Ministry of Education and Science of the Republic of Kazakhstan
Suleyman Demirel University

UDC.....

Anuar Tastembekov

Development of algorithms and software for recognition of airway inflammation based on image processing

THESIS
Presented in Partial Fulfillment for the
Degree of Master of Science in Computing Systems and Software
(degree code: 6M070400)
Department of Computer Sciences
Faculty of Engineering and Natural Sciences

Supervisor: **Edilkhan Amirgaliyev**

Kaskelen, 2020

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Anuar Tastembekov

Abstract

Every person is always happy to make sure that everything is good with his body and that he is not ill. It is the desire of doctors and their patients not to face the consequences of diseases, but to timely detect or prevent them makes diagnostic procedures so popular. The X-ray is one of the most common methods for diagnosing various lung diseases and is prescribed much more often than other types of examinations - magnetic resonance imaging or computed tomography. In this thesis, we develop an algorithm and application to analyze X-ray images to predict airway inflammation. We preprocess the data by cropping the lung area to have more accurate results at analysis. Our goal is to help the medical system in a mass preliminary inspection during the SARS-COV-2 pandemic. The results obtained in this work can be used in the future to help doctors to conduct the examination more efficiently, which will reduce the time spent on each person.

Аңдатпа

Әр адам әрдайым денесінің жақсы екеніне және оның ауырмайтынына көз жеткізуге қуанышты. Бұл дәрігерлер мен олардың пациенттерінің аурудың салдарымен бетпе-бет келмеуін қалайды, бірақ оларды дер кезінде анықтау немесе алдын-алу диагностикалық процедураларды соншалықты танымал етеді. Рентгенография өкпенің әртүрлі ауруларын диагностикалаудың кең таралған әдістерінің бірі болып табылады және басқа тексерулерге қарағанда жиі тағайындалады - магниттік-резонансты томография немесе компьютерлік томография. Бұл тезисте біз тыныс жолдарының қабынуын болжау үшін рентген суреттерін талдау алгоритмі мен қосымшасын жасаймыз. Талдау кезінде дәлірек нәтиже алу үшін өкпенің аймағын кесу арқылы деректерді алдын-ала өңдейміз. Біздің мақсатымыз - SARS-COV-2 пандемиясының кезіндегі медициналық жүйеге алдын ала тексеруден өтуге көмектесу. Осы жұмыста алынған нәтижелер болашақта дәрігерлерге емтиханды тиімдірек өткізуге көмектеседі, бұл әр адамға жұмсалатын уақытты қысқартады.

Аннотация

Каждый человек всегда рад убедиться, что с его телом все хорошо и что он не болен. Врачи и их пациенты стремятся не сталкиваться с последствиями заболеваний, а своевременно выявлять или предотвращать их, что делает диагностические процедуры такими популярными. Рентген является одним из наиболее распространенных методов диагностики различных заболеваний легких и назначается гораздо чаще, чем другие виды исследований - магнитно-резонансная томография или компьютерная томография. В этой диссертации мы разрабатываем алгоритм и приложение для анализа рентгеновских изображений для прогнозирования воспаления дыхательных путей. Мы предварительно обрабатываем данные, обрезая область легких, чтобы получить более точные результаты при анализе. Наша цель - помочь медицинской системе провести массовый предварительный осмотр во время пандемии SARS-COV-2. Полученные в этой работе результаты могут быть использованы в будущем, чтобы помочь врачам более эффективно проводить обследование, что сократит время, затрачиваемое на каждого человека.

Acknowledgements

I wanted to thank everyone who helped to complete this work. Special thanks to my thesis supervisor for constant support and useful discussion; to the external reviewer for very useful feedback that helped to significantly improve the current work.

To my family

Contents

Nomenclature

CIS The Commonwealth of Independent States

CNN Convolutional neural networks

CT Computed tomography

FCN Fully Convolutional Network

MRI Magnetic resonance imaging

MSE Root Mean Squared Error

ReLU Rectified Linear Unit

SARS-CoV-2 Severe acute respiratory syndrome-related coronavirus 2

1. Introduction

1.1 Motivation

Every person would be happy to make sure that everything is in good with his body and that he is not ill with anything. It is the desire of doctors and their patients not to face the consequences of diseases, but to timely detect or prevent them makes diagnostic procedures so popular. Especially doctors from all over the world, as well as their wards, can be grateful to William Roentgen (see Figure 1.1), whose discoveries make it possible to undergo X-ray and fluorography today [1].

Figure 1.1: Wilhelm Conrad Röntgen (27 March 1845 – 10 February 1923)

X-ray is one of the most common methods for diagnosing various lung diseases and is prescribed much more often than other types of examinations - magnetic resonance imaging or computed tomography. The principle of obtaining an x-ray image is simple - a beam of rays emanating from the radiation tube of the apparatus, passing through the human body to varying degrees, is projected onto the film (see Figure 1.2). In fact, the method resembles the process of making

an ordinary photograph, but due to the ability of organs to transmit X-rays differently, a picture is obtained in which soft tissues have shades of gray, air cavities are displayed in black, and bones that absorb radiation, on the contrary, are bright white . Radiography can be overview - in cases where it is necessary to consider the lungs as a whole or sight when the organ fragment is examined.

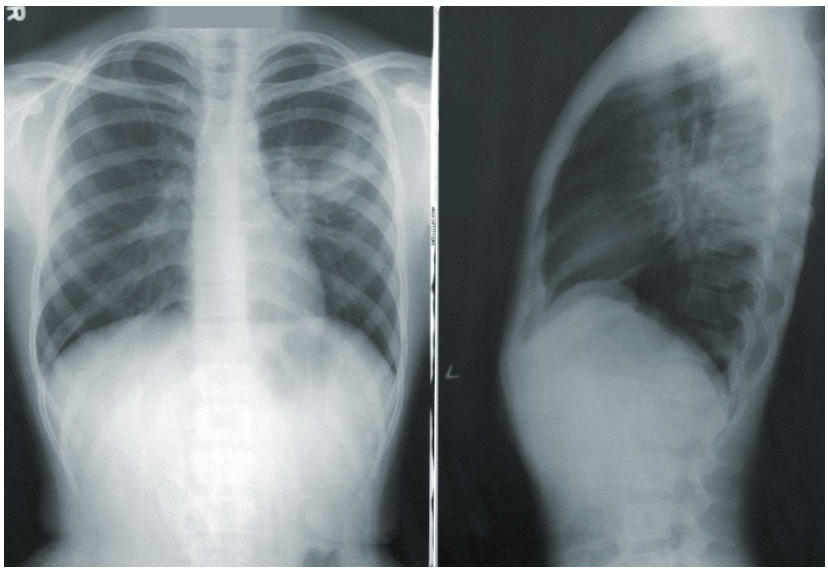

Figure 1.2: An X-ray images of a human chest

Computed tomography is also based on x-ray radiation, which passes through the body from several angles at once. The received “frames” are processed by a computer and “connected” into a common image. The informational content of CT in studying the state of the lungs is much higher, but the cost of the procedure is 3-4 times more expensive than x-ray, and the radiation dose is more significant, so a similar study is prescribed, mainly to clarify the diagnosis.

With magnetic resonance imaging, images are obtained by applying a magnetic field to the body. Despite its harmlessness to the body, MRI also has a high cost and limitations in use: for example, the patient may have a pacemaker, some types of metal implants and prostheses.

What research should be used in a particular case, should be determined by the doctor. In general, MRI and CT are inappropriate to use for prophylactic purposes and to familiarize themselves with the general condition of the lungs, for this X-ray is most often used.

Often people do not know the difference between the concepts of "x-ray" and "fluorography", but in fact these are different procedures. Fluorography, which is common for the population of the CIS countries, is actually an outdated

method of examination, in which the image from the radiographic apparatus is photographed on film. Such a diagnosis is the most accessible for the masses because of its cheapness, but also the most inaccurate - the clarity of the image is several times less than x-ray and does not allow to identify many lung problems. Therefore, in this work, all the data were obtained using x-rays. That will allow you to get more accurate results in the end.

1.2 Aims and Objectives

Currently, lung x-rays are used by doctors to diagnose various pathologies of the bronchopulmonary system. This method is effective for detecting diseases such as pneumonia, tuberculosis, cancerous tumors, fungal diseases, as well as for detecting any lung abnormalities.

The ultimate goal of the procedure is to get an x-ray with which the doctor can diagnose and prescribe treatment. Only a professional radiologist can decrypt the image correctly, who, in the form of dimming and enlightenment, the intensity of the lines and their shade, makes a conclusion about the state of internal organs. For example, lung cancer in the images is characterized by rounded darkening of different diameters with clear boundaries. Large shadows with blurry edges can indicate fungal, cardiovascular diseases, pneumonia. Tuberculosis is characterized by intense lines of the lungs, as well as many small dark areas.

But here the following problem arises - the lack of specialists as a result of the large flow of customers that we saw recently during the SARS-CoV-2 pandemic. Various methods are used to reduce the burden on the health system, such as preliminary questionnaires and passing rapid tests. In this paper, we want to automate the analysis of x-rays for preliminary detection of the diagnosis in a patient using modern machine learning methodology. The created program will help doctors conduct the examination more efficiently, which will reduce the time spent on each person.

1.3 Thesis Outline

We develop an algorithm and application to analyze X-ray images to predict airway inflammation. An important starting point for analyzing lung images

is segmentation of lung x-ray images (see Figure 1.3). This is a significant step forward to give an accurate analysis of lung x-rays, such as diagnosing pneumonia, tuberculosis, and SARS-CoV-2. We propose finding the boundaries of lung x-ray images utilizing one of the most popular architecture in the finding object boundaries in images called U-Net. The network comprises a contracting path and an expansive path, which gives it the u-formed shape. This architecture can be prepared with a small number of images, and it surpasses many other architectures.

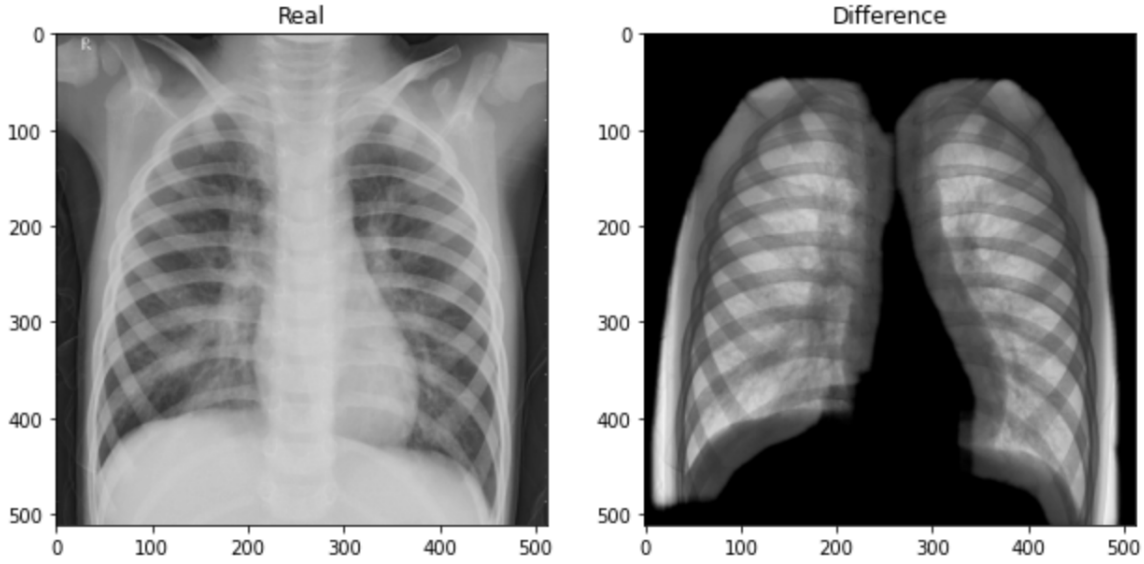

Figure 1.3: Predicted of a lung boundaries

Precisely distinguishing and identifying the airway inflammation types is a significant medical issue, and automated strategies can be utilized to save time and decrease mistakes. Automated analysis strategy provides additional assistance to doctors to simplify the diagnosis of the disease. For all of the above reasons, we have developed an effective CNN-based strategy for recognizing the various types of lung diseases on x-ray images.

2. Preliminaries

The aim of this work is the application of machine learning methods in the search for pathologies on x-ray images of the lungs and conducting relevant studies. To achieve this goal the following tasks are set:

1. Prepare data for segmentation: concatenate mask images, convert images to numpy array and store in pickle format.

2. Develop deep learning model for segmentation: build the most optimal model architecture and implement it.

3. Develop deep learning model for analyzing segmented image: build the most optimal model architecture and implement it.

We used two different sources of input images for trinaing segmentation model. The whole dataset[2] contains 801 X-Ray lung images from No.3 Hospital in Shenzhen, Guangdong region, China (662 images), and from the tuberculosis control program of the Department of Health and Human Services of Montgomery County, MD, USA (139 images).

Geographically distant sources of images from different countries, scanners, and nations would help to acquire more sparse and convincing results.

The architecture that we used to find boundaries of lungs - the U-Net - is one of the popular architectures in the class of image segmentation. Thus, adopted architecture was used in this work.

The next step, analysing preprocessed images was developed over ZFNet architecture by utilizing the dataset for training consisting of 5232 images that were collected by Kermany D, Goldbaum M, Cai W et al.[3]

The architecture is similar to the AlexNet network created by Alex Krizhevsky. However, ZFNet has more filters on the layer and nested convolutional layers. The

network includes convolutions, wide pooling, dropout, data aggregation, ReLU activation functions, and stochastic gradient descent.

3. Overview

This chapter provides the necessary concepts from the theory of machine learning, describes the methods of preprocessing x-ray images, as well as the data used in this work.

3.1 Machine Learning

Machine learning every day occupies an increasing place in our life due to the huge range of its applications. Starting from the analysis of traffic jams and ending with self-driving cars, more and more tasks are being shifted to self-learning cars [4].

Machine learning is considered a branch of artificial intelligence, the main idea of which is that the computer does not just use a pre-written algorithm, but learns how to solve the problem.

Any working machine learning technology can be conditionally assigned to one of three levels of accessibility. The first level is when it is available only to various technological giants of the Google or IBM level. The second level is when a student with some knowledge can use it. The third level is when even a granny is able to control her.

Machine learning is now at the junction of the second and third levels, due to which the rate of change of the world with this technology is growing every day.

Most machine learning tasks can be briefly divided into supervised learning and unsupervised learning [5]. If you imagined a programmer with a whip in one hand and a piece of sugar in the other, you were a little mistaken. By "teacher" here is meant the very idea of human intervention in data processing. When teaching with a teacher, we have data on the basis of which we need to predict something, and some hypotheses. When learning without a teacher, we only have

data whose properties we want to find. In the examples, you will see the difference a little clearer.

3.2 Machine Learning tasks

Machine learning has general tasks that can solve. Below is the general list of issues:

3.2.1 Regression

Based on various signs to predict a material response. In other words, the answer may be 1, 5, 23.575 or any other real number, which, for example, can represent the value of the apartment. Examples: predicting the value of a stock in six months, predicting the profit of a store next month, predicting the quality of wine in a blind test.

3.2.2 Classification

Based on various signs to predict a categorical answer. In other words, there are a finite number of answers in such a task, as in the case of determining whether a patient has cancer or determining whether a letter is spam. Examples: handwriting recognition of a text, determination of whether a person or a cat is in a photograph.

3.2.3 Clusterization

Splitting data into similar categories. Examples: the breakdown of customers of a cellular operator by solvency, the breakdown of space objects into similar ones (galaxies, planets, stars, and so on).

3.2.4 Dimensionality reduction

Train to describe our data not with N signs, but with a smaller number (usually 2-3 for subsequent visualization). As an example, in addition to the need for visualization, data compression can be cited.

3.2.5 Anomaly detection

On the basis of signs to learn to distinguish between distinguishing anomalies from "non-anomalies." It seems that this task is no different from the classification problem. But the peculiarity of identifying anomalies is that we either have very few or no examples of anomalies for training the model, so we cannot solve such a problem as the classification problem. Example: identifying fraudulent bank card transactions.

3.3 Deep Neural Networks

In machine learning, there are a large number of algorithms, some of which are quite universal. For example, including the support vector method, boosting over decision trees, and the deep neural networks.

A neural network is a network of neurons, where each neuron is a mathematical model of a real neuron. Neural networks began to be very popular in the 80s and early 90s, however, in the late 90s, their popularity fell sharply. However, recently it is one of the advanced technologies used in machine learning, used in a huge number of applications. The reason for the return of popularity is simple: the computing abilities of computers have increased.

A neuron in an neural network is a mathematical function (e.q the Sigmoid and ReLU functions), which receives some value at the input and the output obtained using the mathematical function.

4. Data preprocessing

One of the key stages of deep learning is the data preprocessing phase. Depending on the specifics of the study area and the goals pursued, one or another pre-processing methods is selected. In this paper, we will use the following methods of X-ray pre-processing: enhancement of image contrast and segmentation of the lungs.

4.1 Enhancement of image contrast

One of the significant factors affecting image quality is the level of image contrast. If the contrast is too low or too high, then the image is unlikely to be suitable for training. To increase the contrast of the image using methods of contrast enhancement.

In this work we used a general solution to increase the local contrast of x-ray images:

$$L_(LCE) = \frac{L - L'}{\sqrt{L^2 - (L')^2}}$$

where L is the original image, and L' is the image to which Gaussian blur applied.

4.2 Segmentation of Chest Image

Getting to the analysis of x-ray images of the lungs we begin with the selection of the contour of the lungs. Thus, by the task of lung segmentation is meant the recognition and subsequent separation of the lung region from the heart and other organs present in the images. To solve this problem, a convolutional neural network based on the UNet architecture was chosen. This architecture has proven

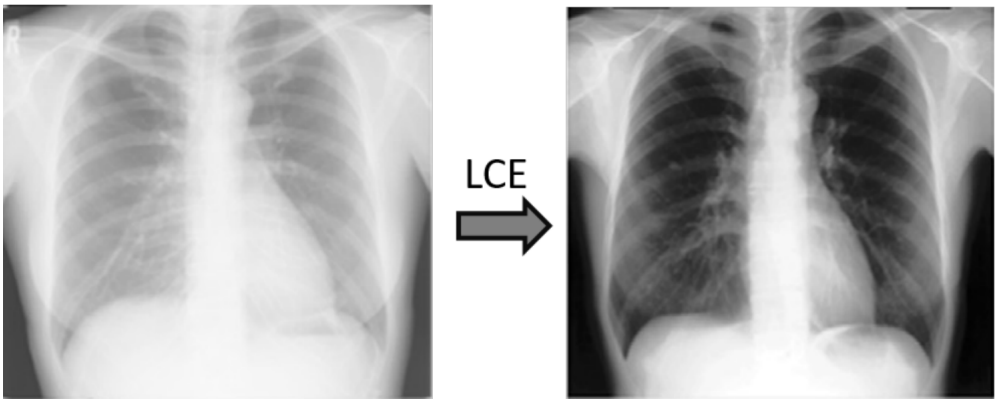

Figure 4.1: An example of the local contrast enhancement method

to be the best in applying to the problems of segmentation of biomedical images [6].

The disadvantages of this architecture include the requirement for high computing power, as well as the relatively weak distribution of images by type of pathology. The first drawback is eliminated by the availability of appropriate equipment that allows training using a graphic processor (GPU). The second - with a high degree of probability, will not affect the learning outcomes, since in this paper we consider the task of binary classification, namely the task of determining the presence or absence of pathology in the X-ray image.

The steps of lung segmentation method involved in this work is applied to the original images:

1. Search for lung boundaries by training a convolutional neural network based on UNet for lung segmentation on the initial data set with manually prepared “masks” (lung boundaries).

2. Determination of lung boundaries in the form of black-and-white “masks” using a trained convolutional neural network for each source image from a data set.

3. Separation of the area of interest (left and right lungs) from the original images.

Please follow the next chapter to get more detailed information about the lung segmentation architecture step explained here.

Figure 4.2: An example of the lung segmentation

5. Segmentation of Chest Image

In essence, segmentation is a procedure that divides an image into regions. This method is an image processing approach that helps us isolate objects and surfaces in images. Segmentation is mainly used in applications, for example, for the remote detection or recognition of tumors in medicine.

It is known that the segmentation of medical images is one of the tough issues in the field of image analytics and preprocessing [7]. The segmentation of the image goes before different phases of image analysis, thus any mistakes in defining incorrect object boundaries affect all subsequent steps.

Despite the issue of lung segmentation in lung x-rays, usually in several studies, the results of automatic extraction of the lung boundaries stay weak in many cases. This is particularly true with lung segmentation, which is influenced by obsessive processes and, what's more, extraordinary changes related with age and environment. The issue of accurate segmentation has been additionally disturbed in a situation of mass screening of people.

There are many traditional methods of automatic segmentation. For example, methods for detecting points, lines and edges, morphological, threshold, pixel-based, regional, cluster approaches, and so on. Various methods have been created for segmentation using convolutional neural networks, which have become more effective in solving more complex problems with image segmentation. In this article, we will look at one such architecture: U-Net [6].

Deep Learning approaches. There is a huge interest in studying Deep Learning in recent years. This approach is generally explained as a part of Machine Learning that works on algorithms that are based on a high abstraction level of input that gives huge flexibility.

Here we cannot provide general image training using classic convolutional neural networks through tag labels because these problems need information on localization/positioning using pixel approaches. And usually, deep learning methodologies need huge data sets to train the model. But we often cannot control the amount of data that must be collected for this image processing problem. In this particular situation, accessibility means time, money, and hardware.

U-Net here is more successful than other traditional models, from the point of view of architecture and from the point of view of pixel-based image segmentation formed from convolutional layers of a neural network. It is even effective with a small amount of data set.

5.1 Materials

We used two different sources of input images for our experiment. The original and corresponding mask images of the lung are shown Figure 5.2, and Figure 5.1.

The whole dataset contains 801 X-Ray lung images [8] from No.3 Hospital in Shenzhen, Guangdong region, China (662 images), and from the tuberculosis control program of the Department of Health and Human Services of Montgomery County, MD, USA (139 images).

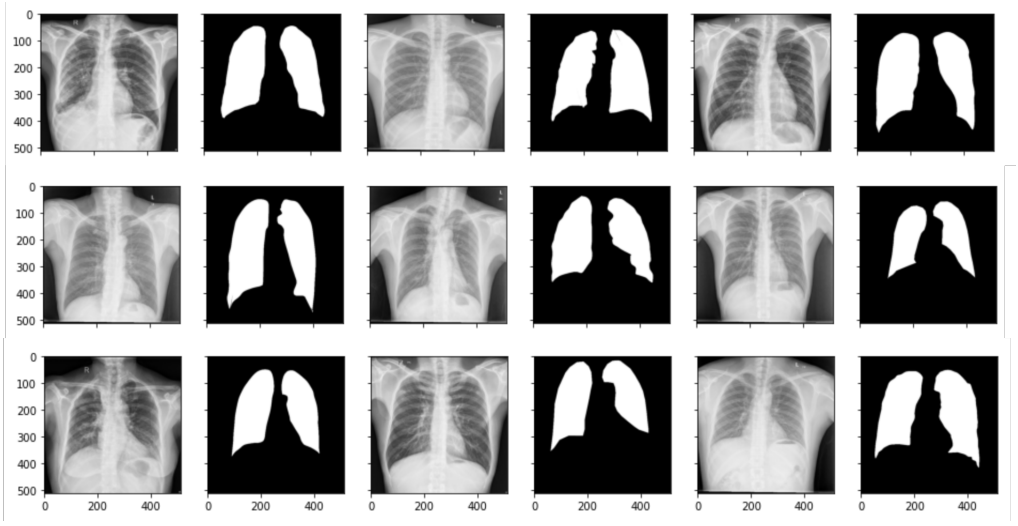

Figure 5.1: Examples of original and mask images from No.3 Hospital in Shenzhen, Guangdong province, China

Geographically distant sources of images from different countries, scanners, and nations would help to acquire more sparse and convincing results. Machine learning architectures have enormously improved the present state. Especially,

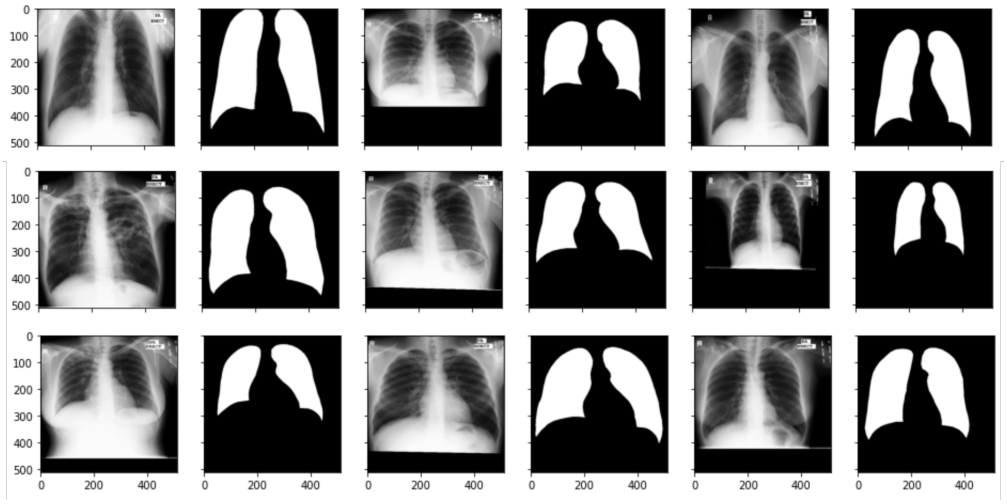

Figure 5.2: Examples of original and mask images from the Department of Health and Human Services of Montgomery County, MD, USA

Deep Learning as a part of machine learning started to be very popular in pattern recognition problems after Hinton proposed his new architecture in 2006 [9]. This methodology has won numerous international contests of pattern recognition challenges since then [10].

5.2 Methods

Image classification with CNN has made colossal improvement in images segmentation issues since FCN was introduced in 2015 [11]. In addition, when the CNN architecture became famous for its dense prediction without fully connected layers, this new approach helped for segmentation tasks for any image and was a lot faster than the previous methods of image segmentation.

The U-Net, is one of the popular architectures in the class of image segmentation. Thus, adopted architecture was used in this work. It is an architecture proposed by Ronneberger et al [12]. based on the Fully Connected Network architecture. The basis of the U-Net can be said as a union of convolution layers at the contracting path and deconvolution layers at the expansive path (see Figure 5.6).

The contracting path is similar to the classical convolutional neural network architecture, it uses Max-pooling and Convolution layers with activation function. For this purpose we used ReLU. In the expansive path, it has up-convolution and convolution layers with ReLU after the upsampling feature map. Because of losing

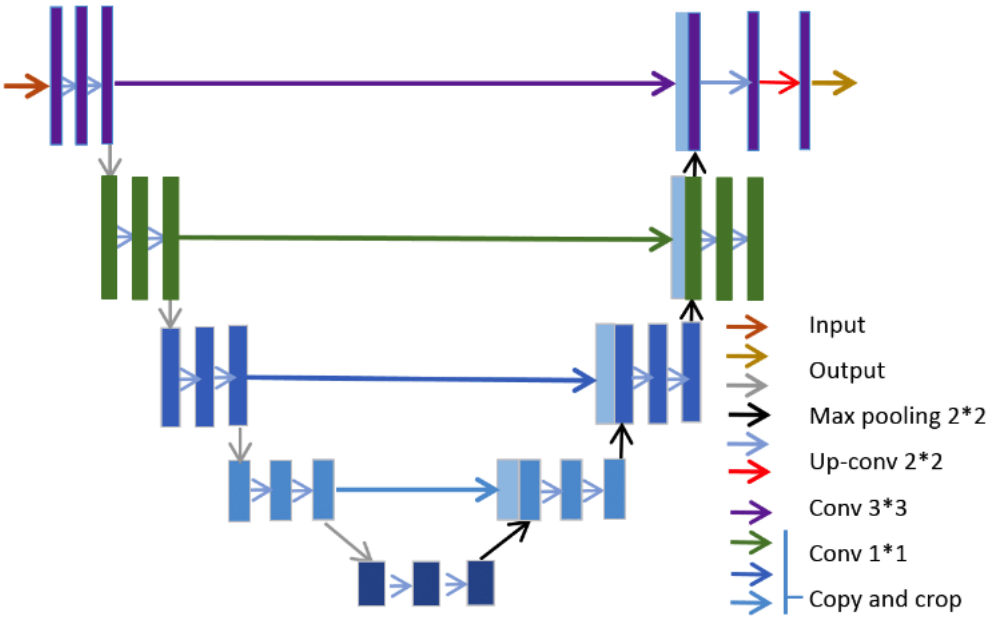

Figure 5.3: U-Net architecture

adjacent pixels data on each convolution the model crops the map of objects from the extraction path and concatenate with the layers in the expansive path

The input images and masks are used to train the U-Net model. At the testing we input images to predict a boundary, and the output we apply to original image to crop interested area.

5.3 Results

Our adopted U-Net architecture was trained on GPU (AMD Radeon Pro 5500M with 4GB of memory) and developed by using Keras framework, high-level API based on TensorFlow. The hyperparameters of the network:

- Batch size: 4
- Number of epochs: 11
- Learning rate: 1e-3
- Optimizer: Adam

With our network, the average accuracy has increased to 96.82%. Figure 5.4 shows the results of experiments in which we show the resulting segmentation using our network.

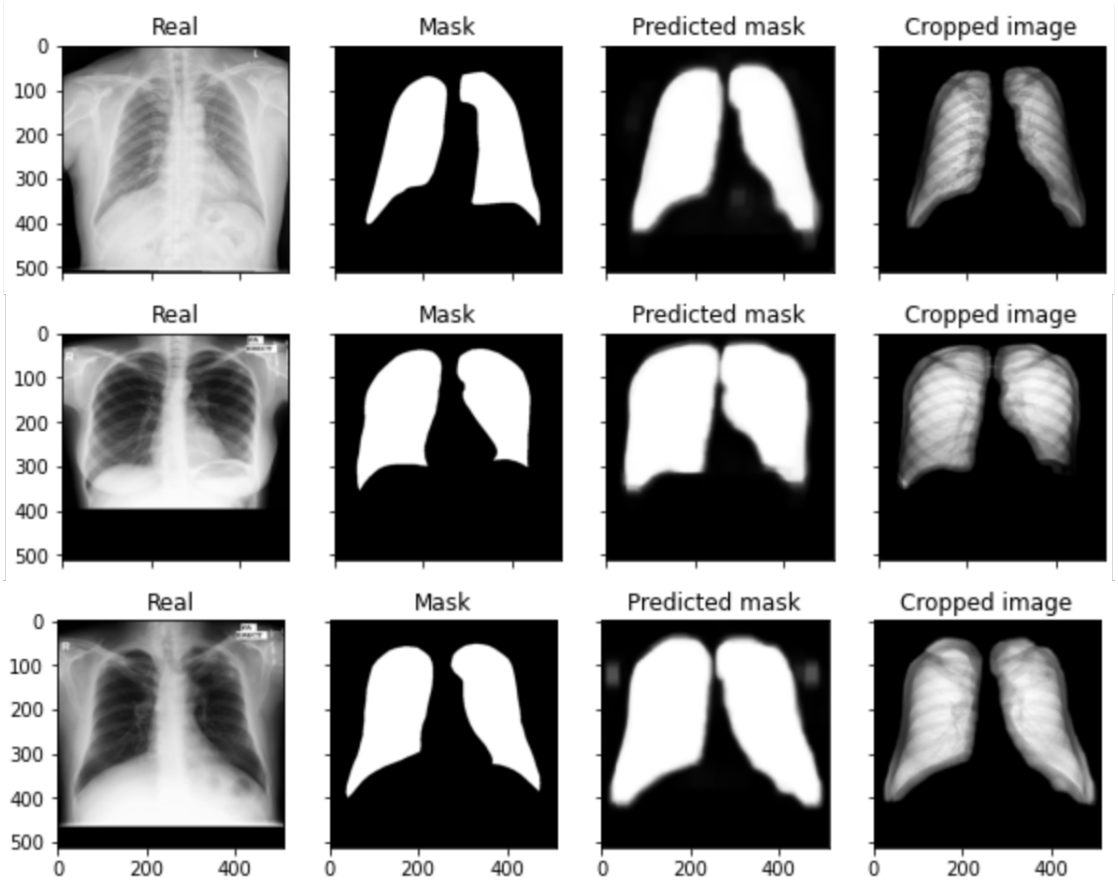

Figure 5.4: First column is the real image; second is manually created mask; third is predicted lung area; and the last is the real image cropped by the prediction mask.

By using the U-Net architecture the results we get segmentation with 96.82% accuracy and 0.0708 loss at the testing stage (see Figure 5.5).

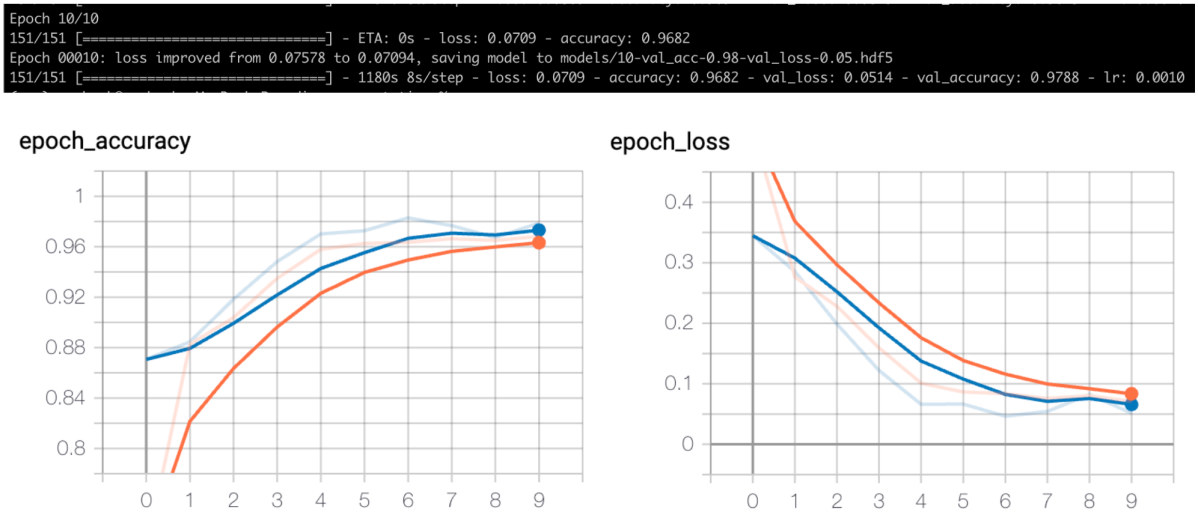

Figure 5.5: U-Net architecture training results

Layer (type)	Output Shape	Param #	Connected to
input_2 (InputLayer)	[(None, 256, 256, 1)	0	
conv2d_5 (Conv2D)	(None, 256, 256, 32)	320	input_2[0][0]
conv2d_6 (Conv2D)	(None, 256, 256, 32)	9248	conv2d_5[0][0]
max_pooling2d_3 (MaxPooling2D)	(None, 128, 128, 32)	0	conv2d_6[0][0]
conv2d_7 (Conv2D)	(None, 128, 128, 64)	18496	max_pooling2d_3[0][0]
conv2d_8 (Conv2D)	(None, 128, 128, 64)	36928	conv2d_7[0][0]
max_pooling2d_4 (MaxPooling2D)	(None, 64, 64, 64)	0	conv2d_8[0][0]
conv2d_9 (Conv2D)	(None, 64, 64, 128)	73856	max_pooling2d_4[0][0]
conv2d_10 (Conv2D)	(None, 64, 64, 128)	147584	conv2d_9[0][0]
max_pooling2d_5 (MaxPooling2D)	(None, 32, 32, 128)	0	conv2d_10[0][0]
conv2d_11 (Conv2D)	(None, 32, 32, 256)	295168	max_pooling2d_5[0][0]
conv2d_12 (Conv2D)	(None, 32, 32, 256)	590080	conv2d_11[0][0]
max_pooling2d_6 (MaxPooling2D)	(None, 16, 16, 256)	0	conv2d_12[0][0]
conv2d_13 (Conv2D)	(None, 16, 16, 512)	1180160	max_pooling2d_6[0][0]
conv2d_14 (Conv2D)	(None, 16, 16, 512)	2359808	conv2d_13[0][0]
conv2d_transpose (Conv2DTranspo	(None, 32, 32, 256)	524544	conv2d_14[0][0]
concatenate (Concatenate)	(None, 32, 32, 512)	0	conv2d_transpose[0][0] conv2d_12[0][0]
conv2d_15 (Conv2D)	(None, 32, 32, 256)	1179904	concatenate[0][0]
conv2d_16 (Conv2D)	(None, 32, 32, 256)	590080	conv2d_15[0][0]
conv2d_transpose_1 (Conv2DTrans	(None, 64, 64, 128)	131200	conv2d_16[0][0]
concatenate_1 (Concatenate)	(None, 64, 64, 256)	0	conv2d_transpose_1[0][0] conv2d_10[0][0]
conv2d_17 (Conv2D)	(None, 64, 64, 128)	295040	concatenate_1[0][0]
conv2d_18 (Conv2D)	(None, 64, 64, 128)	147584	conv2d_17[0][0]
conv2d_transpose_2 (Conv2DTrans	(None, 128, 128, 64)	32832	conv2d_18[0][0]
concatenate_2 (Concatenate)	(None, 128, 128, 128	0	conv2d_transpose_2[0][0] conv2d_8[0][0]
conv2d_19 (Conv2D)	(None, 128, 128, 64)	73792	concatenate_2[0][0]
conv2d_20 (Conv2D)	(None, 128, 128, 64)	36928	conv2d_19[0][0]
conv2d_transpose_3 (Conv2DTrans	(None, 256, 256, 32)	8224	conv2d_20[0][0]
concatenate_3 (Concatenate)	(None, 256, 256, 64)	0	conv2d_transpose_3[0][0] conv2d_6[0][0]
conv2d_21 (Conv2D)	(None, 256, 256, 32)	18464	concatenate_3[0][0]
conv2d_22 (Conv2D)	(None, 256, 256, 32)	9248	conv2d_21[0][0]
conv2d_23 (Conv2D)	(None, 256, 256, 1)	33	conv2d_22[0][0]

Total params: 7,759,521
Trainable params: 7,759,521
Non-trainable params: 0

Figure 5.6: Segmentation model

6. Lung Nodule Analysis

This chapter provides some of the features of the individual work steps. As the main tool, as well as during segmentation, Keras and Tensorflow auxiliary machine learning libraries were used.

6.1 Materials

There are many datasets containing x-rays of the lungs, both healthy and pathological. After studying several examples of suitable data sets, preference was given to a dataset prepared by the Clinical Center at the US National Institutes of Health (NIH Clinical Center) [13]. The dataset is referred to as the NIH Chest X-rays Dataset.

The dataset is referred to as the NIH Chest X-rays Dataset. The key the selection criterion was the number of images contained in the set, since one of the features of deep learning methods is the need for as much as possible source data. The selected dataset is freely available and contains a little over 112 thousand pictures.

Images are presented in PNG format with a color depth of 8 bits. Each image has a resolution of 1024x1024 pixels. But to optimize the training we resized the images to 512x512 pixels.

After analyzing the contents of the dataset, the ratio of healthy images (about 60% of the dataset) to images containing any pathology (about 40% of the dataset) was established.

Below are the examples of images from the NIH Chest X-rays Dataset dataset:

Before submitting x-rays for analyze step, you must carefully prepare the images to achieve the best result. The very first step was to analyze the contents of the pictures. It turned out that many images are not suitable for training on

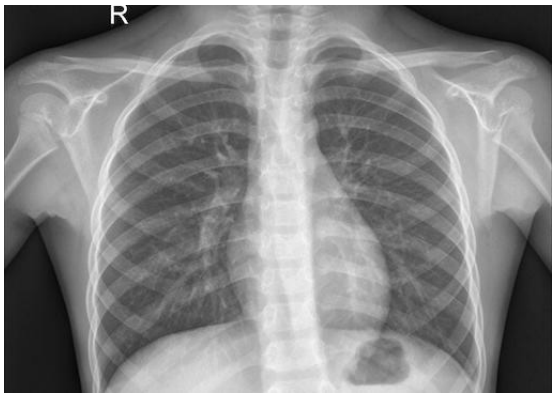

Figure 6.1: Healthy image

Figure 6.2: Pathological image

them neural networks. In this connection, it was decided to manually delete the "bad" images. The picture was deemed "bad" in the following cases:

- The picture was not taken in direct projection
- The position of the chest in the picture is too distorted
- The picture contains a lot of foreign objects in the chest area, such as: wires from medical devices, pins, buttons, pacemakers and so on
- Image too blurry or overshot

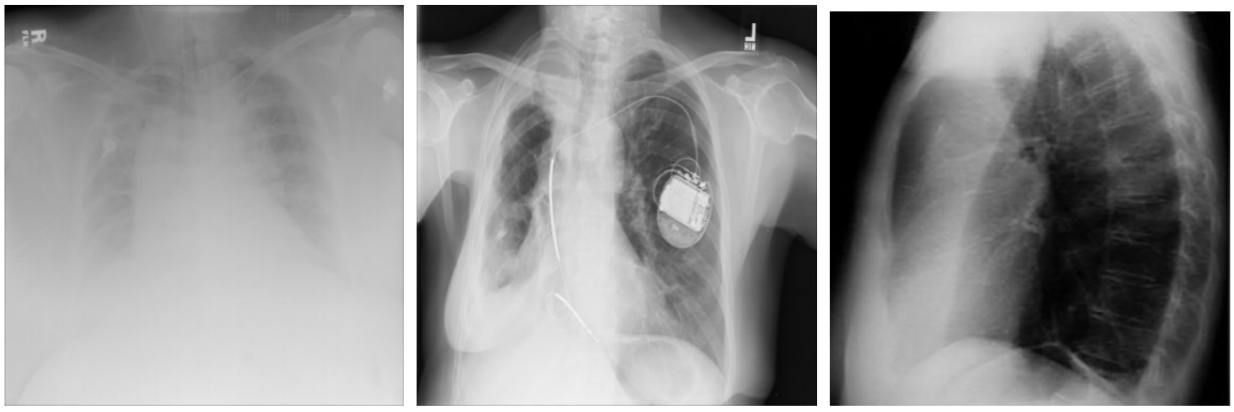

Figure 6.3: An example of "bad" images

As a result, a significant portion of the images were eliminated. Despite this, the number of remaining images fully satisfies the needs of the neural network. At the result we had 1350 "Normal" and 3884 "Pneumania" lung images.

Next, you need to compress the images, because the higher the resolution of the image, the slower the training. All images were resized to a resolution of 512x512 pixels.

In addition to resizing, a very useful technique is to increase the contrast of individual parts of the image (LCE) and segment lung image boundaries. This techniques allows you to improve the quality of images, making the outlines of the corresponding structures (lungs, ribs, hearts, etc.) more clear and distinguishable, which, in turn, improves the accuracy of other algorithms. A description of this methods can be found in the previous chapters.

6.2 Approaches

For several years, competitions have been held to find anomalies in computed tomography of the lungs, such as Luna16 and Date Scene Bowl 2017.

The use of computed tomography is a more accurate method than fluorography, but much more complex, which complicates its use in large quantities.

6.2.1 CheXNet

After the publication of the NIH Chest X Ray dataset [13], work appeared on the application of machine learning methods to solve the problem of finding pathologies specifically on fluorography of the lungs. Several groups conducted work on the search for anomalies. One of the most successful work is CheXNet [14] neural network from Stanford ML Group.

The ChestXNet neural network receives a picture as an input, consists of 121 convolutional layers similar in structure to the VGG16 neural network [7] and, based on the results of its work, gives the probability of pathology in the picture.

The study conducted several series of training. In one series, the result was compared with the expert assessment of several doctors, and in the second, they were compared with related works.

As part of the first training, pathologies were divided into 4 groups, the dataset was divided in the proportions of 93% for training, 6% for validation, and the remaining 1% for testing, which corresponds to 28744, 1672 and 389 patients or 98637, 6351 and 420 images. An F1-measure was used to calculate the accuracy of the model, and an accuracy of 0.435 was obtained.

As part of the second training, pathologies were divided into 14 groups, and the sample in the proportions of 70%, 10% and 20% was divided into training, validation and testing, respectively. For assessment, BCE and MSE was calculated and an average accuracy of 0.8 for different types of pathologies was obtained.

For a clear interpretation, maps of singular points were extracted, which were displayed as the last convolutional layer. Then, the special points most characteristic for the classification of pathologies were extracted, scaled to the size of the original image, and superimposed on it.

6.2.2 Autoencoders

The classical method of classification using machine learning is to mark up the dataset and use a neural network that can return the probability of belonging to a particular class. In the field of medical data processing, this approach requires a lot of work of medical experts at the stage of data preparation, and interpretation of the results of the classifier requires additional work.

An alternative approach is to use generative neural networks to train them to generate images without pathologies. To identify pathologies, the input image is processed by a neural network, the output is an image without pathology, and pathologies are searched as the difference between the initial and output image.

Applying this approach requires different data preparation from the first. In contrast to the first approach, not images with pathologies are required, on which areas with pathologies are clearly marked, but rather healthy images without anomalies, which reduces the burden on doctors when marking up data. Also, the use of generative neural networks simplifies the interpretation of the result of the neural network. The result of the work of the neural network is a snapshot of the dimension of the input image, which is a visual interpretation of its work.

Autoencoder [15] is a combination of two neural networks (see Figure 6.4): encoding and decoding. An image is fed to one input, and it learns to display from the space of the original image into a hidden space. The second input is an entity from a hidden space, and it learns to map it to the space of the original image. Since the dimension of the hidden space is less than the space of the original image, the neural network selects only important signs. Thus, trained only on images without pathologies, the neural network will not be able to generate images with anomalies.

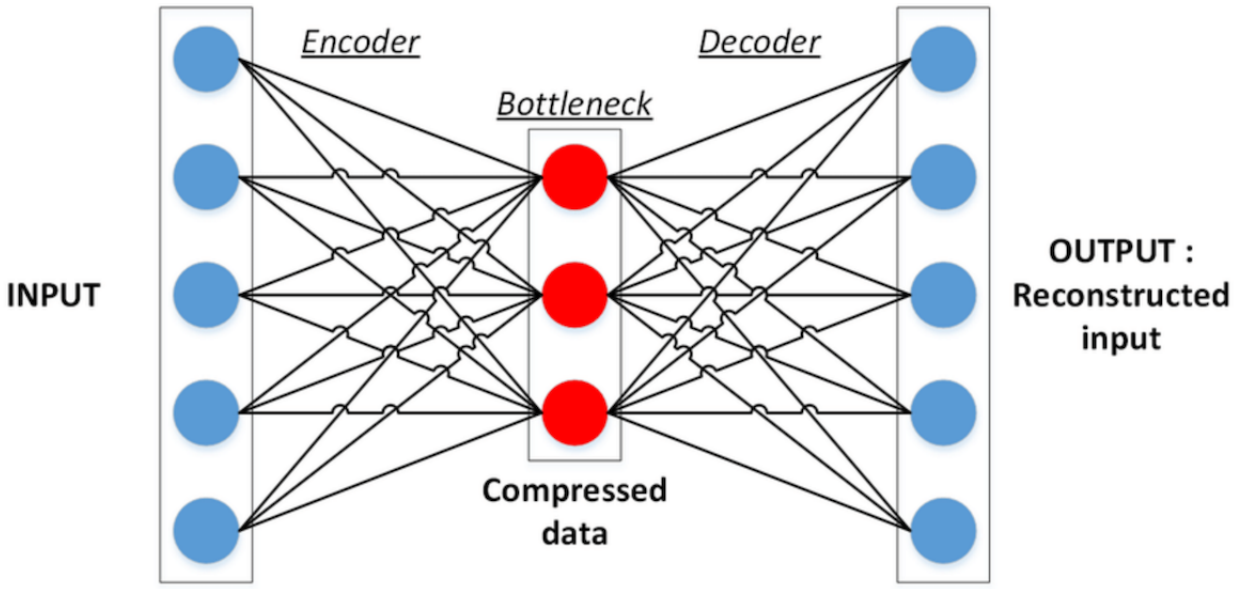

Figure 6.4: Autoencoder schema

6.2.3 Convolutional Neural Network

Convolutional neural networks sound like a strange combination of biology and mathematics with an admixture of informatics, but no matter how it sounds, these networks are one of the most influential innovations in the field of computer vision. Neural networks first attracted widespread attention in 2012, when Alex Krizhevsky won the ImageNet contest, lowering the classification error record from 26% to 15%, which was then a breakthrough.

CNN is a class of deep neural networks of direct distribution, successfully used in the recognition of visual images, including image recognition. The network operates in such a way that, taking an image, the network highlights some specific features of this image, after which it proceeds to obtain some more abstract details, and so on, up to highlighting high-level abstractions. At the same time, in the process of training, the network independently reconfigures and builds the most suitable a hierarchy of abstract features, neglecting unimportant details and highlighting the most significant. As a rule, a convolutional neural network consists of a sufficiently large number of different layers.

Here are the main types of layers that are usually included in CNN composition: convolutional layer, pooling layer, activation layer and fully connected layer [16] (See Figure 6.5).

One of the stages of this work is the classification of x-rays into two classes: healthy lungs and lungs containing any pathology. Any classification task is posed

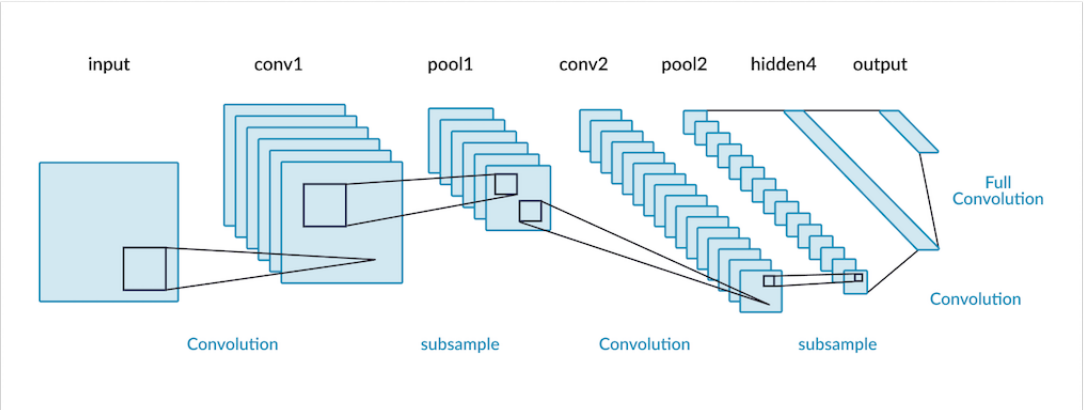

Figure 6.5: CNN Architecture example

as follows: there are many objects divided in some way into classes. A finite set of objects is given for which it is known which classes they belong to. This set is called the training set. The class affiliation of the remaining objects is not known. It is required to construct an algorithm capable of indicating the name of the class to which an arbitrary object from the original set belongs [16]. To verify the operation of the trained algorithm, one should also prepare a validation set. It is on this set that the accuracy of classification will be evaluated. Figure 6.6 shows the most common scheme containing the main stages of the operation of such algorithms:

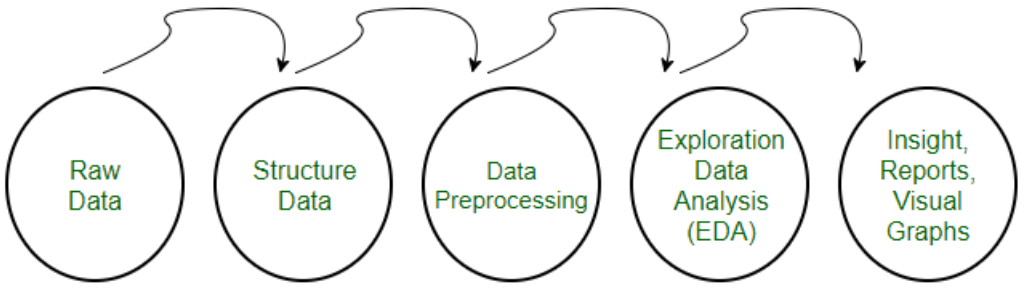

Figure 6.6: Common Machine Learning Stages

6.2.4 Classification algorithms

The intermediate latent representation of the image can be perceived as a vector of features that the autoencoder selected during processing. The resulting vectors can be classified by various machine learning methods.

Random forest

The first classification algorithm used is a random forest [17]. It consists in the use of an ensemble of decisive trees, giving out the probability that the image belongs to the class. The final decision is made by voting, in which all the trees participate.

ZFNet

In 2013, ZFNet was invented by Dr. Rob Fergus and his Ph.D. The student at that moment was Dr. Matthew D. Zeiler of New York University. That's why it is called ZFNet, which is based on their surnames Zeiler and Fergus, and in the 2014 ECCV document it is called "Visualization and Understanding of Convolutional Networks". ZFNet has significantly improved the image classification error rate compared to AlexNet [18], the winner in ILSVRC 2012. And Clarifai has only a slight improvement over ZFNet. However, when we talk about the ILSVLC 2013 winner deep learning network, we usually talk about ZFNet.

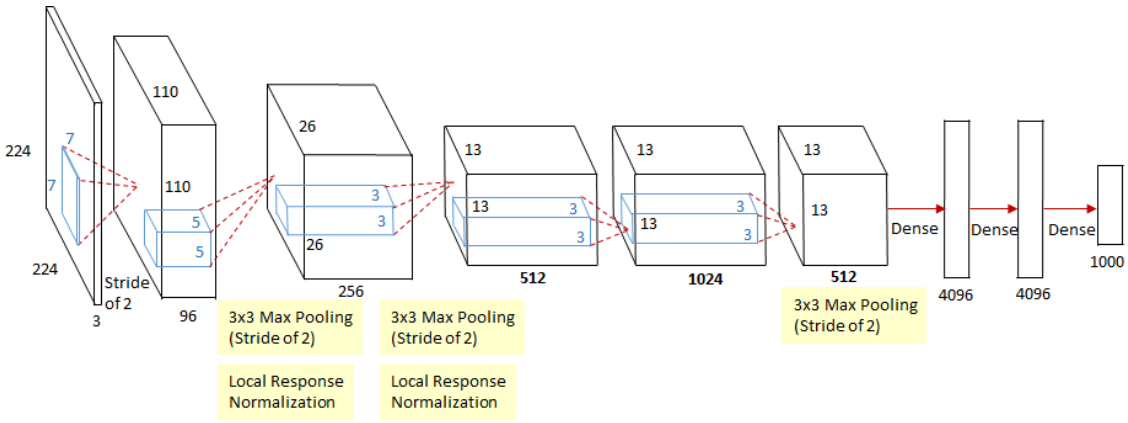

Figure 6.7: ZFNet architecture

The network contains eight layers: 5 convolutional layers and 3 fully connected layers. The structure of this neural network is shown in Figure 6.7. This neural network uses the ReLU activation function. The ReLU layer is located after each convolutional and fully connected layer.

6.3 Accuracy Measurement Method

The error of the autoencoder is calculated by Binary Cross Entropy [19] and AUC-ROC loss functions. These functions is the best suitable for binary tasks

like detecting whether the image is healthy or ill.

6.3.1 Binary Cross Entropy

Before considering the algebraic formulation of cross entropy, we will deal with the logic underlying it. Suppose that we have a probabilistic model, which is designed to interpret past events and predict the future. For each case in the past, the model assigns an estimate of the probability of an exact repetition of such a case. You can create a model that simply remembers all past cases and assigns them a probability of 1, but such a model will not allow you to learn anything about the future. Thus, an interesting model should in some way simplify past events and assign them a probability of less than 1.

General formula of Binary Cross Entropy function:

$$H_p(q) = -\frac{1}{N} * \sum_{i=1}^{\infty} y_i * \log\left(p(y_i)\right) + (1 + y_i) * \log\left(1 - p(y_i)\right)$$

Taking the Bayesian approach, we can assess the likelihood that the model creates all cases of the phenomenon. If in the future all cases are considered independent, then the probability that the model created an existing set of cases is the result of evaluating all the probabilities for each case in the past, performed using the model.

The mathematical product of thousands of variables, whose value is usually less than 0.5 (suppose we are working with a rather vague phenomenon), should be an incredibly small number. Despite the fact that assessing this value is a very difficult task, it is clear that the number will be extremely small.

Thus, logarithms are used to smooth out this problem, which is also known as "order loss". Logarithms can be used to translate works into sums, which allows you to effectively solve the problem of the disappearance of order.

From the 1990s to the beginning of the 2010s, most statisticians were convinced that the most effective way to optimize the indicator, from a purely mathematical point of view, was to create an optimization algorithm specifically for this indicator. However, deep learning specialists have found that this is not true. Numerical optimization is a very difficult task, and most indicators are not suitable for effective and large-scale numerical optimization. In addition, at the same time, data

processing specialists realized that all the tasks associated with forecasting were problems of numerical optimization.

6.3.2 Root Mean Squared Error

The MSE is the most common loss function. The MSE loss function is widely used in linear regression as an indicator of performance. To calculate the MSE, you need to take the difference between the predicted values and the true ones, square it and average over the entire data set.

To calculate the MSE, we simply take all the error values, calculate their squares of lengths and average them,

$$MSE = \frac{1}{m} * \sum_{i=1}^{\infty} ||\hat{y}^i - y^i||^2$$

where y^i is the actual expected result, and $\hat{y}^i$ is the forecast of the model.

6.4 Methods

In the framework of this work, two models have been developed Convolutional autoencoder model. The model was implemented by using Keras framework, high-level API based on TensorFlow.

The architecture of the neural network (see Figure 6.9) consists of six groups of layers arranged in series, including a convolutional layer with a core of size 7, a normalization layer to accelerate learning, and an activation layer of Leaky ReLU and ends with a sigmoid. The dimension of the hidden space is 4096.

The architecture of the decoding neural network is similar in structure and consists of four groups of layers, consisting of a reverse convolutional layer with a core of size 7, a normalization layer and an activation layer of ReLU and ends with a hyperbolic tangent.

To test the results of the autoencoder, it is enough to simply process a certain amount of X-ray images of the chest.

6.5 Results

For training of the neural networks, all images were randomly divided in the proportions of 85% and 15% into training and testing, which corresponds to 5234 and 624 images.

Our adopted architectures were trained on GPU (AMD Radeon Pro 5500M with 4GB of memory) and developed by using Keras framework, high-level API based on TensorFlow. The hyperparameters of the network:

- Batch size: 32
- Number of epochs: 13
- Learning rate: 1e-5
- Optimizer: Adam

The validation accuracy has increased to 95.76% for ZFNet model. Figure 6.8 shows the results of experiments in which we show the resulting segmentation using our network.

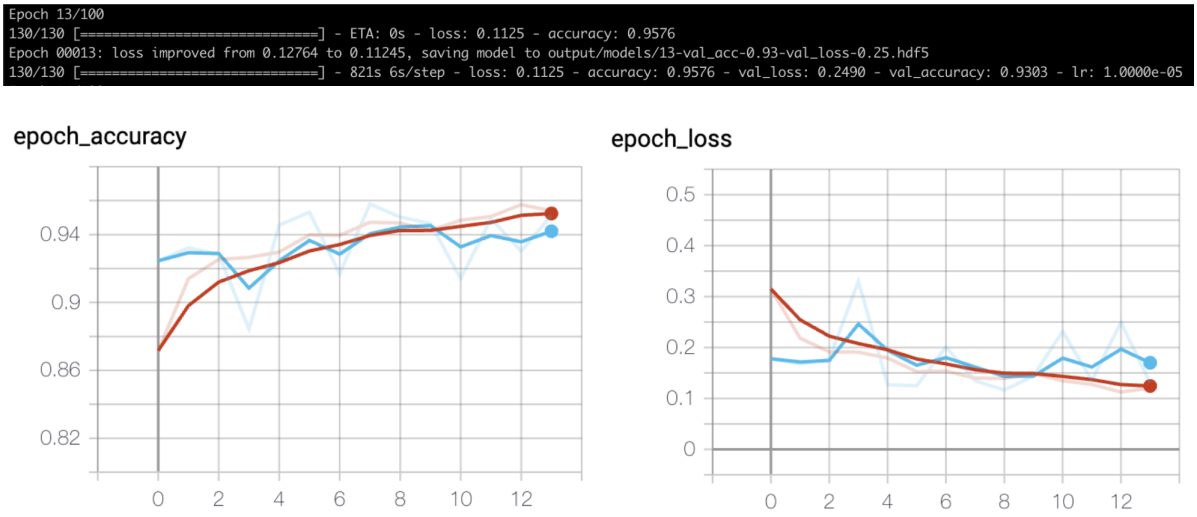

Figure 6.8: Autoencoder architecture training results

The validation accuracy has increased to 61.28% for Random Forest model. The same hyperparameters were used, but number of epochs. It took 37 epochs.

Also, after conducting all experiments, the main metrics for assessing the quality of classification were calculated. The results are shown in table 6.1.

Models	Binary Cross Entropy	Root Mean Square Error
CeXNet	0.84	0.82
ZFNet	0.96	0.93
Random Forest	0.61	0.65

Table 6.1: Models compare

The use of autoencoder neural networks, did not give very good results as traditional classification, deep convolutional neural networks, such as ZFNet and CheXNet. The comparison is shown in Table 6.1. The Random Forest architecture results could be negatively affected by the fact that the model was too simple for such a complecated problem. The adopted ZFNet model might show better results than CheXNet because of the less amount of images and epochs or a little overfitting.

Due to the simplisity and smaller size of the model, ZFNet has a simple structure and short training time. There was also no special configuration of the model, since the task of obtaining the highest possible accuracy and minimal losses in the classification was not posed in this paper. Based on these requirements, a standard convolutional neural network model based on the ZFNet architecture was selected for the application.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 512, 512, 1)]     0
_________________________________________________________________
conv2d (Conv2D)              (None, 253, 253, 96)      4800
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 127, 127, 96)      0
_________________________________________________________________
batch_normalization (BatchNo (None, 127, 127, 96)      384
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 31, 31, 256)       614656
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 16, 16, 256)       0
_________________________________________________________________
batch_normalization_1 (Batch (None, 16, 16, 256)       1024
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 14, 14, 512)       1180160
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 14, 14, 1024)      4719616
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 14, 14, 512)       4719104
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 7, 7, 512)         0
_________________________________________________________________
batch_normalization_2 (Batch (None, 7, 7, 512)         2048
_________________________________________________________________
flatten (Flatten)            (None, 25088)             0
_________________________________________________________________
dense (Dense)                (None, 4096)              102764544
_________________________________________________________________
dropout (Dropout)            (None, 4096)              0
_________________________________________________________________
dense_1 (Dense)              (None, 4096)              16781312
_________________________________________________________________
dropout_1 (Dropout)          (None, 4096)              0
_________________________________________________________________
dense_2 (Dense)              (None, 2)                 8194
=================================================================
Total params: 130,795,842
Trainable params: 130,794,114
Non-trainable params: 1,728
_________________________________________________________________
```

Figure 6.9: CNN autoencoder model

7. Discussion

Rather high values of the results indicate the suitability of the constructed classifier for predicting the selected diagnoses. Moreover, the drawback of the study is not taking into account the whole the data: the results of laboratory tests and signs such as gender and age of the patient. The weakness of the analysis is a small amount of real data, only existing online datasets. But the work analyzed data from different parts of the world, which makes it possible to omit geographical location, nationality and environmental conditions. This study solved the problem of predicting a diagnosis based on x-rays, namely: the presence of pulmonary disease. In our case, several options were chosen as the quality metric in the work: Binary Cross Entropy and Root Mean Square Error, which made it possible to evaluate the algorithms from different angles. Nevertheless, it is worth noting that to improve the results will require more data and a more complex algorithm that will be optimal for applied use.

8. Conclusion

So, we have completely automated the method for determining lung segments in x-ray images and analysis to predict whether the image is healthy or diseased. In the work, we compared several analysis approaches and compared to get the best option for use in the application. The final choice fell on the architecture of ZFNet, for its small occupied memory and sufficient speed in determining the answer.

References

[1] S. M. D. NASA, "X-rays", [Online]. Available: https://science.nasa.gov/ems/11_xrays.

[2] "Tuberculosis chest x-ray image data sets", [Online]. Available: https://lhncbc.nlm.nih.gov/publication/pub9931.

[3] C. W. e. a. Kermany D Goldbaum M, "Large dataset of labeled optical coherence tomography (oct) and chest x-ray images", [Online]. Available: http://dx.doi.org/10.17632/rscbjbr9sj.3.

[4] "Machine learning textbook", [Online]. Available: http://www.cs.cmu.edu/~tom/mlbook.html.

[5] P. Russell Stuart J.; Norvig, *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2010.

[6] T. B. Olaf Ronneberger Philipp Fischer, "U-net: Convolutional networks for biomedical image segmentation", [Online]. Available: https://arxiv.org/abs/1505.04597.

[7] I.H.Bankman, "Handbook of medical image processing and analysis", in. Academic Press, 2009, p. 985.

[8] "Tuberculosis chest x-ray image data sets, u.s. national library of medicine", [Online]. Available: https://lhncbc.nlm.nih.gov/publication/pub9931.

[9] S. O. G. E. Hinton and Y.-W. Teh, "A fast learning algorithm for deep belief nets", in. 2006, pp. 1527–1554.

[10] J. Schmidhuber, "Deep learning in neural networks: An overview", in. 2015, pp. 85–117.

[11] E. S. J. Long and T. Darrell, "Fully convolutional networks for semantic segmentation", in. 2015, pp. 3431–3440.

[12] P. F. O. Ronneberger and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", in. 2015, pp. 234–241.

[13] "Nih clinical center provides one of the largest publicly available chest x-ray datasets to scientific community", [Online]. Available: https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community, (accessed: 05.05.2020).

[14] J. I. e. a. Pranav Rajpurkar, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning", [Online]. Available: https://arxiv.org/abs/1711.05225.

[15] W.-C. C. e. a. Cheng-Yuan Liou, "Autoencoder for words", 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231214003658.

[16] C. A. Goodfellow I. Bengio Y., *Deep Learning*. MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org/.

[17] B. Leo, "Random forests", [Online]. Available: https://link.springer.com/article/10.1023%2FA%3A1010933404324.

[18] A. S. e. a. Krizhevsky, "Imagenet classification with deep convolutional neural networks", 2017. [Online]. Available: https://dl.acm.org/doi/10.1145/3065386.

[19] "Loss functions", [Online]. Available: https://ml-cheatsheet.readthedocs.io/en/latest/loss_functions.html.