging (MRI), and Ultrasounds, etc. In this article, we mainly focus on X-rays images.

X-rays imaging X-rays are a form of electromagnetic waves with high energy that can pass through objects, including the human body. It can create images of the internal body. The images show the body parts in different colors because different tissues absorb different amounts of radiation. Calcium in bones absorbs the most X-rays, so bones appear white on the X-ray image. Fat and other soft tissues absorb less and appear gray. The lungs are filled with air and air absorbs the least, so the lungs appear black. X-rays imaging is used for many purposes. The common use of X-rays is to check for fractures. In addition, chest X-rays can look for mass, pneumonia and COVID-19, etc.

Computerized tomography (CT scan) A CT scan uses a computer connected to an X-ray machine to take X-rays images from different angles of your body. After that, computer processing is used to create cross-sectional images of the soft tissues, blood vessels and bones inside the body. CT scan images provide more detailed information about structures than plain X-rays. A CT scan can be used to visualize almost all parts of the body. It is particularly suitable for quickly examining those who may have internal injuries or internal bleeding due to car accidents or other types of trauma. CT scans are usually not used for screening if you don't have any symptoms.

Ultrasounds Ultrasounds also known as sonogram, is an imaging procedure that uses sound waves to create images of structures inside the body. Most ultrasound examinations are done using a small device called an ultrasound probe outside the body. The probe emits high-frequency sound waves. When sound waves bounce off different body parts, they create "echoes" that are picked up by the probe and turned into a moving image. Ultrasound scan is mainly used to monitor an unborn baby's growth, diagnose a disease, or guide a surgeon during certain procedures.

Magnetic resonance imaging (MRI) MRI is a medical imaging technique that uses a magnetic field and radio waves to produce detailed images of the organs and structures in the human body. It produces cross-sectional images of the internal of the body that helps diagnose various problems. Most MRI scanners are large magnets tubes requiring the patient to lie inside during the scan. Healthcare professionals use MRI scans to examine almost any part of the body. It is very useful testing of the brain and spinal cord. The results of an MRI scan can be used to help diagnose, plan treatments and assess how effective previous treatments have worked.

2.1.1 The history of X-rays

German physicist Wilhelm Konrad Röntgen (1845-1923) discovered X-rays in 1895. While investigating the effects of cathode-ray tubes in his lab, Röntgen uncovered a glow from a screen coated with fluorescent a few feet away from the tube, even though the tube is shielded with black paper from ultraviolet and direct visible light. The cathode-ray tube consisted of a glass bulb with positive and negative electrodes encapsulated. The tube produces fluorescence when the air in the tube is evacuated and a high voltage is

applied. Röntgen deduced that the invisible radiation from the discharge tube passes through the air and causes the screen to glow. He proved that transparent objects were the new form of ray. He called them "X-rays" to indicate that the observed radiation differed from previously recorded. The X-ray could pass through substances like human tissue and paper and cast shadows on solid objects like bone or metal. Röntgen dramatically demonstrated this by producing a photograph of the bones in his wife's hand and the ring on her finger. His discovery of X-ray has caused excitement in the scientific community and the public worldwide. This has given rise to new possibilities in physics and the study of the structure of matter. Besides, there is great enthusiasm for the potential application of X-rays to aid medical and surgical procedures.

X-rays became an important medical diagnostic tool, allowing doctors to see inside the human body for the first time without surgery. In the month following the announcement of this discovery, surgeons used medical radiographs to guide their work. In June 1896, only six months after Roentgen announced his discovery, doctors used X-rays on the battlefield to locate the bullets of wounded soldiers. X-ray technology is used in many areas, such as medicine, materials analysis and airport security scanners.

2.1.2 The types of X-rays

When examining a patient's health, many different types of X-rays are needed for various reasons. For example, blood tests and physical exams don't give your doctor enough clues to discover what's causing your symptoms. The doctor will use image-generating diagnostic tools such as X-rays. Common examples of X-rays used in radiology include:

- Chest X-rays take images of the lungs, heart, arteries, and ribs. The doctor will look for pneumonia, COVID-19, tuberculosis, lung cancer, heart size problems, broken ribs ,or spine with a chest X-ray. Abdominal X-rays can visualize organs and structures within the abdomen, including the stomach, spleen ,and intestine. This will help the doctor to identify conditions that might cause abdominal discomforts, such as kidney stones or any damage to abdominal tissues.

- Skeletal X-rays are used to detect fractured or inflammation bones in the body. Only the damaged joint will be scanned such as knees, ankles, shoulders, wrists, or hips.

- Dental X-rays provide images of the teeth and mouth to examine oral health. Structures such as teeth, tissues, and fluids will appear shades

of gray on the X-ray.

2.2 Lung Infection

Lung infection can be understood as an infection in the lung area. It will happen when a virus, fungus or bacteria enters your lungs and causes inflammation. There are many types of lung infections.

2.2.1 Pneumonia

One of the most common lung infections is pneumonia. It inflames the tissue in one or both lungs. The air sacs may be filled with fluid or pus causing coughing, fever, and difficulty breathing. It is most often caused by bacteria but may also be caused by viruses or fungus. The severity of pneumonia ranges from mild to life-threatening. It is most serious in infants, children, the Elderly, and people with low immune systems. A chest X-ray is often used to diagnose pneumonia and look for lung inflammation.

2.2.2 COVID-19

Coronavirus disease 2019 (COVID-19) is one of the major lung infection diseases that almost everyone knows about. It is a highly contagious respiratory disease that originated from the SARS-CoV-2 virus. Since the beginning of the pandemic, the virus that causes COVID-19 has constantly been changing. We have seen some variants such as Alpha, Beta, Delta, and Omicron. It will spread from person to person through coughs, talks, or sneezes. The symptoms caused by COVID-19 include fever, sore throat, shortness of breath, etc. COVID-19 can cause pulmonary complications, such as pneumonia. In most severe cases, it can lead to inflammation and fibrosis of the lung. These COVID-19 survivors often have persistent abnormalities on lung imaging for six months to a year after infection.

By January 2023, the confirmed cases of COVID-19 were 657,977,736, including 6,681,433 deaths reported to WHO. Therefore, early and accurate diagnosis of COVID-19 is important to control the spread of the disease and reduce its mortality rate. A commonly used method to detect COVID-19 is reverse transcription polymerase chain reaction (RT-PCR). Besides, The use of X-ray images is also a common way to detect how the virus affects the lungs [9].

2.3 Literature Review

The field of medical image analysis encompasses the utilization of computer algorithms to process and evaluate medical images. This can include techniques such as image segmentation, registration, and visualization. Common applications include diagnostic imaging, treatment planning, and image-guided surgery. A significant area of research in medical image analysis is the advancement of deep learning algorithms for image segmentation and classification. These methods, built on convolutional neural networks, have demonstrated exceptional performance in various medical imaging tasks. Another area of research is the employment of medical images for diagnosis and disease detection. This includes using machine learning algorithms to categorize images based on features such as texture, shape, and intensity and natural language processing techniques to extract information from radiologic reports. Furthermore, research in medical image registration, which aligns and combines multiple images of the same body structure or organ, is widely used in image-guided surgery, radiotherapy and other clinical applications.

Medical image analysis can be used in many areas of healthcare, such as neurology, cardiology, oncology and orthopedics. Medical imaging technology has been helpful in the early detection, diagnosis, and treatment of diseases. Medical imaging is used not only for diagnostic purposes but also for guiding. Medical images guide physicians in procedures to improve precision and accuracy. In clinical practice, medical images interpretation has mostly done by human experts, such as physicians and radiologists. However, due to the impact of COVID-19 and the potential fatigue of physicians, which results in conventional artificial methods challenging to satisfy the growing social needs. In addition, physicians may misinterpret diseases because of inexperience, leading to missed diagnosis. Thus, physicians and researchers have recently benefited from computer-aided diagnostic approaches. Although the progress of computational medical image analysis is belated compared to medical imaging technologies, it has been significantly improved with the application of machine learning techniques.

The common method to analyze and classify of X-ray images relies on radiologists. Healthcare professionals generate and collect large numbers of X-ray images and deliver them to radiologists for analysis. This conventional artificial method requires substantial effort from the radiologist, which is highly subjective and misinterpreted diseases. Another method is to use automated technology for identification to help the diagnosis process. As soon as scanning medical images and loading them into a computer was possible, the researchers had already set up a system for automatic analysis.

Since the 1960s, papers describing automatic abnormality detection sys-

tems for chest X-ray images have been published. Lodwick et al. developed a coding system to transfer radiographs of lung cancer to an electronic data processing system. This made it possible to explore the use of computers as a helpful tool for radiological diagnosis [10]. From the 1970s to the 1990s, medical image analysis was done by converting visual images on radiographs into digital sequences and mathematical modeling to construct rule-based image processing systems that can then be manipulated and evaluated by digital computers [11]. Kruger et al. described a computer algorithm that aims to automatically detect feature extraction and normal-abnormal diagnosis for chest radiographs [12]. After the 1990s, supervised techniques became popular for image analysis. The approach uses a set of graphs that fit the data to constitute the training data, feature extraction, and statistical classifiers to develop the system [13].

In recent years, machine learning has been designed and used to analyze medical datasets. Deep learning has become the technique of choice for image analysis and greatly impacts medical imaging. Especially in recent years, the development of computers has provided inexpensive and variety of means to store and share data. It forms the basis of successful commercial medical image analysis systems. Model-based image analysis and computer-aided diagnosis systems have played a role in helping the doctor by providing accurate and rapid results. To automate disease detection, the classification of X-ray images should be automated as an initial step. When we look at a picture, we can easily describe and recognize what is shown. However, this is not an easy task for the computer. Image classification is the task of classifying and assigning labels to images. In simple terms, we want a computer to analyze an image and determine what "category" the image belongs to, such as food, animals, buildings, etc. In the early stage of image classification, the computer decomposes the image into individual pixels or vectors for classification. Two pictures of the same thing may look very different on different backgrounds, and the computing power was quite limited at the time. This makes it a challenge to get high correct rates on classified images. Thus, deep learning is a key technique for image classification. It can be an effective tool for X-ray image classification.

Figure 2.1 shows a taxonomy of traditional machine learning-based medical image analysis algorithms. Traditional machine-learning techniques have been widely used in medical image analysis. Some common traditional machine learning approaches used in the literature include:

- Support Vector Machines (SVMs) have been used to classify breast masses in mammography images as benign or malignant.
- Random Forest (RF) have been used for image classification and seg-

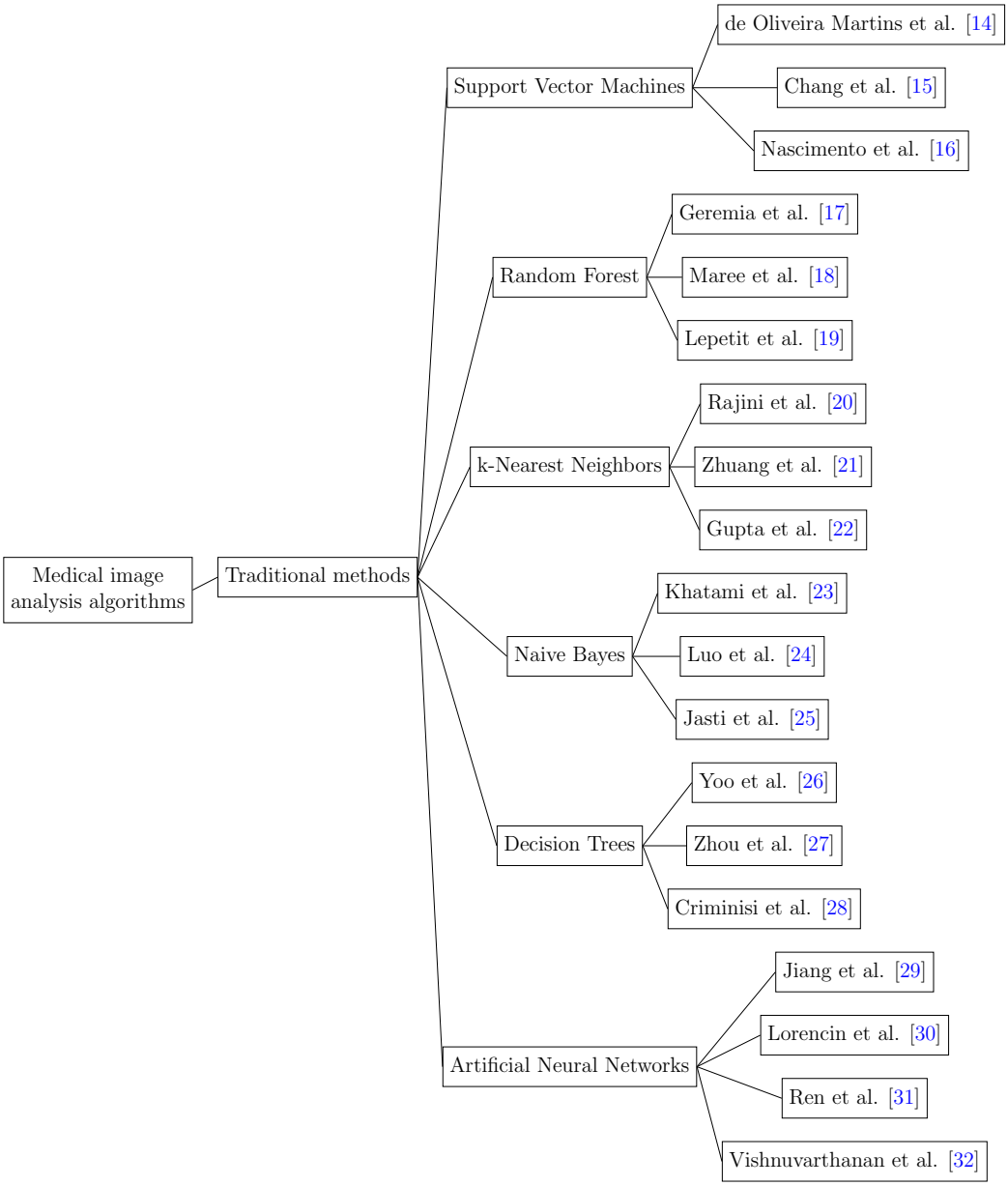


Figure 2.1: A taxonomy of traditional machine learning-based medical image analysis algorithms

mentation, such as classifying brain tumors in magnetic resonance imaging (MRI) images and segmenting the left ventricle in cardiac magnetic resonance images.

- k-Nearest Neighbors (k-NN) have been used for image retrievals, such

as retrieving similar X-ray images from a large database for comparison and diagnosis.

- Naive Bayes (NB) have been used for feature selection, such as selecting the most relevant imaging features for diagnosing a specific disease.
- Decision Trees (DTs) have been used for feature extraction, such as extracting features from ultrasound images to diagnose liver diseases.
- Artificial Neural Networks (ANNs) have been used for image registration, such as aligning pre-operative and post-operative images for image-guided surgery.

In addition, unsupervised learning algorithms such as clustering [33–35] are also commonly applied in medical analysis.

Deep learning, specifically Convolutional Neural Networks (CNNs), has been widely used in medical image analysis in recent years. Figure 2.2 shows a taxonomy of deep learning-based medical image analysis algorithms. Some examples of deep learning approaches used in the literature include:

- U-Net is a popular CNN architecture for image segmentation, such as segmenting brain tumors in MRI images and lung nodules in CT images.
- ResNet is a CNN architecture for image classification, such as classifying skin lesions as malignant or benign.
- DenseNet is a CNN architecture for image classification, such as classifying chest X-rays as normal or abnormal.
- Mask R-CNN is a CNN architecture used for object detection and segmentation, such as detecting and segmenting tumors in CT images.
- Generative Adversarial Networks (GANs) have been used for image synthesis and enhancement, such as generating synthetic CT images from MRI images and enhancing low-dose CT images.
- Attention-based CNNs have been used for image captioning, such as generating captions for radiology images to help radiologists in their diagnostic tasks.
- 3D CNNs have been used for volumetric image segmentation and analysis, such as segmenting the heart and vessels in cardiac MRI images.

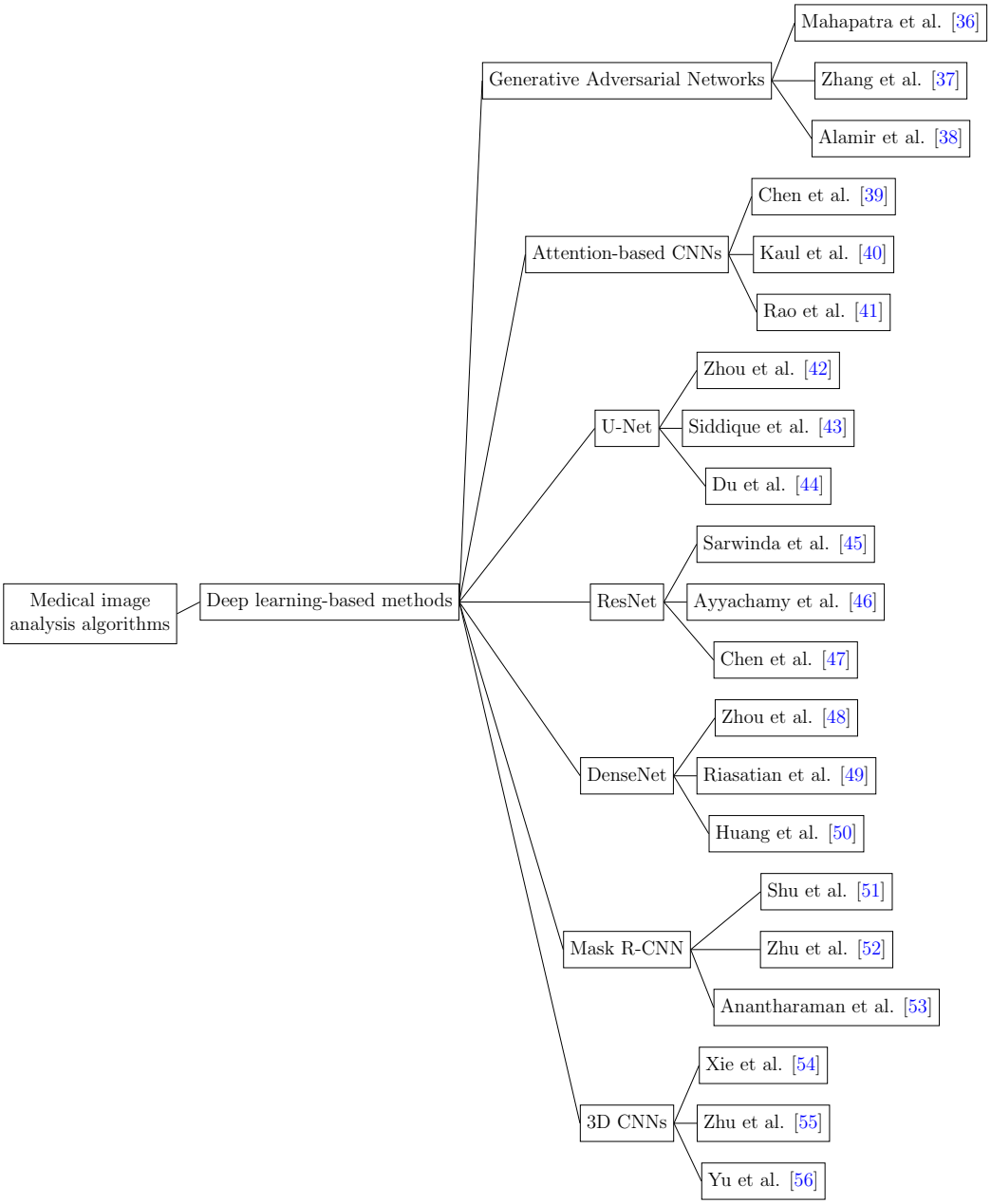


Figure 2.2: A taxonomy of deep learning-based medical image analysis algorithms

Deep learning is a machine learning technique that allows computers to learn from data. Most deep learning methods use computer systems known as neural networks. This concept has gained the interest of researchers for its good performance and has become the best solution for many problems

in medical image classification. Basha. et al. proposed a system using neural networks, back propagation neural networks and support vector machine classifiers to classify the X-ray images. The overall accuracy of the X-ray image classification system was 92.3% [57]. Ozturk et al. proposed a new model for automatic COVID-19 detection from X-rays images using deep learning methods [58]. Hemdan et al. introduce a new deep-learning framework called COVIDX-Net which can assist radiologists in diagnosing COVID-19 [59]. The authors also evaluate the performance of deep neural network models proposed over the recent years for medical image classification [60]. Z. Karhan. et al. use the ResNet50 model to classify Covid-19 and non-Covid-19 Chest X-ray images with 99.5% accuracy [61]. Deep neural networks are widely used for classification and have been successful in the medical field.

Chapter 3

Methodology

3.1 Computer Image

An image is a visual representation of objects, while a digital image consists of an array of elements called pixels. It can be stored and processed by a computer. Each of these pixels can be represented by a single number or a small set of numbers that describe some property of this pixel such as the colors or brightness. The numbers are arranged in an array of rows and columns corresponding to the vertical and horizontal positions of the pixels in the image. A grayscale image can be represented as a 2-dimensional matrix of numbers. A color image normally has three colors. The RGB image is a color image made of red, green and blue, so the pixels of a color image can be made into a three-dimensional vector. There are also non-optical images, such as ultrasound or x-ray, in which the intensity of the sound or X-ray is recorded. One of the advantages of digital images over traditional images is the ability to transmit them electronically almost instantaneously and to change easily from one medium to another, such as from a web page to a computer screen to a printer. Another even greater advantage is the ability to change them to suit your needs. Digital images often produce large files and are often compressed to make the files smaller. Compression takes advantage of the fact that many nearby pixels in an image have similar colors or brightness [62].

Image processing describes digitally transforming an image and performing specific operations to obtain useful information. In applying some predetermined signal processing methods, image processing systems typically treat images as two-dimensional signals. There are many types of image processing. Pattern recognition measures various patterns around objects in an image. Recognition means detecting or distinguishing objects in an image.

Retrieval is browsing or searching an extensive image database for images that resemble the original image. image enhancement is used to create an enhanced image from the original image. Visualization aims to identify objects that are not visible in the image.

3.2 Neural Networks

Neural networks are networks or circuits composed of artificial neurons or nodes. Their names and structures are inspired by the human brain and mimic how biological neurons signal to each other. Human brain cells are also known as neurons. They form a complex, highly interconnected network and transmit electrical signals to each other to help humans solve problems [63]. Similarly, artificial neural networks are composed of artificial neurons that work together to process information. Artificial Neural Networks (ANNs) are a type of Neural Network. ANNs are the subset of machine learning and are at the core of deep learning.

3.2.1 Neuron

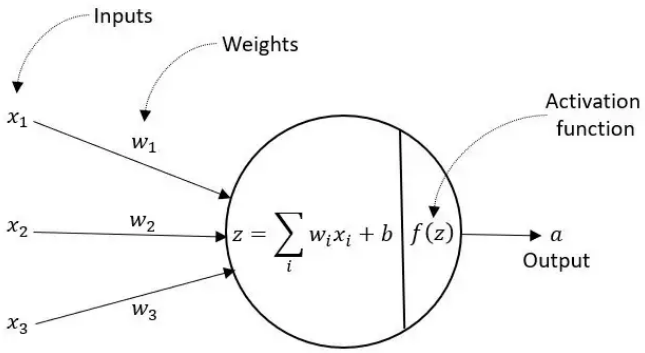


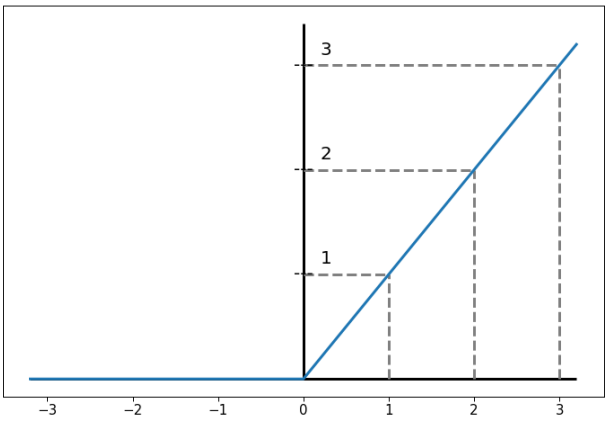
Figure 3.1: Neuron[1]

Artificial neural networks consist of many neurons, which are also called nodes. It receives input data from other neurons or external environments and calculates the output. As shown in Figure 1 below, each neuron contains three parts: weight (w), bias (b), and activation function (f). Each input into a neuron is associated with weights and biases. All initial weights are

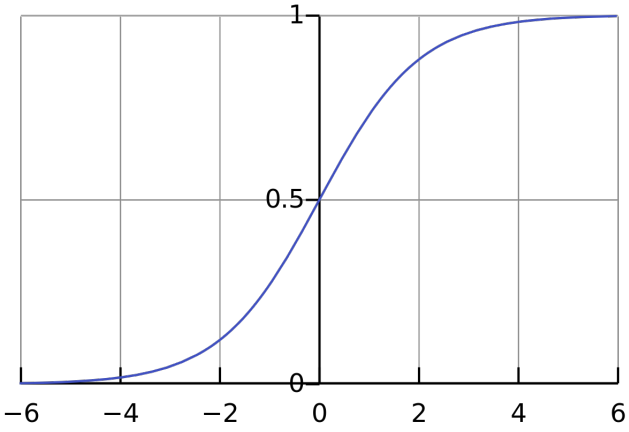
[1] https://levelup.gitconnected.com/a-review-of-the-math-used-in-training-a-neural-network-9b9d5838f272

random, while biases are 0. Figure 3.1 demonstrates a neuron computes a weighted sum with biases of the inputs first, and then an activation function is applied to determine the final output.

The activation functions play a role in transforming linear input values to non-linear ones to enable the modeling of complex tasks. Figure 3.2 shows two activation functions used in our experiment: ReLU and softmax. The ReLU activation function in a CNN rectifies negative values to zero while the softmax activation function outputs a probability distribution over classes in multiclass classification.



(a) ReLU[2]



(b) softmax[3]

Figure 3.2: Activation function

[2]https://tungmphung.com/rectifier-linear-unit-relu/
[3]https://medium.com/@AbhiramiVS/softmax-vs-logsoftmax-eb94254445a2

3.2.2 Layers

ANNs consist of an input layer, one or more hidden layers, and an output layer, as shown in Figure 3.3. Each artificial neuron is connected to another neuron and has an associated weight and threshold.

Input Layer There is no computation performed in the input layer. The input dataset enters from the input layer into the artificial neural network. Neurons in the input layer process and analyze the dataset and feed it into the hidden layer.

Hidden Layer Artificial neural networks can have more than one hidden layer. Each hidden layer analyzes and calculates the output from the previous layer (input layer or hidden layer). The output from the hidden layer will enter the output layer.

Output Layer The output layer takes the input from the hidden layer and performs the computation, giving the final result of all data processing in the artificial neural network. The output layer can have single or multiple neurons based on whether it is a binary classification problem or multiple class classification problem. If there is a binary classification problem (yes/no), the output layer will have an output neuron that will give a result of 1 or 0. However, if there is a multi-class classification problem, the output layer may include more than one output neuron.

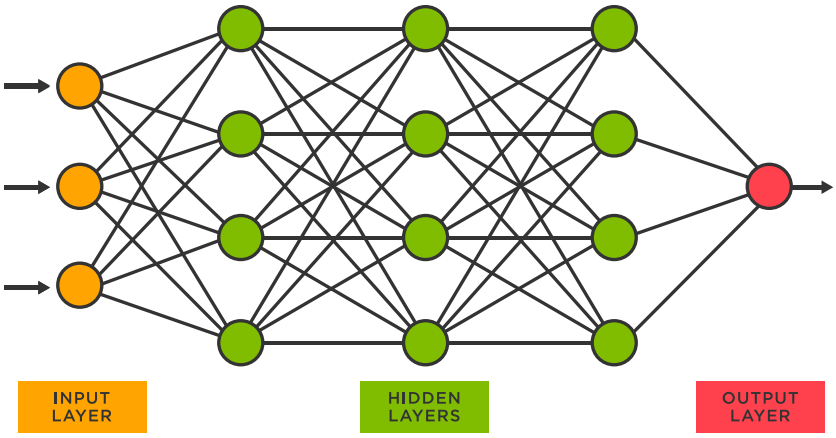


Figure 3.3: Three types of layers[4]

[4] https://www.tibco.com/reference-center/what-is-a-neural-network

3.3 Convolutional Neural Networks

CNNs are a class of ANNs commonly used in recommender systems, image and video recognition, medical image analysis and classification with superior performance [64]. CNNs can extract relevant features from images that are useful for identifying objects in images [65]. This new approach is easier to handle without losing the features that are essential to making good predictions. There is a standard structure of the CNN model that consists of alternating convolutional and pooling layers. The last layer of the CNN model is a fully connected layer as shown in Figure 3.4. In this section, we will briefly review how CNNs work.

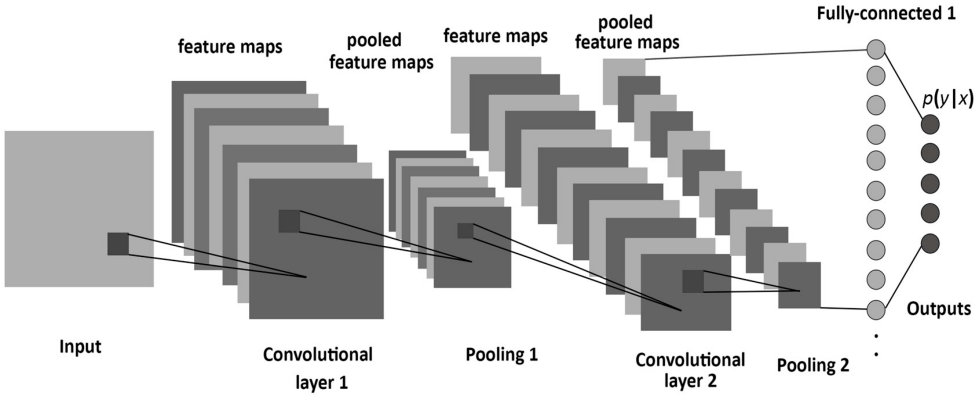


Figure 3.4: CNN structure[1]

3.3.1 Convolutional layer

The convolutional layer is the first layer and the core building block of CNNs. This layer of the process involves input images, kernels, and feature maps [66]. Let's assume that the input will be a black-and-white image made up of a matrix of pixels between 0-255. As shown in Figure 3.5, the small square in green represents a pixel. The number 0 means black color and the number 1 means white. The square in orange is often called a *feature detector*, *kernel*, or *filter*. It will move through the receptive domain of the image and check for the presence of features. This process is called convolution. Although the size of the kernel can vary, the kernel size is usually a 3*3 matrix. The kernel is applied to a region of the image and the dot product between the input pixels, and the kernel is calculated. This dot product is an element value input to a new matrix. The formula used for the calculation is shown in Equation 3.1.

$$J(x,y) = K * I = \sum_{n,m} K(n,m)I(x-n, y-m) \tag{3.1}$$

where J is the feature map, K is the kernel, I is the input images, and n,m is the kernel index. From the equation, we know that the sum of the products of the individual elements of I and the individual elements of K is calculated to obtain each element of J. Each element of J is calculated by multiplying the sum of the elements of I and K.

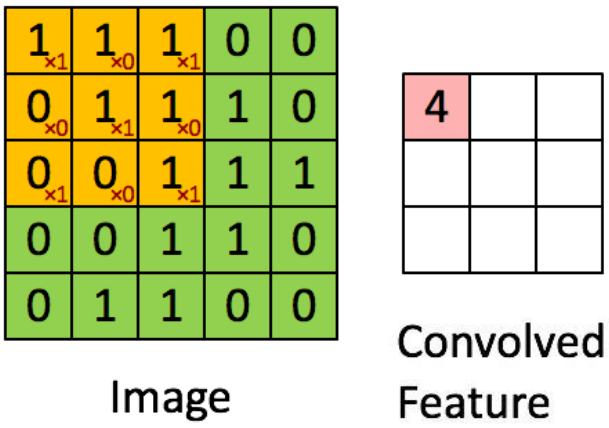


Figure 3.5: Convolutional layer[5]

The kernel is then shifted by a span and repeated until the entire image is swept. The final output of all dot products is called the convolved feature. The feature map contains feature values and relative position information. For example, in face detection from an image, the eyes, nose, and mouth are arranged from top to bottom, so the corresponding feature values are extracted in this order. After each convolution operation, the CNN performs a ReLU transformation of the convolved feature, introducing nonlinearity to the model. Convolutional layers are not only applicable to the input image but also follow the output of other convolutional layers. The stacking of convolutional layers allows for a hierarchical decomposition of the input image. In this case, filters that operate on the first level of convolved features may extract combinations of low-level features, such as features consisting of multiple lines to represent object shapes.

[5] https://glassboxmedicine.com/2020/08/03/convolutional-neural-networks-cnns-in-5-minutes/

3.3.2 Pooling Layer

To reduce the computational burden and avoid overfitting, the pooling layers are used to reduce the spatial size of the feature map. The method uses the value of one neuron to represent a region until all the neurons are represented, which implements the compression of the feature map as shown in Figure 3.6, where the filter size equals 2x2. The pooling operation sweeps a filter over the entire feature map, but the difference is that this filter does not have any weights compared to the kernel. There are two common types of pooling: Max pooling and Average pooling. In this study, we only use Max pooling.

Max pooling: As the filter moves over the feature map, it returns the maximum value from each pooling region and sends it to the output array. It completely discards noisy activations and performs denoising while dimensionality reduction.

Assuming that the pooling region is R, the max pooling P_m is expressed as:

$$X = [x_i | i \in R] \tag{3.2}$$

$$P_m = max(X_R) \tag{3.3}$$

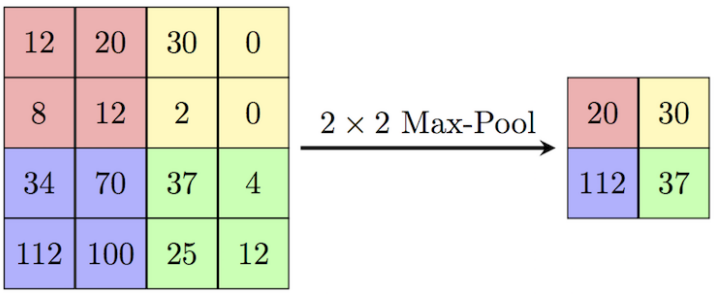


Figure 3.6: Max pooling layer[6]

Although much information is lost in the pooling layer, it also has many benefits for CNNs. It helps reduce complexity and improve efficiency.

3.3.3 Fully connected layers

The fully Connected Layer is the last in the convolutional neural network. In the partially connected layer mentioned above, the pixel values of the input image are not directly connected to the output layer. However, in the fully

[6] https://computersciencewiki.org/index.php/Max-pooling_/_Pooling

connected layer, each neuron in the output layer connects to every neuron in the previous layer, as shown in Figure 3.7. We can add multiple layers as we want, depending on the depth of the classification model [67].

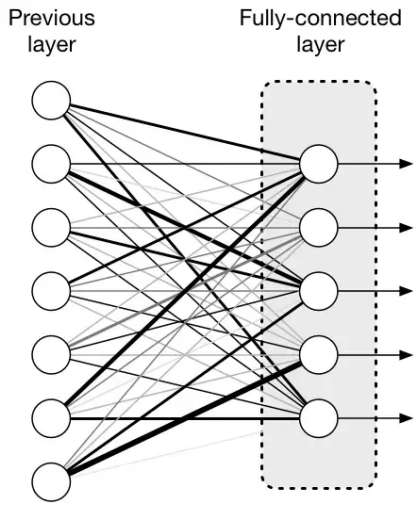


Figure 3.7: Fully connected layers[7]

Compared to other image classification algorithms, CNNs use relatively less preprocessing. This means that the neural network optimizes the kernels by automatic learning. However, these filters are designed manually in traditional algorithms. This feature extraction independent of a priori knowledge and human intervention is a major advantage.

[7] https://medium.com/@tecokids.monastir/fully-connected-layer-with-dynamic-input-shape-70c869ae71af

Chapter 4

Experimentation

In the previous sections, we have introduced medical imaging and the architecture of CNN. This chapter will use chest X-rays images as a representative and report the utility of CNN on a chest X-ray dataset named "COVID-QU-Ex", a dataset that includes normal, COVID-19, and other lung infection images. We first introduce the COVID-QU-Ex baseline dataset for showing the accuracy of the CNN-based X-ray classification model in Section 4.1. Next, we introduce the model we used in this reproduce experiment and list details of our configurations and experimental settings in Section 4.2 and 4.3. We also explain our evaluation metrics in Section 4.4. The classification accuracy and a 5-fold cross-validation result are demonstrated in Section 4.5. Lastly, we visualize the predictions using saliency maps to explore our comprehension of the implementation of CNN.

4.1 Baseline dataset

In our experiment, we adopt the **COVID-QU-Ex**[1] dataset consisting of 33,920 chest X-ray images up to February 2022, compiled by researchers of Qatar University. This dataset contains the following:

- 10,701 **Normal** cases
- 11,956 **COVID-19** infections
- 11,263 **Non-COVID** infections (viral or bacterial pneumonia)

The detail of this dataset is shown in Table 4.1:

[1] https://www.kaggle.com/datasets/anasmohammedtahir/covidqu

Table 4.1: Distribution of the experimental datasets

#Datatset	Normal	COVID-19	Non-COVID
Train	6,849	7,658	7,208
Val	1,712	1,903	1,802
Test	2,140	2,395	2,253
Total	10,701	11,956	11,263

Besides, corresponding ground-truth lung segmentation masks are provided for the entire dataset so medical doctors can better detect, localize, and quantify COVID-19 infection from X-ray images.

4.2 Model Architecture

We use the model proposed by Ahmed et al. [68] as the COIVD classification model. The model structure is shown in Figure 4.1. It contains five convolutional layers, followed by batch normalizations, max-pooling layers, and dropouts. All chest X-ray scans are resized to (200 * 200) before inputting into the model. We process the Dense layer with 512 neurons as a fully connected layer. Following it is the last prediction, with three neurons representing each class of chest X-rays. We process ReLU in each layer as the activation function and softmax in the final dense layer.

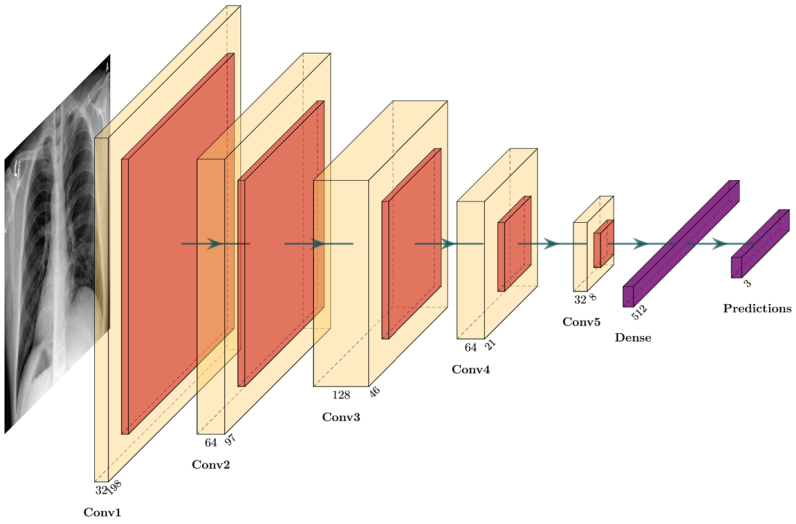


Figure 4.1: CovID Model Archtutre.

Additionally, Table 4.2 lists the details of each layer, including layer types, output shapes, and the parameter amount of each layer.

Table 4.2: Description of Each Layer in the Model

Layer (type)	**Output Shape**	**Param #**
conv2d_1(Conv2D)	(None, 198, 198, 32)	896
batch_normalization_1(BatchNormalization)	(None, 198, 198, 32)	128
max_pooling2d_1(MaxPooling2D)	(None, 99, 99, 32)	0
dropout_1(Dropout)	(None, 99, 99, 32)	0
conv2d_2(Conv2D)	(None, 97, 97, 64)	18496
batch_normalization_2(BatchNormalization)	(None, 97, 97, 64)	256
max_pooling2d_2(MaxPooling2D)	(None, 48, 48, 64)	0
dropout_2(Dropout)	(None, 48, 48, 64)	0
conv2d_3(Conv2D)	(None, 46, 46, 128)	73856
batch_normalization_3(BatchNormalization)	(None, 46, 46, 128)	512
max_pooling2d_3(MaxPooling2D)	(None, 23, 23, 128)	0
dropout_3(Dropout)	(None, 23, 23, 128)	0
conv2d_4(Conv2D)	(None, 21, 21, 64)	73792
batch_normalization_4(BatchNormalization)	(None, 21, 21, 64)	256
max_pooling2d_4(MaxPooling2D)	(None, 10, 10, 64)	0
dropout_4(Dropout)	(None, 10, 10, 64)	0
conv2d_5(Conv2D)	(None, 8, 8, 32)	18464
batch_normalization_5(BatchNormalization)	(None, 8, 8, 32)	128
max_pooling2d_5(MaxPooling2D)	(None, 4, 4, 32)	0
dropout_5(Dropout)	(None, 4, 4, 32)	0
flatten(Flatten)	(None, 512)	0
dense(Dense)	(None, 512)	262656
batch_normalization_6(BatchNormalization)	(None, 512)	2048
dropout_5(Dropout)	(None, 512)	0
predictions(Dense)	(None, 3)	1539

4.3 Experiment Configuration

The original model was coded under TensorFlow and Kears. Thus, our implementation was under the same platform. However, due to the change of some source packages, we reprogrammed some parts to be compatible with the latest version of Python and Kears. We list the version of the Python environment and some key packages we used in this experiment:

(1) python == 3.10.8; (2) keras == 2.10.0; (3) tf_keras_vis == 0.8.4; (4) numpy == 1.23.5; (5) pandas == 1.5.2; (6) scikit-learn == 1.2.0

Simultaneously, considering the chest X-ray images amounts have increased from 1,389 to 33,920, we adjust the hyperparameters with the following settings:

- image size: (200, 200)
- epoch: 25
- batch size: 64
- learning rate: 0.001
- dropout rates: 0.25 (convolutional layers) / 0.50 (dense layer)
- optimizer: RMSProp[2]
- loss: categorical cross-entropy[3]

The whole experiment was conducted on JAIST High-Performance Computer (HPC) named Kagayaki. We set the queue class as GPU-S, which contains 2 NVIDIA A40 GPUs for every node. All our programs ran under the container environment called jupyter/tensorflow-notebook[4] from Docker. Because of the scattered sources, we also renamed and sorted all images of the COVID-QU-Ex dataset from 1 to end. Thanks to the powerful Kagayaki with its GPU nodes, the training and cross-validation stages only cost approximately 2 and 8 hours per round.

4.4 Evaluation Metrics

To evaluate our sentiment classification performance, we used a three classes confusion matrix commonly used to describe the data classification result as shown in Table 4.3. Generally, for data prediction results, there are 4 cases, including True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). True (T) and False (F) denote the predicted result is right and wrong, respectively. In contrast, Positive (P) and Negative (N) denote the positive and negative cases of the predicted samples, respectively.

In our model, we denoted three classes $\mathbf{T} = \{Norm, Cov, Non\}$ for Normal, COVID-19, and Non-COVID. The formulations of TP, FP, TN, and FN under each label are defined as in Equation 4.1, 4.2, and 4.3, respectively.

- For label is $Norm$:

$$\begin{aligned} TP_{Norm} &= a & TN_{Norm} &= e + f + h + i \\ FP_{Norm} &= d + g & FN_{Norm} &= b + c \end{aligned} \tag{4.1}$$

[2] https://keras.io/api/optimizers/rmsprop/

[3] https://www.tensorflow.org/api_docs/python/tf/keras/losses/CategoricalCrossentropy

[4] https://hub.docker.com/r/jupyter/tensorflow-notebook

Table 4.3: Confusion matrix for three sentiment labels.

Confusion Matrix		Actual		
		Norm	Cov	Non
Predicted	Norm	a	d	g
	Cov	b	e	h
	Non	c	f	i

- For label is Cov:

$$\begin{aligned} TP_{Cov} &= e & TN_{Cov} &= a+c+g+i \\ FP_{Cov} &= b+h & FN_{Cov} &= d+f \end{aligned} \tag{4.2}$$

- For label is Non:

$$\begin{aligned} TP_{Non} &= i & TN_{Non} &= a+b+d+e \\ FP_{Non} &= c+f & FN_{Non} &= g+h \end{aligned} \tag{4.3}$$

Macro Performance To obtain $Precision$, $Recall$, and F_1 score under macro criteria, we first calculated the precision and recall for each label $t \in T$ as follows:

$$Precision_t = \frac{TP_t}{TP_t + FP_t} \qquad Recall_t = \frac{TP_t}{TP_t + FN_t} \tag{4.4}$$

And the F1 score under label t is defined as the following Equation:

$$F1_t = \frac{2 * Precision_t * Recall_t}{Precision_t + Recall_t} \tag{4.5}$$

Thus, the $Macro\ F_1$ is defined as:

$$Macro\ F1 = \frac{\sum_{t \in T} F1_t}{3} \tag{4.6}$$

Accuracy We use a general evaluating indicator, accuracy, to describe how the model performs across all classes. It refers to the ratio of correctly predicted samples to the total number of predicted samples without considering whether they are positive or negative cases. The definition of an average accuracy across all classes T is described in Equation 4.7 and 4.8:

$$Accuracy_t = \frac{TP_t + TN_t}{TP_t + TN_t + FP_t + FN_t} \tag{4.7}$$

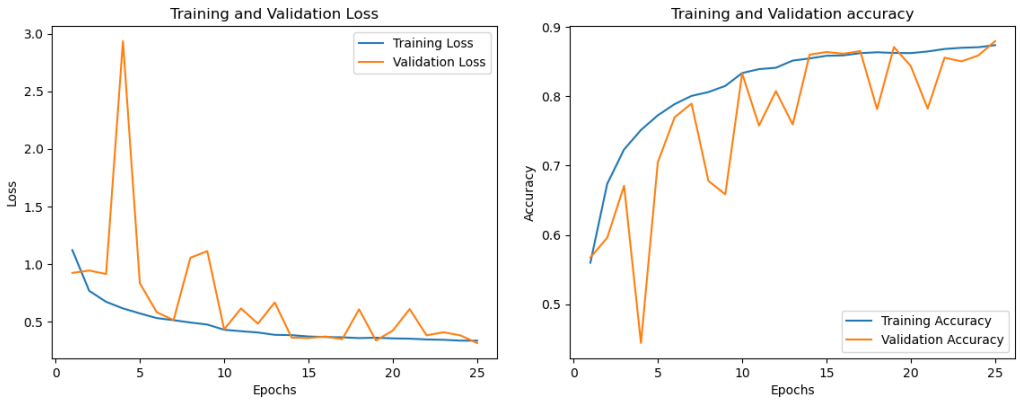
$$Accuracy = \frac{\sum_{t \in T} Accuracy_t}{3} \tag{4.8}$$

4.5 Experiment Results

In this section, we present our experimental results from different perspectives. The changing of loss and accuracy during training are shown in Section 4.5.1. After training, we implement the 5-Fold cross-validation result and summarize them in Section 4.5.2. Finally, to better understand the result and the mechanism of how CNN works, we prepare visualized results in the last Section 4.5.3.

4.5.1 Initial Model Performance

For the initial step, Figure 4.2a and 4.2b shows the reproduced model successfully converged and reached high accuracy on the new dataset based on our settings.



(a) loss (b) accuracy

Besides, Table 4.4 has proven the model still achieved an accuracy of 88.60% on the testing set, with a Macro F1-score of 88.52%.

Table 4.4: Initial model performance (%)

	Precision	Recall	F1
Normal	80.60	91.45	85.68
COVID-19	95.42	91.44	93.39
Non-COVID	90.41	82.87	86.48
Macro avg	88.81	88.59	**88.52**

Figure 4.3 shows the confusion matrix of the initial model performance, which indicates that the model on the large-scale dataset can also distinguish

different infections (COVID-19 and viral & bacterial pneumonia) with high accuracy for each category of chest X-ray.

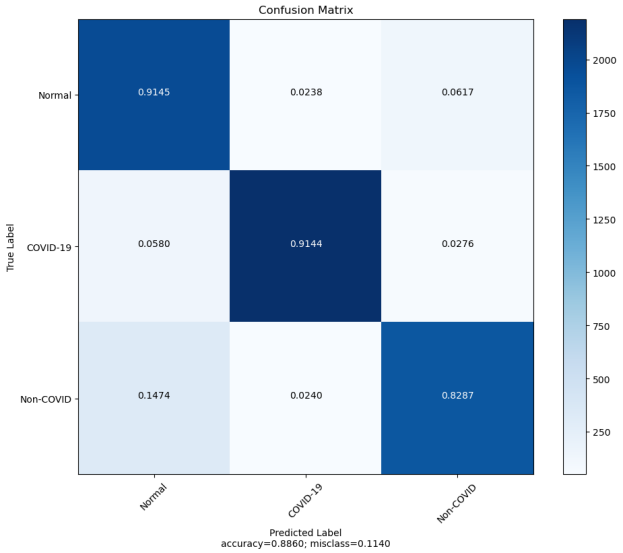


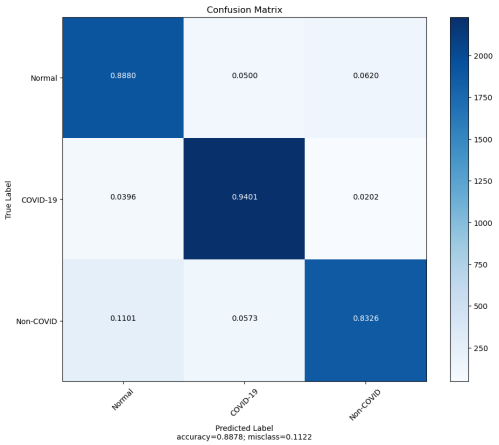
Figure 4.3: Conftusion matrix

4.5.2 5-fold cross-validation

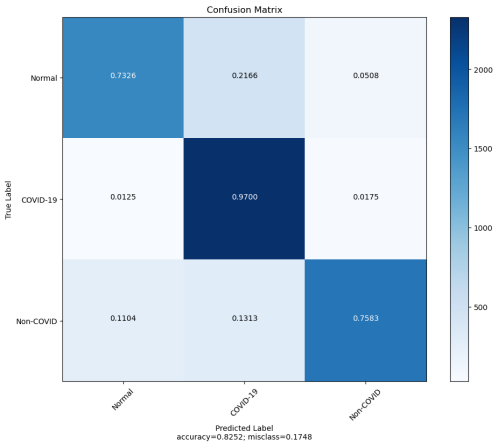
In addition, we implemented an extra 5-fold cross-validation in our model to better measure the performance. Table 4.5 reflects a similar result to the initial F1-score and accuracy, with an average F1-score of 86.06% and an accuracy of 86.30%. Besides, we plot the accuracy of each class for each fold into Figure 4.4 to monitor the accuracy changing between each fold. The average accuracy indicates the model can successfully detect lung infections, especially when identifying COVID-19.

Table 4.5: 5-fold cross-validation performance (%)

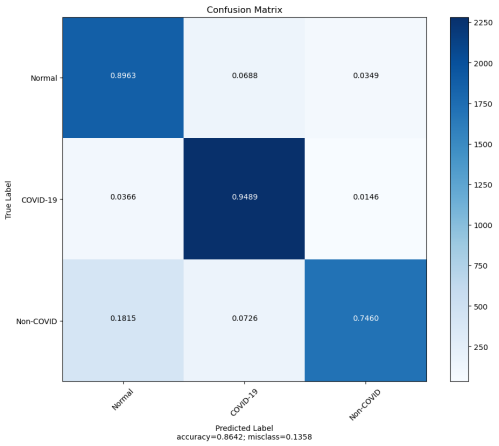
Fold	**Precision**	**Recall**	**Macro F1**	**Accuracy**
Fold-1	88.80	88.69	88.66	88.78
Fold-2	84.08	82.03	82.21	82.52
Fold-3	86.98	86.37	86.14	86.42
Fold-4	86.14	85.30	85.25	85.47
Fold-5	88.39	88.08	88.04	88.30
Average	86.88	86.09	86.06	86.30



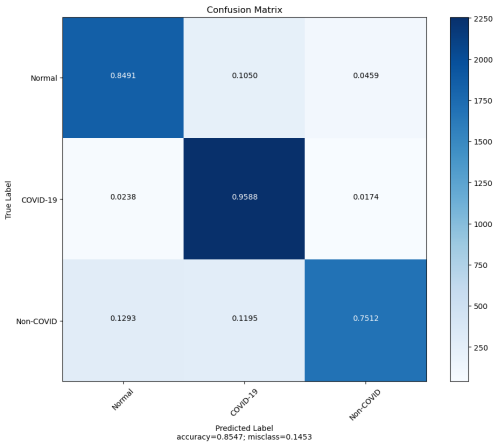
(a) Fold 1



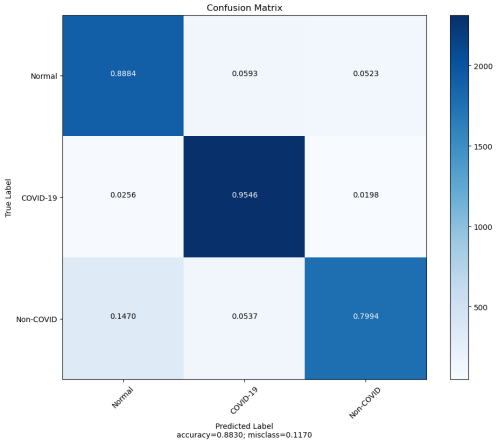
(b) Fold 2



(c) Fold 3



(d) Fold 4

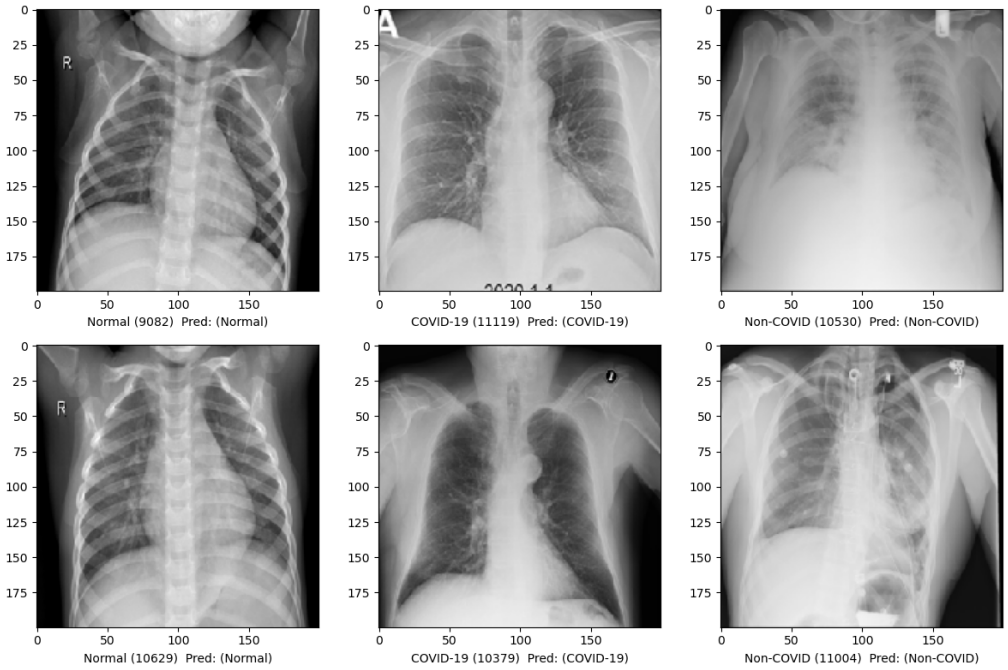


(e) Fold 5

Figure 4.4: 5-flod confusion matrices.

4.5.3 Result Visualzation

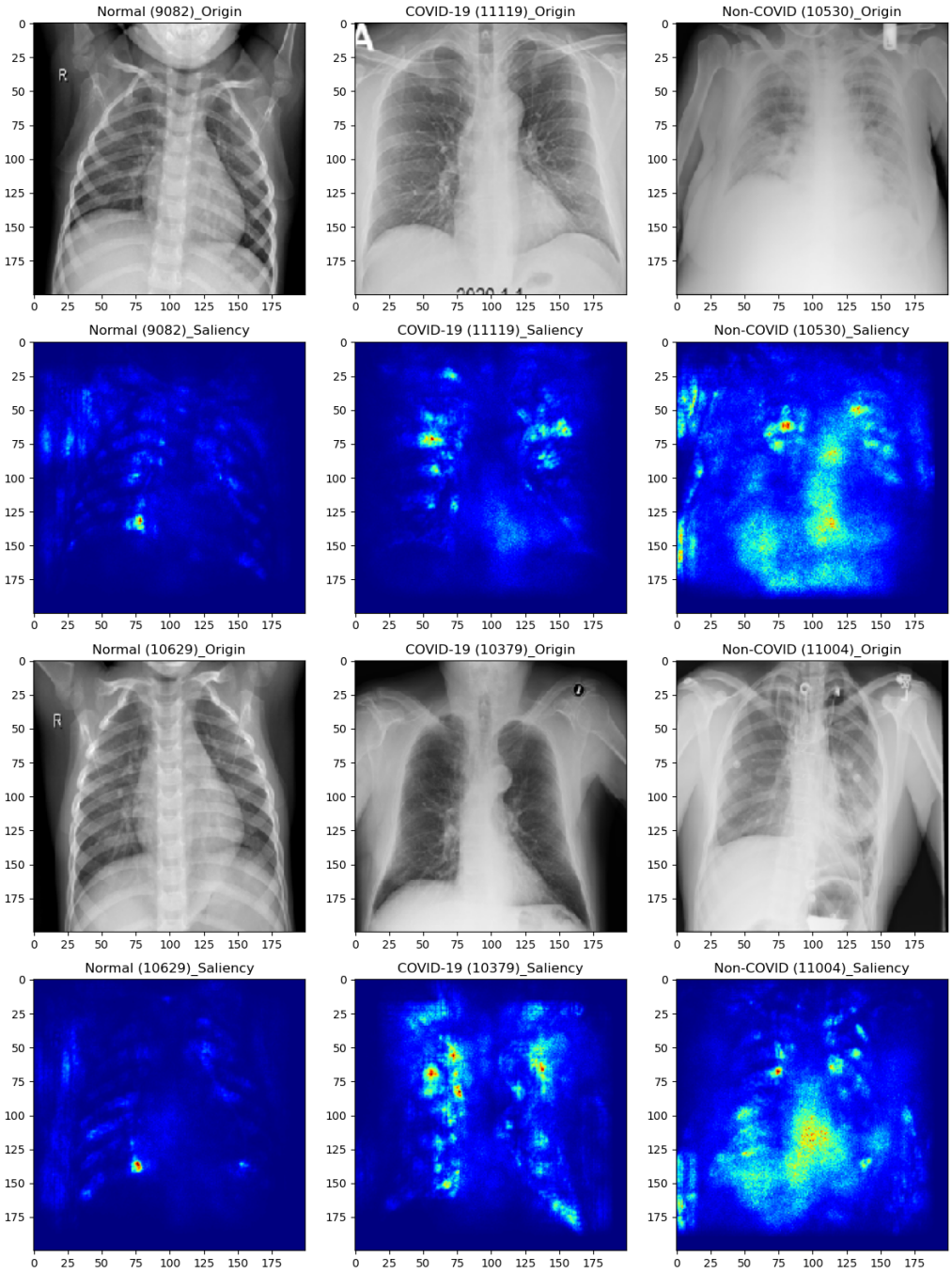
In order to better understand the chest X-ray classification results, we visualize the outputs of our reproduced model. Two chest X-ray images were randomly selected from the test set under each category, and the corresponding prediction results near image names are demonstrated in Figure 4.5. These predictions will help radiologists quickly obtain the initial conclusions about patients' situations.



(a) Normal (b) COVID-19 (c) Non-COVID

Figure 4.5: Chest X-ray images with corresponding predictions.

Despite the above results that can satisfy radiologists' primary demands, it is essential to figure out how the model makes these predictions. The saliency map [69], a visualization technique used to identify the areas in an image that are most important for a given classification task, can help us better understand the feature of CNN and tell us what area impacts the predictions of the model. We believe the highlighted areas may provide insight into the most important features of chest X-ray images. Following the previous step, Figure 4.6 shows the visualized chest X-ray images using saliency maps, summarizing the original images with corresponding saliency maps for comparison. Apparently, the location, area, and color lightness under each category are significantly different.



(a) Normal (b) COVID-19 (c) Non-COVID

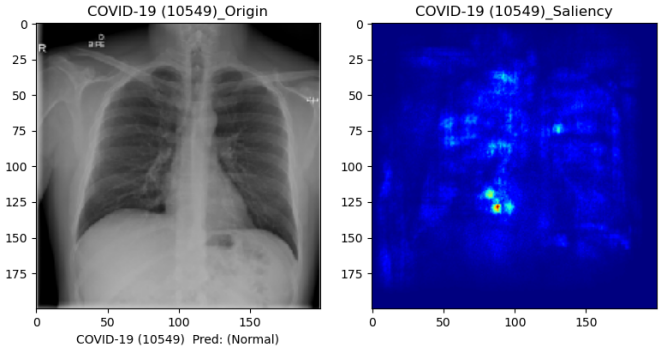
Figure 4.6: Saliency maps highlighted areas for chest X-ray images.

We infer the highlighted areas in Figure 4.6 have the greatest impact on the prediction results of the model. From analyzing all saliency maps, it is evident that:

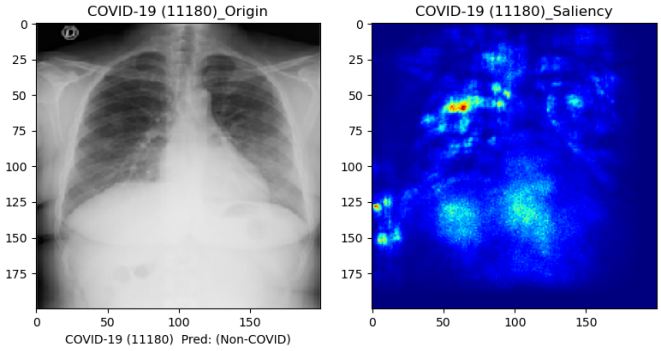
Normal
: few and insignificant highlighted areas are found for Normal cases.

COVID-19
: for COVID-infected patients, the highlighted areas are scattered over most of the patient's lungs.

Non-COVID
: for non-COVID-19 infected patients (bacterial or viral pneumonia), the highlighted areas are concentrated in the lungs and the cardiophrenic angle area.



(a) COVID-19 misclassed as Normal.



(b) COVID-19 misclassed as Non-COVID.

Figure 4.7: Misclassed COVID-19 X-ray images.

We also observe some misclassed cases in this experiment as shown in Figure 4.7. In these misclassed examples, the model still utilizes the same

logic and pattern for image classification in chest X-ray images. These examples of classification errors can support the model's interpretability, but the classification results still need to be discussed with the physician.

Chapter 5

Conclusion

5.1 Summary

The image classification problem is a classical and highly practical value research topic in the field of computer vision. Its applications are also very wide, from autonomous driving to face recognition. It is no exaggeration to say that the application of deep learning on image classification has attracted the public's attention, thus leading us to an era of "AI everywhere".

In the era of machine learning, image classification relied on the accuracy of feature extractors. However, with the popularity of deep learning represented by neural networks, especially the application of convolutional neural networks to image classification problems, its accuracy has surpassed the previous manual feature extraction simultaneously, and the end-to-end learning mode has saved a lot of workloads.

On the other hand, as the science of the human body, any breakthrough in medical research needs to be discussed and validated seriously. This poses three major challenges for expanding the application of deep learning to the medical field:

Accuracy Even though the model can predict 99.9% correct and 0.1% incorrect, the incorrect classification results consequently lead to misdiagnosis and delayed treatment, ultimately endangering patients' lives. A correct classification result is crucial for the next treatment, which is the fundamental requirement.

Reliability The current models are proven to perform outstandingly in clean images while poor performance in noisy data. However, it is impossible to input clean images at all times. Therefore, we want the classification model to perform equally well on cleaned and noisy images.

Interpretability Although deep learning models can predict correct classification results, their "black-box" operation mechanism makes it difficult for doctors to trust and accept them.

To clear the above problems, we not only introduce medical knowledge, including medical image analysis and lung infection, but deep learning models for classification problems such as convolutional neural networks in this survey. Moreover, we reproduce a chest X-ray image classification experiment to show the cutting-edge study of image classification on COVID-19 diagnosis. Since the revolution AlexNet brought in 2012, many studies have proven the amount of data significantly impacts the performance of neural networks [70, 71]. The larger, more comprehensive dataset is our major improvement to the original work of Ahmed et al. [68]. Therefore, our contributions to this work are concentrated as follows:

1. As the dataset size increase by over 24 times, we believe our experimental results can better reflect the model performance when used in practice, although our average accuracy is 4.34% lower than the original work.

2. With more species of lung infection added (Lung Opacity → Non-COVID), the application of this model will be widely expanded to more scenarios.

3. Thanks to the continuous updates of saliency map techniques, the visualized results compared to the original work are more intuitive, especially the normal cases are obviously difference from lung-infected cases. Without any medical background, we can easily distinguish between a normal and lung-infected saliency map.

The results demonstrate that the classification model can still achieve high accuracy and partially explain the basis of prediction. We hope that this system will assist radiologists in reducing their burden.

5.2 Future Work

We will first collect more data on lung infections in our future work. In addition to COVID-19 and pneumonia, we expect other X-ray images with similar symptoms, such as influenza. We believe these new images will significantly expand the application of classification models.

Besides, we also noticed that some state-of-the-art models, such as ViT (Vision Transformer), are becoming increasingly popular in computer vision.

In future experiments, we will use ViT instead of CNN as the model substrate to test whether the classification accuracy can be improved.

Finally, we hope to collaborate with radiologists to analyze the highlighted areas considered important in the classification model, so that we may discover new diagnostic methods and fix some potential defects in the model.

To summarize, our future work focuses on improving the generalizability and performance of the model and trying to contribute to the interpretability of the deep learning model for the Chest X-ray images classification problem.

Publications

Conferences

[1] Yu, F., Nguyen, B., Dinh, D., Wu, X., Yu, Y., Fujinami, T. (2022). The Impact of Virtual Reality on Gaming Experience: A Perspective from Half-Life on Steam. In: Tareq Ahram and Redha Taiar (eds) Human Interaction & Emerging Technologies (IHIET-AI 2022): Artificial Intelligence & Future Applications. AHFE (2022) International Conference. AHFE Open Access, vol 23. AHFE International, USA. doi: http://doi.org/10.54941/ahfe100859

Bibliography

[1] S. Albelwi and A. Mahmood, "A framework for designing the architectures of deep convolutional neural networks," *Entropy*, vol. 19, no. 6, p. 242, 2017.

[2] H. Greenspan, B. Van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.

[3] Y. Mintz and R. Brodie, "Introduction to artificial intelligence in medicine," *Minimally Invasive Therapy & Allied Technologies*, vol. 28, no. 2, pp. 73–81, 2019.

[4] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, pp. 221–248, 2017.

[5] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *Ieee Access*, vol. 6, pp. 9375–9389, 2017.

[6] R. C. Chen, T. T. Tan, and L. P. Chan, "Adapting to a new normal? 5 key operational principles for a radiology service facing the covid-19 pandemic," 2020.

[7] J. J. Naidich, A. Boltyenkov, J. J. Wang, J. Chusid, D. Hughes, and P. C. Sanelli, "Impact of the coronavirus disease 2019 (covid-19) pandemic on imaging case volumes," *Journal of the American College of Radiology*, vol. 17, no. 7, pp. 865–872, 2020.

[8] M. A. Khafaji, M. A. Safhi, R. H. Albadawi, S. O. Al-Amoudi, S. S. Shehata, and T. Fadi, "Artificial intelligence in radiology," *Saudi Medical Journal*, vol. 43, no. 1, pp. 53–60, 2022.

[9] Z. Y. Zu, M. D. Jiang, P. P. Xu, W. Chen, Q. Q. Ni, G. M. Lu, and L. J. Zhang, "Coronavirus disease 2019 (covid-19): a perspective from china," *Radiology*, vol. 296, no. 2, pp. E15–E25, 2020.

[10] G. S. Lodwick, T. E. Keats, and J. P. Dorst, "The coding of roentgen images for computer analysis as applied to lung cancer," *Radiology*, vol. 81, no. 2, pp. 185–200, 1963.

[11] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[12] E. Çallı, E. Sogancioglu, B. van Ginneken, K. G. van Leeuwen, and K. Murphy, "Deep learning for chest x-ray analysis: A survey," *Medical Image Analysis*, vol. 72, p. 102125, 2021.

[13] R. P. Kruger, J. R. Townes, D. L. Hall, S. J. Dwyer, and G. S. Lodwick, "Automated radiographic diagnosis via feature extraction and classification of cardiac size and shape descriptors," *IEEE Transactions on Biomedical Engineering*, no. 3, pp. 174–186, 1972.

[14] L. de Oliveira Martins, E. C. da Silva, A. C. Silva, A. C. de Paiva, and M. Gattass, "Classification of breast masses in mammogram images using ripley's k function and support vector machine," in *Machine Learning and Data Mining in Pattern Recognition: 5th International Conference, MLDM 2007, Leipzig, Germany, July 18-20, 2007. Proceedings 5*, pp. 784–794, Springer, 2007.

[15] R.-F. Chang, W.-J. Wu, W. K. Moon, Y.-H. Chou, and D.-R. Chen, "Support vector machines for diagnosis of breast tumors on us images," *Academic radiology*, vol. 10, no. 2, pp. 189–197, 2003.

[16] C. D. L. Nascimento, S. D. d. S. Silva, T. A. d. Silva, W. C. d. A. Pereira, M. G. F. Costa, and C. F. F. Costa Filho, "Breast tumor classification in ultrasound images using support vector machines and neural networks," *Research on Biomedical Engineering*, vol. 32, pp. 283–292, 2016.

[17] E. Geremia, D. Zikic, O. Clatz, B. H. Menze, B. Glocker, E. Konukoglu, J. Shotton, O. Thomas, S. Price, T. Das, *et al.*, "Classification forests for semantic segmentation of brain lesions in multi-channel mri," *Decision Forests for Computer Vision and Medical Image Analysis*, pp. 245–260, 2013.

[18] R. Marée, L. Wehenkel, and P. Geurts, "Extremely randomized trees and random subwindows for image classification, annotation, and retrieval," *Decision Forests for Computer Vision and Medical Image Analysis*, pp. 125–141, 2013.

[19] V. Lepetit and P. Fua, "Keypoint recognition using random forests and random ferns," *Decision Forests for Computer Vision and Medical Image Analysis*, pp. 111–124, 2013.

[20] N. H. Rajini and R. Bhavani, "Classification of mri brain images using k-nearest neighbor and artificial neural network," in *2011 International conference on recent trends in information technology (ICRTIT)*, pp. 563–568, IEEE, 2011.

[21] J. Zhuang, J. Cai, R. Wang, J. Zhang, and W.-S. Zheng, "Deep knn for medical image classification," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23*, pp. 127–136, Springer, 2020.

[22] T. Gupta, T. K. Gandhi, R. Gupta, and B. K. Panigrahi, "Classification of patients with tumor using mr flair images," *Pattern Recognition Letters*, vol. 139, pp. 112–117, 2020.

[23] A. Khatami, S. Araghi, and T. Babaei, "Evaluating the performance of different classification methods on medical x-ray images," *SN Applied Sciences*, vol. 1, pp. 1–7, 2019.

[24] Y. Luo, H. Carretta, I. Lee, G. LeBlanc, D. Sinha, and G. Rust, "Naïve bayesian network-based contribution analysis of tumor biology and healthcare factors to racial disparity in breast cancer stage-at-diagnosis," *Health Information Science and Systems*, vol. 9, pp. 1–14, 2021.

[25] V. D. P. Jasti, A. S. Zamani, K. Arumugam, M. Naved, H. Pallathadka, F. Sammy, A. Raghuvanshi, and K. Kaliyaperumal, "Computational technique based on machine learning and image processing for medical image analysis of breast cancer diagnosis," *Security and communication networks*, vol. 2022, pp. 1–7, 2022.

[26] S. H. Yoo, H. Geng, T. L. Chiu, S. K. Yu, D. C. Cho, J. Heo, M. S. Choi, I. H. Choi, C. Cung Van, N. V. Nhung, *et al.*, "Deep learning-based decision-tree classifier for covid-19 diagnosis from chest x-ray imaging," *Frontiers in medicine*, vol. 7, p. 427, 2020.

[27] W. Zhou, H. Wang, C. Yang, Y. Bai, D. Wang, and Y. Zhan, "Decision tree based medical image clustering algorithm in computer-aided diagnoses," *Journal of Computational Methods in Sciences and Engineering*, vol. 15, no. 4, pp. 645–651, 2015.

[28] A. Criminisi and J. Shotton, *Decision forests for computer vision and medical image analysis*. Springer Science & Business Media, 2013.

[29] J. Jiang, P. Trundle, and J. Ren, "Medical image analysis with artificial neural networks," *Computerized Medical Imaging and Graphics*, vol. 34, no. 8, pp. 617–631, 2010.

[30] I. Lorencin, N. Andelic, J. Spanjol, and Z. Car, "Using multi-layer perceptron with laplacian edge detector for bladder cancer diagnosis," *Artificial Intelligence in Medicine*, vol. 102, p. 101746, 2020.

[31] J. Ren, "Ann vs. svm: Which one performs better in classification of mccs in mammogram imaging," *Knowledge-Based Systems*, vol. 26, pp. 144–153, 2012.

[32] G. Vishnuvarthanan, M. P. Rajasekaran, P. Subbaraj, and A. Vishnuvarthanan, "An unsupervised learning method with a clustering approach for tumor identification and tissue segmentation in magnetic resonance brain images," *Applied Soft Computing*, vol. 38, pp. 190–212, 2016.

[33] D.-T. Dinh, V.-N. Huynh, and S. Sriboonchitta, "Clustering mixed numerical and categorical data with missing values," *Information Sciences*, vol. 571, pp. 418–442, 2021.

[34] H. Ng, S. Ong, K. Foong, P.-S. Goh, and W. Nowinski, "Medical image segmentation using k-means clustering and improved watershed algorithm," in *2006 IEEE southwest symposium on image analysis and interpretation*, pp. 61–65, IEEE, 2006.

[35] T. Schultz and G. L. Kindlmann, "Open-box spectral clustering: applications to medical image analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2100–2108, 2013.

[36] D. Mahapatra, B. Bozorgtabar, and R. Garnavi, "Image super-resolution using progressive generative adversarial networks for medical image analysis," *Computerized Medical Imaging and Graphics*, vol. 71, pp. 30–39, 2019.

[37] T. Zhang, J. Cheng, H. Fu, Z. Gu, Y. Xiao, K. Zhou, S. Gao, R. Zheng, and J. Liu, "Noise adaptation generative adversarial network for medical image analysis," *IEEE transactions on medical imaging*, vol. 39, no. 4, pp. 1149–1159, 2019.

[38] M. AlAmir and M. AlGhamdi, "The role of generative adversarial network in medical image analysis: An in-depth survey," *ACM Computing Surveys*, vol. 55, no. 5, pp. 1–36, 2022.

[39] S. Chen, G. Bortsova, A. García-Uceda Juárez, G. Van Tulder, and M. De Bruijne, "Multi-task attention-based semi-supervised learning for medical image segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Confer-*

ence, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22, pp. 457–465, Springer, 2019.

[40] C. Kaul, S. Manandhar, and N. Pears, "Focusnet: An attention-based fully convolutional network for medical image segmentation," in *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*, pp. 455–458, IEEE, 2019.

[41] A. Rao, J. Park, S. Woo, J.-Y. Lee, and O. Aalami, "Studying the effects of self-attention for medical image analysis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3416–3425, 2021.

[42] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pp. 3–11, Springer, 2018.

[43] N. Siddique, P. Sidike, C. Elkin, and V. Devabhaktuni, "U-net and its variants for medical image segmentation: theory and applications," *arXiv preprint arXiv:2011.01118*, 2020.

[44] G. Du, X. Cao, J. Liang, X. Chen, and Y. Zhan, "Medical image segmentation based on u-net: A review," *Journal of Imaging Science and Technology*, 2020.

[45] D. Sarwinda, R. H. Paradisa, A. Bustamam, and P. Anggia, "Deep learning in image classification using residual network (resnet) variants

for detection of colorectal cancer," *Procedia Computer Science*, vol. 179, pp. 423–431, 2021.

[46] S. Ayyachamy, V. Alex, M. Khened, and G. Krishnamurthi, "Medical image retrieval using resnet-18," in *Medical imaging 2019: imaging informatics for healthcare, research, and applications*, vol. 10954, pp. 233–241, SPIE, 2019.

[47] S. Chen, K. Ma, and Y. Zheng, "Med3d: Transfer learning for 3d medical image analysis," *arXiv preprint arXiv:1904.00625*, 2019.

[48] T. Zhou, X. Ye, H. Lu, X. Zheng, S. Qiu, and Y. Liu, "Dense convolutional network and its application in medical image analysis," *BioMed Research International*, vol. 2022, 2022.

[49] A. Riasatian, M. Babaie, D. Maleki, S. Kalra, M. Valipour, S. Hemati, M. Zaveri, A. Safarpoor, S. Shafiei, M. Afshari, *et al.*, "Fine-tuning and training of densenet for histopathology image representation using tcga diagnostic slides," *Medical Image Analysis*, vol. 70, p. 102032, 2021.

[50] Z. Huang, X. Zhu, M. Ding, and X. Zhang, "Medical image classification using a light-weighted hybrid neural network based on pcanet and densenet," *Ieee Access*, vol. 8, pp. 24697–24712, 2020.

[51] J.-H. Shu, F.-D. Nian, M.-H. Yu, and X. Li, "An improved mask r-cnn model for multiorgan segmentation," *Mathematical Problems in Engineering*, vol. 2020, pp. 1–11, 2020.

[52] G. Zhu, Z. Piao, and S. C. Kim, "Tooth detection and segmentation with mask r-cnn," in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pp. 070–072, IEEE, 2020.

[53] R. Anantharaman, M. Velazquez, and Y. Lee, "Utilizing mask r-cnn for detection and segmentation of oral diseases," in *2018 IEEE international conference on bioinformatics and biomedicine (BIBM)*, pp. 2197–2204, IEEE, 2018.

[54] Y. Xie, J. Zhang, C. Shen, and Y. Xia, "Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*, pp. 171–180, Springer, 2021.

[55] J. Zhu, Y. Li, Y. Hu, K. Ma, S. K. Zhou, and Y. Zheng, "Rubik's cube+: A self-supervised feature learning framework for 3d medical image analysis," *Medical image analysis*, vol. 64, p. 101746, 2020.

[56] H. Yu, L. T. Yang, Q. Zhang, D. Armstrong, and M. J. Deen, "Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives," *Neurocomputing*, vol. 444, pp. 92–110, 2021.

[57] C. Zeelan Basha, T. Maruthi Padmaja, and G. Balaji, "Automatic x-ray image classification system," in *Smart Computing and Informatics*, pp. 43–52, Springer, 2018.

[58] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of covid-19 cases using deep neural networks with x-ray images," *Computers in biology and medicine*, vol. 121, p. 103792, 2020.

[59] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images," *arXiv preprint arXiv:2003.11055*, 2020.

[60] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Physical and engineering sciences in medicine*, vol. 43, no. 2, pp. 635–640, 2020.

[61] Z. Karhan and A. Fuat, "Covid-19 classification using deep learning in chest x-ray images," in *2020 Medical Technologies Congress (TIPTEKNO)*, pp. 1–4, IEEE, 2020.

[62] M. Nixon and A. Aguado, *Feature extraction and image processing for computer vision*. Academic press, 2019.

[63] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," *IEEE transactions on neural networks and learning systems*, 2021.

[64] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, *et al.*, "Recent advances in convolutional neural networks," *Pattern recognition*, vol. 77, pp. 354–377, 2018.

[65] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, 2016.

[66] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 international conference on engineering and technology (ICET)*, pp. 1–6, Ieee, 2017.

[67] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," *arXiv preprint arXiv:1603.07285*, 2016.

[68] F. Ahmed, S. A. C. Bukhari, and F. Keshtkar, "A deep learning approach for covid-19 8 viral pneumonia screening with x-ray images," *Digital Government: Research and Practice*, vol. 2, no. 2, pp. 1–12, 2021.

[69] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *arXiv preprint arXiv:1312.6034*, 2013.

[70] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, M. Hasan, B. C. Van Essen, A. A. Awwal, and V. K. Asari, "A state-of-the-art survey on deep learning theory and architectures," *Electronics*, vol. 8, no. 3, p. 292, 2019.

[71] A. Najah, F. F. Mustafa, and W. S. Hacham, "Building a high accuracy transfer learning-based quality inspection system at low costs," *Al-Khwarizmi Engineering Journal*, vol. 17, no. 1, pp. 1–12, 2021.