

VIETNAM NATIONAL UNIVERSITY, HO CHI MINH CITY
HO CHI MINH CITY UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE AND
ENGINEERING

★ ★ ★



UNDERGRADUATE THESIS

BUILDING A SECURITY MODULE
FOR WEB FIREWALL

COMPUTER SCIENCE COMMITTEE

Supervisors: DR. NGUYEN AN KHUONG, PH.D.
MR. LE DINH THUAN, M.Sc.
MR. NGUYEN VAN HOA, B.ENG.
MR. LE NGUYEN MINH KHOI, B.ENG.
Examiner: THS. TRẦN GIANG SƠN

Students: TRINH CONG VU 1852882
LE THIEN THANH 1814009

Ho Chi Minh City, June 2023

COMMITMENT

We vouch for the fact that the work in this dissertation was completed in accordance with the guidelines established by the university and was not completed for submission to any other academic bodies. The works are our own, unless otherwise noted by a particular citation in the text.

HO CHI MINH CITY, NOVEMBER 2022

ACKNOWLEDGEMENT

First and foremost, we want to express our appreciation to our supervisors, professor Nguyen An Khuong and Le Dinh Thuan, for their patience, inspiration, and extensive expertise, all of which were crucial in assisting us with the thesis. Throughout the entire process of working and producing the thesis, their advice was helpful.

We would like to express our thankfulness to Nguyen Van Hoa and Le Nguyen Minh Khoi for their particular assistance and important information sharing. We cannot complete a fantastic thesis without their assistance.

In addition, we want to express our gratitude to our families and friends who have helped and supported us greatly during this thesis and our time at university.

THESIS GROUP MEMBERS

PREFACE

With the increased exchange of information and other activity on the World Wide Web, the Web has become the primary platform for attackers to cause havoc. Effective methods for detecting Web threats are crucial for ensuring Web security. With the explosion of data, it is becoming more challenging to manually detect and prevent cyber threats; however, the use of machine learning-based ways to fight cybersecurity is on the rise, with broad uses of machine learning in cybersecurity providers.

In this thesis, we propose a machine learning-based strategy to cyber-attack defense. Typically, rule-based WAFs have been widely employed. They do, however, have a significant false-positive rate. Typically, rule-based WAFs have been widely employed. They do, however, have a significant false-positive rate. As a consequence, we are building a machine learning-based module to validate requests that have been labeled suspicious by WAFs to enhance the WAFs and provide better surveillance.

The malicious request validator is based on the assumption that genuine requests to a website generally fit into the same category. The module uses a Convolutional Neural Network to classify the suspected request and evaluates whether it falls into the same category as the standard requests observed. The ultimate decision on whether to block a request is made using this result and combining it with the request content analyzer.

The false-positive rate (or the false alarm rate) is the most crucial requirement for modules in experimentation. Precision and accuracy are other key criteria. We also accounted for processing time with the suspicious request detector, as WAFs are supposed to be instant.

LIST OF FIGURES

2.1	How does WAF work.	8
2.2	Comparison between WAF and network firewall	10
2.3	CNN Architecture	12
2.4	Gradient descent demonstration	13
2.5	Learning rate	14
2.6	Loss function behaviour with different learning rate	14
4.1	Logo Viện nghiên cứu chống ung thư đường tiêu hoá ở Pháp.	20
4.2	Ảnh trực quan gói dữ liệu <i>3D-IRCADb-01</i>	20
4.3	Nhãn phân đoạn bị thiếu trên một số lớp ảnh ở bệnh nhân số 2.	22
4.4	Giá trị nhãn phân đoạn tĩnh mạch chủ trên bệnh nhân số 1	24
4.5	Giá trị nhãn phân đoạn tĩnh mạch chủ trên bệnh nhân số 3	24
5.1	Returned information	30
5.2	Compulsory query	30
5.3	Vulnerable PHP Code	31
5.4	The hacker's query and result	32
5.5	The code is protected against SQL injections	33
5.6	Malicious request validator architecture	37
5.7	Decision model for the combination of CNN and the Regression model . . .	37
6.1	Nội suy được sử dụng để lấp các lớp bị thiếu trong khối ảnh CT.	40
6.2	Mô phỏng quá trình trích xuất phần gan trong khối ảnh CT.	41
6.3	Phân phối mức sáng của các điểm ảnh thuộc nền cơ quan gan và mạch máu gan cùng ngưỡng giới hạn trái và phải.	42
6.4	Biến đổi cường độ sáng điểm ảnh để làm rõ mạch máu.	43
6.5	Làm giàu dữ liệu bằng phép trích xuất khối ngẫu nhiên.	43
6.6	Kiến trúc của bộ mã nguồn <i>Insight Deep Learning</i>	44
6.7	Trực quan hóa kết quả trên ứng dụng Slicer.	46
7.1	Kết quả hệ thống mạch máu của trường hợp tốt nhất trong thí nghiệm 2 .	54
7.2	Kết quả hệ thống mạch máu của trường hợp tốt nhất trong thí nghiệm 2 và nhãn phân đoạn.	55
7.3	Kết quả tìm đường chính giữa và điểm phân nhánh của trường hợp tốt nhất trong thí nghiệm 2.	56
7.4	Kết quả hệ thống mạch máu của trường hợp xấu nhất trong thí nghiệm 2.	57
7.5	Kết quả hệ thống mạch máu của trường hợp xấu nhất trong thí nghiệm 2 và nhãn phân đoạn.	58
7.6	Kết quả tìm đường chính giữa và điểm phân nhánh của trường hợp xấu nhất trong thí nghiệm 2.	59
7.7	Kết quả hệ thống mạch máu của trường hợp tốt nhất trong thí nghiệm 8 .	60
7.8	Kết quả hệ thống mạch máu của trường hợp tốt nhất trong thí nghiệm 8 và nhãn phân đoạn.	61

LIST OF TABLES

4.1	Tình trạng nhãn phân đoạn của cơ quan gan và hệ thống mạch máu.	23
4.2	Giá trị nhãn phân đoạn của cơ quan gan và hệ thống mạch máu.	25
5.1	Table 4.1	37
6.1	Thông tin phần cứng hệ thống máy tính.	45
7.1	Bảng phân chia các bệnh nhân thành các tập dữ liệu.	48
7.2	Thông số các thí nghiệm so sánh tính hiệu quả trước và sau điều chỉnh số tầng ở mô hình U-Net sử dụng convolution 3D.	50
7.3	Kết quả so sánh tính hiệu quả trước và sau điều chỉnh số tầng ở mô hình U-Net sử dụng convolution 3D.	50
7.4	Thông số các thí nghiệm tham khảo.	51
7.5	Kết quả các thí nghiệm tham khảo.	51
7.6	Kết quả thí nghiệm 2 của từng bệnh nhân trong tập kiểm tra.	51
7.7	Thông số các thí nghiệm đề xuất.	52
7.8	Kết quả các thí nghiệm đề xuất.	52
7.9	Kết quả thí nghiệm 8 của từng bệnh nhân trong tập kiểm tra.	52
7.10	Thông số các thí nghiệm có sự kết hợp của DenseNet và ResNet.	53
7.11	Kết quả các thí nghiệm DenseUNet và ResUNet.	53
A.1	Thông tin về tập dữ liệu <i>3D-IRCADb-01</i>	71

LIST OF ABBREVIATIONS

The following list shows the abbreviations that will be used in the text of this thesis.

BG	Background
FG	Foreground
GVLab	Graphics and Computer Vision Laboratory
HPCC	High Performance Computing Center
IoU	Intersection over Union
RAM	Random Access Memory
YAML	YAML Ain't Markup Language

TABLE OF CONTENTS

Preface	vii
List of Figures	ix
List of Tables	xi
List of Abbreviations	xiii
Chapter 1 INTRODUCTION	1
1.1 Motivation and problem statement	2
1.2 Objectives	4
1.3 Challenges	4
1.4 Tentative structure of the thesis	4
Chapter 2 BACKGROUNDS	7
2.1 Web Application Firewall	8
2.2 Machine Learning Model	11
2.3 Deep Neural Network	12
Chapter 3 LITERATURE REVIEW	17
Chapter 4 DATASET	19
4.1 Giới thiệu chung	20
4.2 Các vấn đề cần xử lý	21
Chapter 5 PROPOSED APPROACH	27
5.1 API security problems	28
5.2 Designs	34
Chapter 6 IMPLEMENTATION	39
6.1 Tiền xử lý dữ liệu	40
6.2 Xây dựng mã nguồn hệ thống	44
6.3 Đặc tả phần cứng	45
6.4 Hậu xử lý kết quả phân đoạn	45
6.5 Trực quan hóa kết quả thí nghiệm	46
Chapter 7 EXPERIMENTS	47
7.1 Chuẩn bị dữ liệu	48
7.2 Phương pháp đánh giá	48
7.3 Kết quả thí nghiệm	50

Chapter 8 CONCLUSION	67
8.1 Kết quả đạt được	68
8.2 Hạn chế và hướng phát triển	69
A THÔNG SỐ TẬP DỮ LIỆU 3D-IRCADB-01	71
B KẾ HOẠCH THỰC HIỆN LUẬN VĂN	73
References	75
List of Keywords	77

1

INTRODUCTION

In this chapter, we introduce issues in the field of cyber security, clarifies the necessity of building a firewall system. Next, we present an overview of the goals, challenges and structure of the thesis.

Table of Contents

1.1	Motivation and problem statement	2
1.2	Objectives	4
1.3	Challenges	4
1.4	Tentative structure of the thesis	4

1.1 Motivation and problem statement

Websites have become necessary for every business, brand, institution, organization, and individual. Web-based applications offer the general public and enterprises fast and simple services, they may be used for social media, email, banking, online shopping, education, or entertainment. Web applications are popular for a variety of reasons¹. For starters, web-based applications are convenient for users because they can be accessed from any location that has an internet connection. Second, a web application is a low-cost method for corporations because it does not require any special hardware or software and can be easily scaled up or down as needed. Furthermore, web-based apps frequently offer subscription-based models, which can be more cost-effective for businesses in the long run. Third, because web apps do not need to be downloaded or installed on a device, they are generally faster and more responsive than native apps. These are instead run on remote servers and accessed through a web browser. As a result, web apps can fully utilize the most recent advances in server-side technology, which helps in performance and speed. Why would someone want to attack web applications? Web applications are an attractive target for remote attackers because they can access the web app anywhere with an internet connection. Besides, they frequently handle sensitive information, such as login credentials and financial data. In short, web applications are targeted because they are visible, accessible, and offer a multitude of potential payoffs for an attacker. Cybercriminals can attack web applications for many reasons, including system flaws caused by incorrect coding, misconfigured web servers, application design flaws, or failure to validate forms. These flaws and vulnerabilities enable attackers to gain access to databases containing sensitive information.

The action conducted by attackers, which can do harm to the web application can be called cyber attacks. Cyber attacks are malicious attempts to access a person's or an organization's computer systems, networks, or data without authorization. The vulnerability to cyberattacks and our reliance on technology and connectivity are growing in tandem. Unlike viruses that would shut down a system for a few hours a few years ago, the consequences of cyber attacks today can include stolen data, destroyed networks, and thousands, if not millions, of dollars in recovery efforts. Cyber attacks can harm businesses, governments, and society². Among them, businesses and companies are more prone to being attacked by cybercriminals because they have more holes and gaps in their security that make them vulnerable to attacks. Cyber attacks affect a company's productivity, revenue, and reputation. Regarding productivity, nearly every business that suffers a cyberattack must suspend part or all of its operations until the attack is resolved, whether by paying a ransom, removing the malware from the device, network, or system, or restoring a backed-up version of its system. In terms of revenue, the costs of a cyberattack can wreck the economy of a company. The average cost of a data breach for a small to medium-sized business is massive, whether it has to shut down operations for several days, pay a ransom, lose data, replace devices, or pay a security expert to clean all malware out of the system or network. The most serious effect of a cyberattack is the loss of reputation. The most critical consequence of a cyberattack on a business is a loss of

¹ Amy Bros. Sysprobs. *Why are Web Applications Becoming Popular?*. July 2022. <https://www.sysprobs.com/web-applications-becoming-popular>

² CEI-The Digital Office. *The Consequences Of Cyber Attacks And Their Impact On Cybersecurity*. <https://www.copycei.com/consequences-of-cyber-attacks>

reputation. Consider the recent data breaches at Equifax, Target, and J.P. Morgan Chase, which led to the loss of customer data such as social security numbers, account details, and credit card numbers. Despite having the resources to recover, most businesses can not recover from security breaches because they lose their clients' trust and thus business. The biggest concern of a cyberattack on a government entity is the enormous volume of stolen data. This data could include everything from military and national security information to private data about civilians, which could be sold on the dark web and misused by terrorist groups. When cyberattacks happen, they badly affect practically every part of society, whether it's a large corporation or a small business. Consumers pay the price when enterprises, companies, and even nonprofit service providers like hospitals are forced to cover the costs of a cyberattack. There will be shortages that the customer will have to suffer when a company is restricted from providing its service as a result of cyber attacks or data breaches.

Let's take a look at Parachute's statistics¹. Healthcare, throughout the past 12 years, this sector has experienced the most costly data breaches, the costs have even increased by 41.6% from 2020 until 2022. At least 849 healthcare cybersecurity incidents and 571 data breaches were reported in 2022. The average financial loss due to data breaches in healthcare has skyrocketed from around USD 9 million to USD 10.10 million (2022). In the Finance industry, phishing attacks against banks and other financial institutions held the largest share, accounting for 23.2% of all cyberattacks targeting the financial sector. In the first quarter of 2022, ransomware assaults increased by 35% in the financial sectors. On average, financial organizations bore the second-highest data breach costs, at USD 5.97 million, just behind healthcare institutions (2022).

From the above situation, we can see that tools like Firewalls play a crucial role in organizations' security systems. They are a powerful tool for preventing malicious traffic from entering or leaving an organization's systems. Our group wanted to make some efforts to strengthen firewalls' cyber security, so we decided to choose this topic - Building a Security module for Web firewalls.

¹ Parachute. *Cyber Attack Statistics to Know in 2023.* <https://parachute.cloud/cyber-attack-statistics-data-and-trends/>

1.2 Objectives

Given the necessity of cybersecurity in this era, we chose to construct a firewall add-on. This thesis will concentrate on identifying malicious requests that pass through the firewall. We want to apply machine learning to assist the WAF with processing a tremendous amount of requests per second swiftly and efficiently. The developed module should be able to validate the request and identify whether or not it is an attack.

To eliminate the high FP, an inherent weakness of rule-based WAF, and assure the speed of the WAF, which must react nearly instantly to deliver a consistent customer experience, the module will use two fast and simple machine learning models. Each model will run independently to generate an output, then combined the result to achieve an accuracy of at least 95% as well as low latency of 5 milliseconds¹ for the WAF.

1.3 Challenges

In this topic, we have encountered some problems such as:

1.4 Tentative structure of the thesis

The content of the thesis proposal is demonstrated by these 8 following parts:

Chapter 1 introduces issues in the field of cyber security, and clarifies the necessity of building a firewall system. Next, we present an overview of the goals, challenges, and structure of the thesis.

Chapter 2 introduces the background knowledge of this thesis, including information about Web Application Firewalls, Machine Learning Models, and Deep Neural Networks.

Chapter 3 mentions some related studies on the use of machine techniques to detect malicious requests, their strengths and weaknesses, and evaluate an appropriate approach for this thesis.

Chapter 4 provides the dataset that would be used to train and evaluate the malicious request validator, its problems, and pre-processing details

Chapter 5 displays the problems as well as the malicious request validator's input and output. Then we offer the design and architecture of the problem's solutions.

Chapter 6 shows the implementation and deployment processes, including system frameworks and system specifications.

¹ Abdalslam. *Web Application Firewalls (WAF) Statistics, Trends And Facts 2023*. <https://abdalslam.com/web-application-firewalls-waf-statistics>

Chapter 7 discusses the division of the data set before training, assessment techniques, and experimental results. Compare the outcomes of the experiments we suggest to the reference experiments from relevant works.

Chapter 8 summarizes the results for the thesis until now. Finally, we want to present a working plan to improve the thesis.

2

BACKGROUNDS

This chapter introduces the background knowledge of this thesis, including information about Web Application Firewalls, Machine Learning Models and Deep Neural Networks

Table of Contents

2.1	Web Application Firewall	8
2.2	Machine Learning Model	11
2.3	Deep Neural Network	12

2.1 Web Application Firewall

2.1.1 What is WAF ?

WAF stands for **Web Application Firewall**. This firewall solution commonly monitors data packets and filters them for the presence of malware or viruses. It performs the data monitoring/filtering for to and from data packets.

The WAF tool can be distributed using network-based, cloud-based, or host-based architectures. It needs a reverse proxy to make sure that one or more web apps are in front of it while facing forward.

It can be utilized either alone or in conjunction with other applications. WAF may function at a lower level or a higher level depending on the requirement¹.

2.1.2 How does WAF work?

As previously stated, WAF is deployed at the application layer and serves as a two-way firewall. At work, WAF monitors HTTP or HTTPS traffic entering or exiting a specific web app. When WAF detects a malicious object in the traffic, it activates and destroys it. Figure 2.1 demonstrates how WAF works, normal users (top-left and bottom-left) are permitted access to the server with a WAF enabled, but attackers (middle-left) are prevented from doing so.

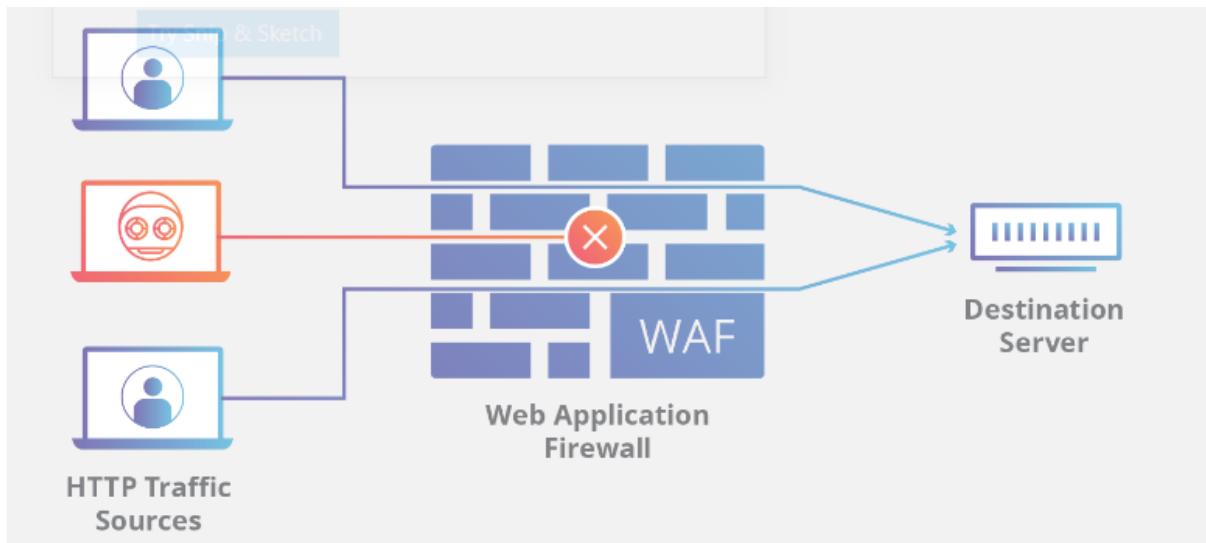


Figure 2.1: How does WAF work.

WAF predefined what is malicious and what is not to make the process easier. WAF adheres to these guidelines throughout the process. WAF primarily analyzes the GET and POST portions of HTTP traffic. GET retrieves data from the server, whereas POST directs data to the server to change its original state¹.

¹ Wallarm. *WAF Meaning*. <https://www.wallarm.com/what/waf-meaning>

2.1.3 WAF vs Firewall

In the modern era of sophisticated cyberattacks and digital innovation, it is essential for organizations to understand the threats they face and what their security measures protect them from. Understanding the value of and distinctions between WAF security and network firewall security is vital for preventing online attacks and other types of network intrusions².

A web application firewall (WAF) protects web applications by intercepting Hypertext Transfer Protocol (HTTP) traffic. This is distinct from a traditional firewall, which acts as a wall between external and internal network traffic.

A WAF stands between external users and web applications to track all HTTP traffic. It then identifies and stops harmful requests from entering users or apps on the web. As a result, WAFs protect key company online applications and servers from zero-day threats and other application-layer attacks. This becomes highly critical as firms invest in new digital efforts, which could expose new web apps and APIs to attacks.

A network firewall guards a secure local-area network against unwanted access to reduce the risk of assaults. Its goal is to distinguish a safe zone from a less secure zone and to control communication between the two. Without it, every device that has a public IP address is exposed to the outside network and potentially vulnerable to attack.

The layer of security that WAF and network-level firewalls operate on is the primary technological distinction between them. Attacks at OSI model Layer 7, or the application level, are protected by WAFs. This covers URL assaults, cookie manipulation, SQL injection, and attacks against JavaScript, ActiveX, and Ajax applications. They also target HTTP and HTTPS, the web application protocols that link web browsers and web servers. Network firewalls secure data transfer and network traffic at OSI model Layers 3 and 4. This covers assaults on the Domain Name System (DNS) and File Transfer Protocol (FTP), as well as Telnet, Secure Shell (SSH), and Simple Mail Transfer Protocol (SMTP). Figure 2.2 displays the attacks which can be blocked by network firewall and WAF.

² Fortinet. *WAF vs. Firewall: Web Application and Network Firewalls*. <https://www.fortinet.com/resources/cyberglossary/waf-vs-firewall>

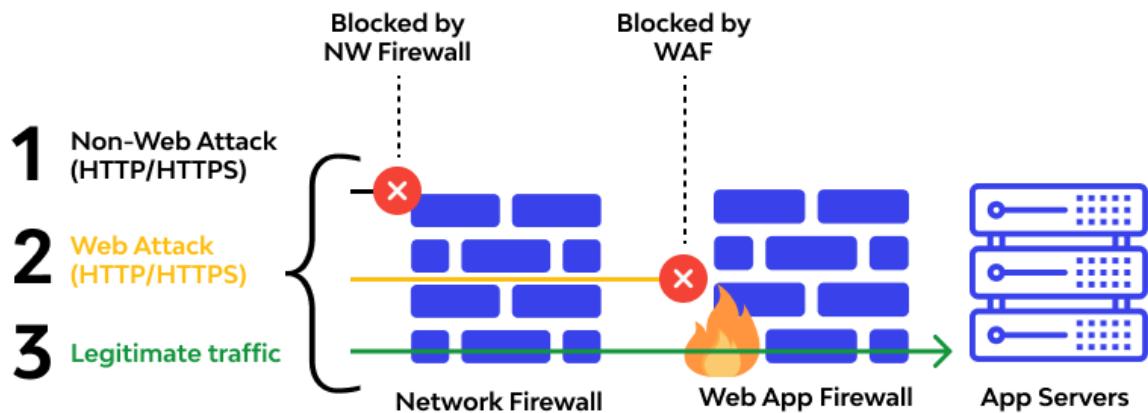


Figure 2.2: Comparison between WAF and network firewall

It's critical to pick a suitable network firewall or WAF to protect against all of the risks that may be present. Businesses cannot be protected from web page attacks by a network firewall alone; WAF capabilities are the sole means of defense. Business organizations risk leaving their larger network vulnerable to attack due to web application vulnerabilities without an application firewall. A WAF cannot protect from attacks at the network layer, so it should supplement a network firewall rather than replace it. Network-based and web-based solutions operate at several layers and protect from multiple types of traffic. As a result, they perform best together rather than against one another. A network firewall usually protects a wider range of traffic types, but a WAF deals with a specific threat that a conventional strategy cannot handle. Having both options is therefore advisable, especially if a company's operating systems and the web interact frequently.

2.2 Machine Learning Model

2.2.1 Logistic Regression

2.2.1.1 Logistic Regression Model

Predictive output of Linear Regression:

$$f(x) = w^T x \quad (2.1)$$

2.2.1.2 Sigmoid Function

Sigmoid function:

$$f(s) = \frac{1}{1 + e^{-s}} \triangleq \sigma(s) \quad (2.2)$$

2.2.1.3 Optimize loss function

The Stochastic Gradient Descent (SGD) algorithm will be used here.

$$\begin{aligned} \frac{\partial z}{z(1-z)} &= \partial s & (2.3) \\ \Leftrightarrow \left(\frac{1}{z} + \frac{1}{1-z}\right) \partial z &= \partial s \\ \Leftrightarrow \log(z) - \log(1-z) &= s \\ \Leftrightarrow \log \frac{z}{1-z} &= s \\ \Leftrightarrow \frac{z}{1-z} &= e^s \\ \Leftrightarrow z &= e^s(1-z) \\ \Leftrightarrow z &= \frac{e^s}{1+e^s} = \frac{1}{1+e^{-s}} = \sigma(s) \end{aligned}$$

2.2.1.4 Updated math formula for logistic sigmoid regression

$$\frac{\partial J(\mathbf{w}; \mathbf{x}_i, y_i)}{\partial \mathbf{w}} = (z_i - y_i) \mathbf{x}_i \quad (2.4)$$

The updated formula (according to the SGD algorithm) for logistic regression:

$$\mathbf{w} = \mathbf{w} + \eta(y_i - z_i) \mathbf{x}_i \quad (2.5)$$

2.3 Deep Neural Network

With their excellent results, broad applicability, and vast growth potential, neural networks are currently the most advanced advancement in artificial intelligence. Feedforward Neural Networks (FNN) and Convolution Neural Networks (CNN) are widely used to make predictions with independent data input. In CNN (Figure 2.3), each input image is passed through a series of convolutional layers (Filters, Pooling, and Fully-connected layers) to extract features before being classified using the Softmax function¹.

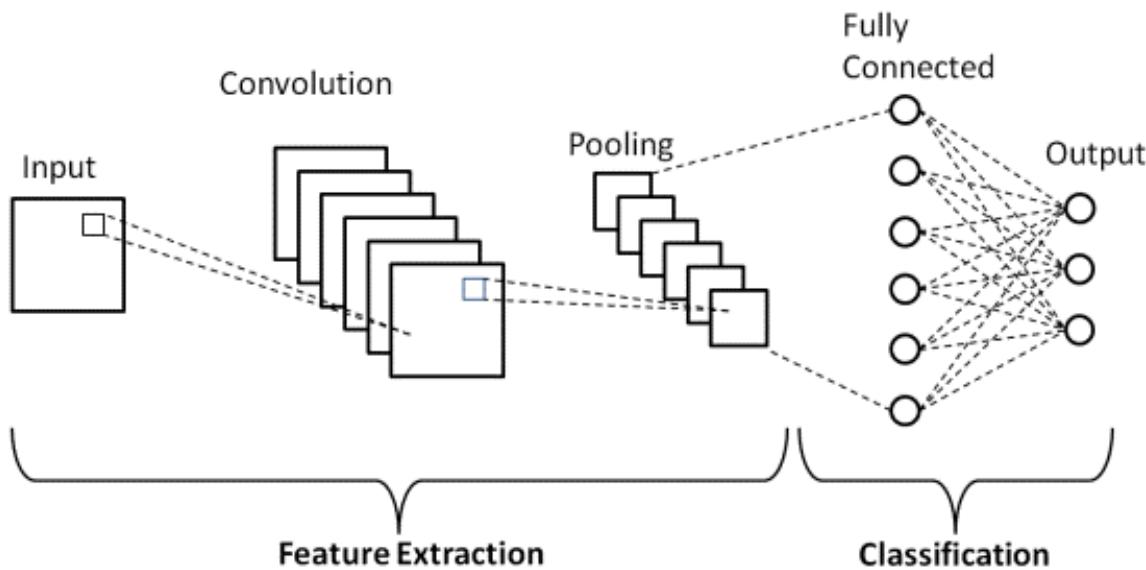


Figure 2.3: CNN Architecture

Layers are the building components of deep neural networks such as CNN and others. A *layer* is a broad phrase that refers to a group of 'nodes' that function together at a given level within a neural network. *Layers* are classified into three types: *input layer*, *hidden layers*, and *output layers*.

The *input layer* contains raw data (each variable as a 'node')

In neural networks, black magic happens in the *hidden layer(s)*. By minimizing an error/-cost function, each layer attempts to learn different elements of the data. The context of 'image recognition', such as a face, is the most intuitive way to understand these levels. The first layer may learn edge detection, the second eye, the third nose, and so on. This isn't exactly what's going on, but the idea is to split the problem down into components that different levels of abstraction can piece together, much to how our own brains work (hence the name neural networks).

The *output layer* is the simplest, usually consisting of a single output for classification problems. Even though it is a single 'node,' it is nevertheless regarded as a layer in a neural network because it might contain numerous nodes. We use six types of layers given by the TensorFlow framework in this thesis: **Embedding** layer, **Conv1D** layer, **MaxPooling1D** layer, **Dropout** layer, **Flatten** layer, and **Dense** layer.

¹ Softmax function: $\sigma(\vec{z})_i = (e^{z_i}) / \sum_{j=1}^K e_j^z$

2.3.1 Gradient descent

Gradient descent is an optimization algorithm that is used to minimize a function by iteratively traveling in the direction of the steepest descent as defined by the gradient's negative. Gradient descent is used in machine learning to update the parameters of our model, especially weights in neural networks.

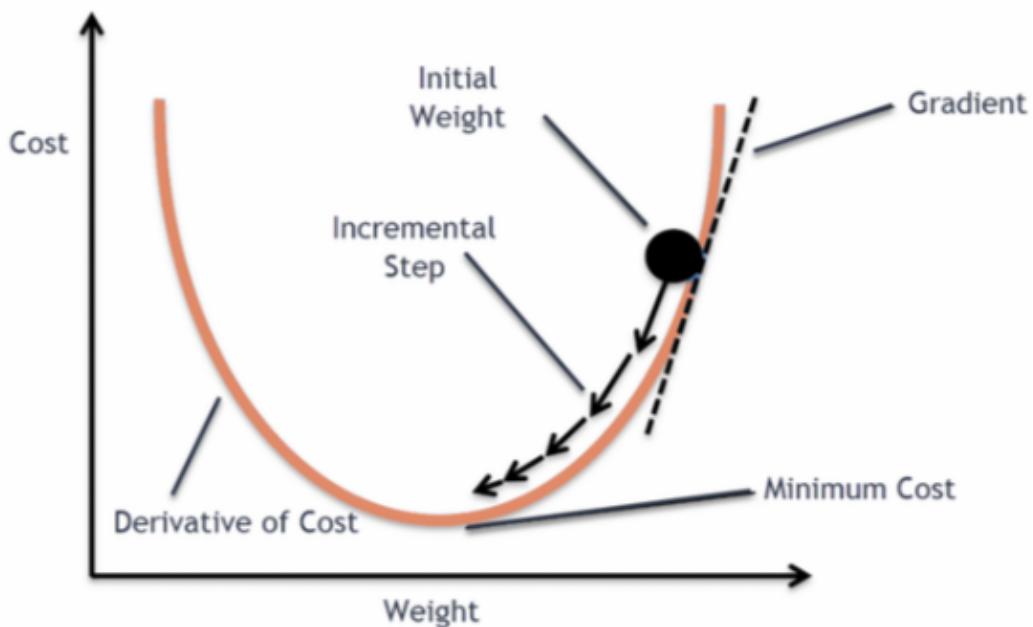


Figure 2.4: Gradient descent demonstration

2.3.2 Learning Rate

The learning rate is the size of each step in each gradient descent cycle. We can cover more territory per step with a high learning rate, but we risk overshooting the lowest spot because the slope of the hill is continually changing. We may reliably go in the direction of the negative gradient with a very low learning rate because we are recalculating it so frequently. A low learning rate is more exact, but calculating the gradient takes time, so we will take a long time to reach the bottom. An example of the learning rate is shown in Figure 2.5.

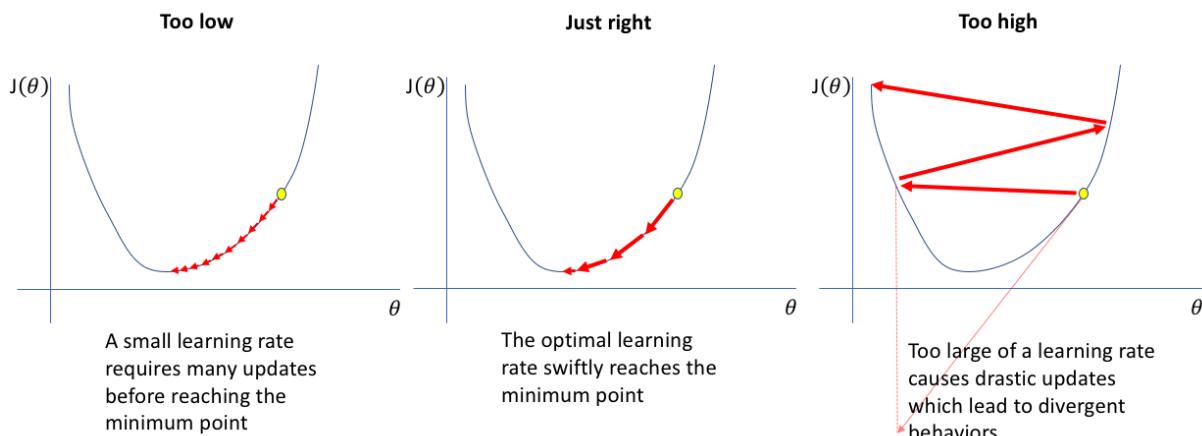


Figure 2.5: Learning rate

2.3.3 Loss Function

A Loss Function (or cost function) indicates how well the model predicts a given set of parameters. Index of the cost function The loss function has its curve and gradients. The slope of this curve indicates how we should adjust our parameters to improve the model's accuracy. If the cost ever rises, we must reduce the value of the learning rate; if the cost falls slowly, we must increase the value of the learning rate. Figure 2.6 shows examples of loss function behavior based on the learning rate.

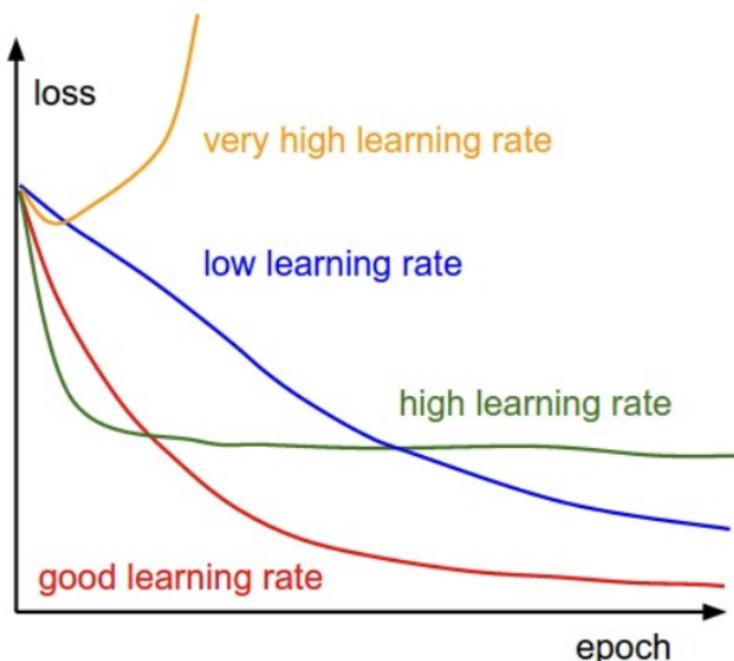


Figure 2.6: Loss function behaviour with different learning rate

2.3.4 Gradient descent optimizer

A method that computes adaptive learning rates for each parameter is Adaptive Moment Estimation (Adam). Adam, like Adadelta and RMSprop, preserves an exponentially decaying average of past squared gradients v_t in addition to an exponentially decaying average of past gradients m_t . Whereas momentum can be thought of as a ball rolling down a hill, Adam behaves more like a heavy ball with friction, preferring flat minima on the error surface. The decaying averages of past and past squared gradients, m_t and v_t , are computed as follows:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (2.6)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (2.7)$$

m_t and v_t are estimates of the first moment (the mean) and the second moment (the uncentered variance) of the gradients respectively, hence the name of the method. As m_t and v_t are initialized as vectors of 0's, the authors of Adam observe that they are biased towards zero, especially during the initial time steps, especially when the decay rates are small.

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (2.8)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (2.9)$$

They then use these to update the parameters:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t \quad (2.10)$$

3

LITERATURE REVIEW

This chapter mentions some related studies on the use of machine techniques to detect malicious requests, their strengths and weaknesses, and evaluate an appropriate approach for this thesis.

Table of Contents

Malicious request detection has been a focus of recent studies, with several malicious request detection systems developed. They propose solutions for particular or different types of harmful requests. For example, reference [1] show us their solution to SQLI attack which is one of the most common malicious requests. Joshi and Geetha [1] detect SQLI attacks using the Naive Bayes machine learning algorithm. Or [2] automatically generates malicious requests containing SQL injection to exploit flaws and take over of database servers. Reference [3] train a Support Vector Machine(SVM) model to detect SQLI attacks.

Fawaz Mereani and Jacob Howe et al[4] have demonstrated that SVM, k-NN, and Random Forest can be used to build classifiers for XSS coded in JavaScript giving high accuracy (up to 99.75%) and precision (up to 99.88%) when applied to a large real-world dataset. One especially intriguing component of this study is that, unlike other studies, a binary measure was employed for all features. This has resulted in greater accuracy and precision than previous tests employing weighted measurements.

Laughter et al [5] integrated the HTTP request features in the process of visiting the website into the detection feature set. By extracting the content of each field in the request header and request body, classification methods such as decision trees and SVM were used to complete the research.

Alshammari, Amirah and Aldribi, Abdulaziz et al [6] present a reliable model running in Real-time to detect malicious data flow traffic on the cloud depending on the ML supervised techniques based on the ISOT-CID dataset that contains network traffic data features. When tested using cross-validation and split-validation, DTREE and Random Forest both produced the best accuracy results. Their two models did not fail in any of the classification processes used to assess different portions or folds of the dataset.

This group of researchers [7] proposed an approach to detect malicious URLs using ensemble learning. They use TF-IDF to pre-process input URLs, then applied Rain Forest Ensemble-Based for prediction and an artificial neural network (ANN) classifier was constructed for decision making. Results show that this approach significantly improved the detection performance, achieving 96.80% compared with the best 90.4% achieved by the URL-based features. The false-positive rate was significantly decreased to 3.1% compared with 12% performed by the URL-based model.

Khoi. Le et al [8] suggested an approach of extracting WAF rules and trained a machine learning with the decision model independent from the rules themselves. This makes the model more self-reliant and the overall result more neutral. Our module is inspired by this approach.

4

DATASET

This chapter introduces the data set used to evaluate the system. We then discuss the issues that must be addressed before this dataset can be used.

Table of Contents

4.1	Giới thiệu chung	20
4.2	Các vấn đề cần xử lý	21

4.1 Giới thiệu chung



Figure 4.1: Logo Viện nghiên cứu chống ung thư đường tiêu hoá ở Pháp.
 (Source: <https://www.ircad.fr/>.)

3D-IRCADb (3D Image Reconstruction for Comparison of Algorithm Database) là tập dữ liệu do Viện nghiên cứu chống ung thư đường tiêu hoá *IRCAD* cung cấp và là tập dữ liệu đáp ứng tốt nhất cho bài toán trong đề tài của chúng tôi. Tập dữ liệu này cung cấp một số bộ ảnh y khoa từ các bệnh nhân được ẩn danh. Với mỗi bộ ảnh, các cơ quan, cấu trúc khác nhau được phân đoạn thủ công bởi các chuyên gia lâm sàng.

3D-IRCADb có hai gói dữ liệu là *3D-IRCADb-01* và *3D-IRCADb-02*. Chúng tôi chọn sử dụng gói *3D-IRCADb-01* [9] để đánh giá hệ thống trong luận văn này. *3D-IRCADb-01* bao gồm ảnh CT của 20 bệnh nhân (10 nam và 10 nữ) với sự xuất hiện của khối u gan trong 75% các trường hợp. Hệ thống mạch máu trong các bộ ảnh CT được chia thành ba phần là tĩnh mạch chủ, tĩnh mạch cửa và động mạch và được phân đoạn riêng cho từng loại mạch máu. Figure 4.2 là ảnh trực quan dữ liệu cho từng bệnh nhân trong gói *3D-IRCADb-01*. Các thông số chi tiết về gói dữ liệu này được chúng tôi trình bày trong Appendix A.

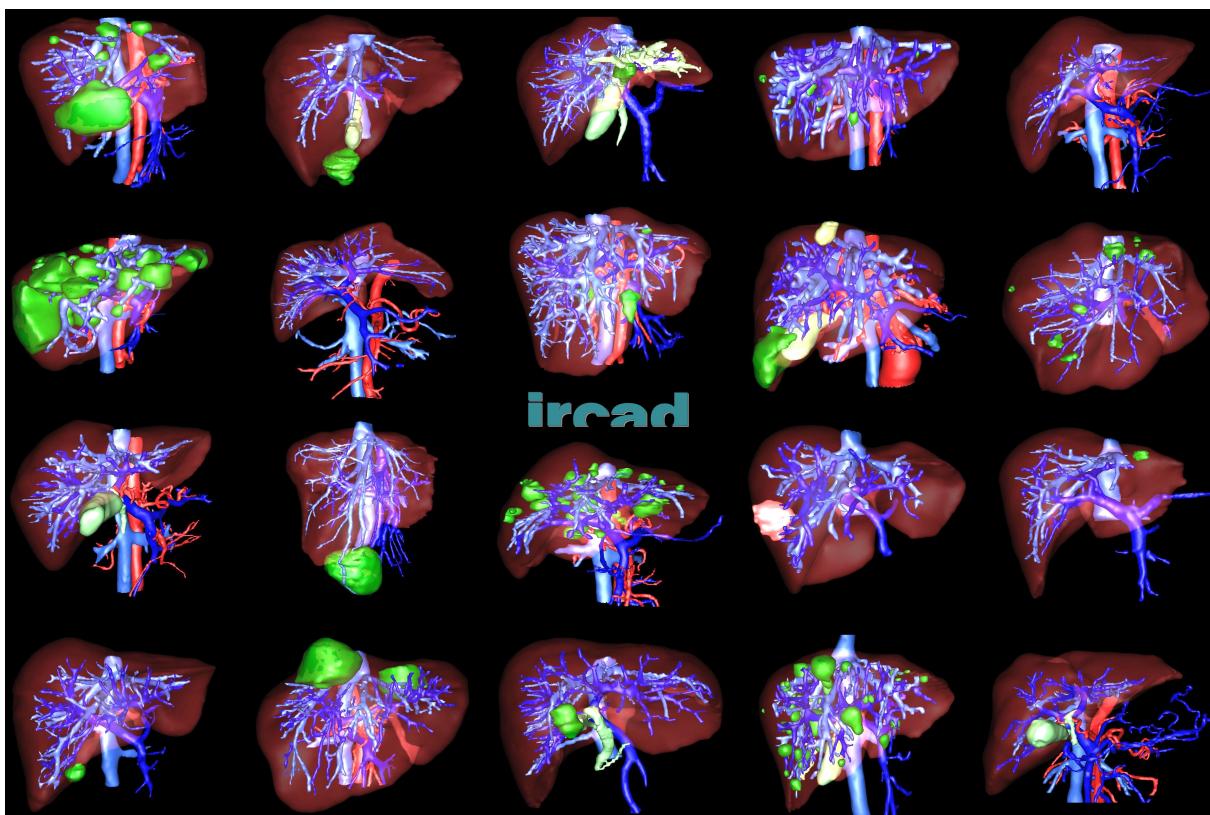


Figure 4.2: Ảnh trực quan gói dữ liệu *3D-IRCADb-01* (Source: [9]).

4.2 Các vấn đề cần xử lý

Trong mục này, chúng tôi trình bày các vấn đề gặp phải đối với tập dữ liệu cần được chú ý và (hoặc) xử lý trước khi có thể sử dụng bao gồm đơn vị hounsfield, nhãn phân đoạn không đầy đủ và giá trị nhãn không nhất quán.

4.2.1 Đơn vị Hounsfield

Nếu trong khoa học máy tính giá trị của một điểm ảnh được đo bằng cường độ sáng (intensity), thì trong y khoa giá trị của một điểm ảnh trên ảnh CT được đo bằng đơn vị hounsfield (xem [10]). Cách tính giá trị hounsfield như sau

$$HU = m * P + b, \quad (4.1)$$

trong đó, HU là giá trị hounsfield, P là giá trị số của điểm ảnh, m là giá trị tại trường (0028,1053) “Rescale slope” và b là giá trị tại trường (0028,1052) “Rescale intercept” được lưu trong file DICOM¹. Đối với bộ dữ liệu *3D-IRCADb-01*, m và b đều có giá trị là 1 nên ta có thể sử dụng trực tiếp không phải thực hiện bước tính giá trị hounsfield.

4.2.2 Nhãn phân đoạn không đầy đủ

Trong đề tài này, chúng tôi quan tâm các thành phần trong tập dữ liệu liên quan đến cơ quan gan và hệ thống mạch máu. Tuy nhiên, sau khi khảo sát tình trạng của tập dữ liệu, chúng tôi phát hiện tập dữ liệu có hai vấn đề về tính đầy đủ của nhãn phân đoạn.

Thứ nhất, nhãn phân đoạn không đầy đủ cho các loại mạch máu. Khi tìm kiếm các tệp tin liên quan đến hệ thống mạch máu trong tập dữ liệu, chúng tôi phát hiện có rất nhiều bệnh nhân không có nhãn phân đoạn cho động mạch.

Thứ hai, nhãn phân đoạn không đầy đủ trên tất cả các lớp ảnh. Figure 4.3 là ví dụ về việc gán nhãn không đầy đủ trên các lớp ảnh cho tĩnh mạch ở bệnh nhân số 2. Figure 4.3a và Figure 4.3b lần lượt là lớp ảnh CT thứ 74 chứa tĩnh mạch và nhãn phân đoạn tương ứng bị gán thiếu. Trong khi Figure 4.3c là lớp ảnh CT liền kề với nhãn phân đoạn tĩnh mạch đầy đủ như Figure 4.3d.

Tables 4.1 tổng hợp tình trạng nhãn phân đoạn của cơ quan gan và hệ thống mạch máu bao gồm tĩnh mạch chủ, tĩnh mạch cửa và động mạch. Ở mỗi đối tượng, bảng cho biết nhãn phân đoạn tương ứng có được cung cấp hay không và những lớp ảnh nào trong tổng số các lớp ảnh được gán nhãn.

¹ DICOM là định dạng chuyên dụng cho ảnh CT.

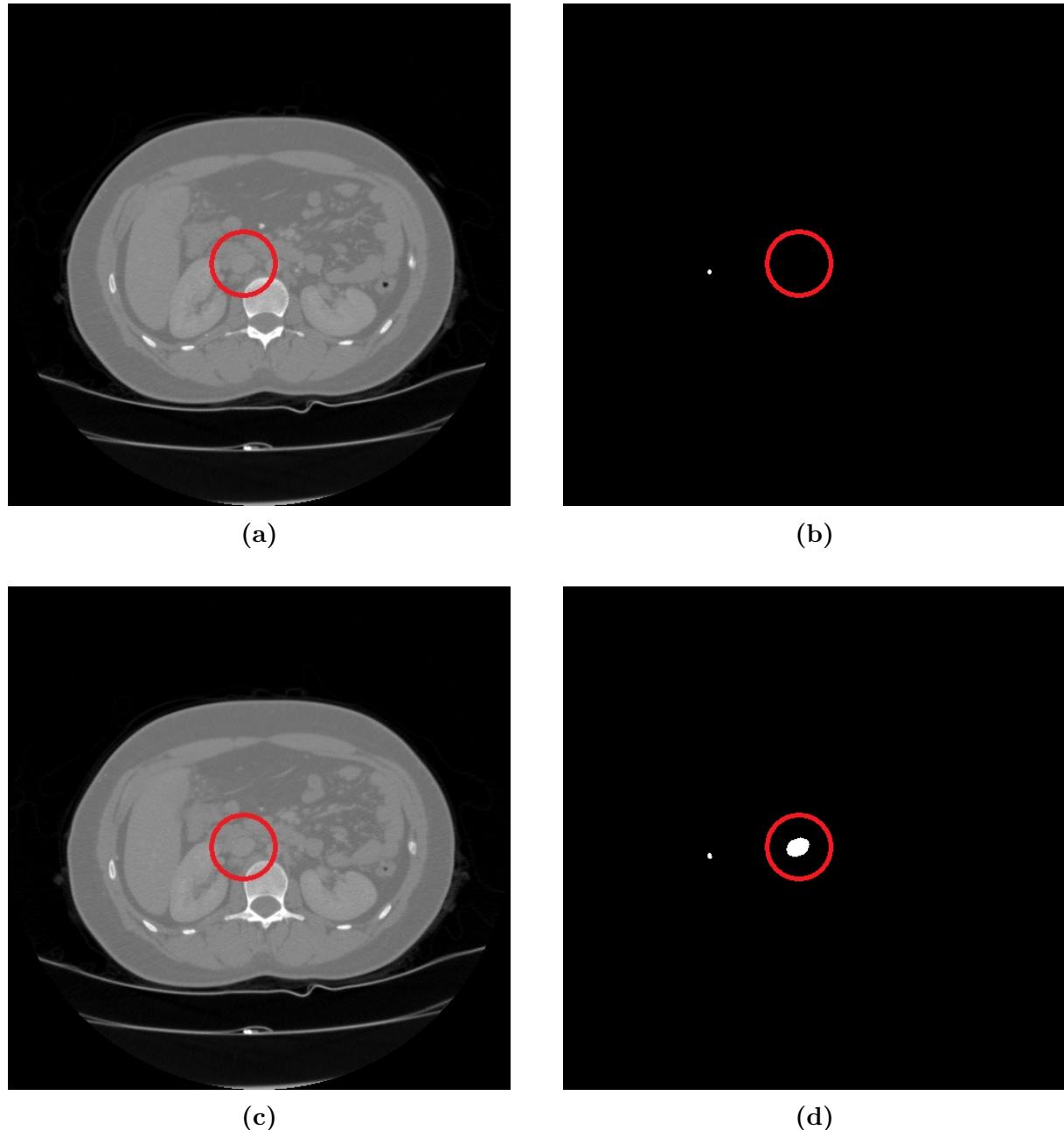


Figure 4.3: Nhãn phân đoạn bị thiếu trên một số lớp ảnh ở bệnh nhân số 2. (a) ảnh CT lớp thứ 74 có tĩnh mạch chủ. (b) nhãn phân đoạn cho ảnh CT lớp thứ 74 không được đánh nhãn tĩnh mạch chủ. (c) ảnh CT lớp thứ 75 có tĩnh mạch chủ. (d) nhãn phân đoạn cho ảnh CT lớp thứ 75 được đánh nhãn tĩnh mạch chủ.

Tables 4.1: Tình trạng nhãn phân đoạn của cơ quan gan và hệ thống mạch máu bao gồm tĩnh mạch chủ, tĩnh mạch cửa và động mạch. Kết quả khảo sát ở mỗi đối tượng cho biết nhãn phân đoạn của đối tượng đó có được cung cấp hay không và danh sách các lớp ảnh được gán nhãn.

STT	Số lớp ảnh	Gan		Tĩnh mạch chủ		Tĩnh mạch cửa		Động mạch	
		Tồn tại	Được gán nhãn	Tồn tại	Được gán nhãn	Tồn tại	Được gán nhãn	Tồn tại	Được gán nhãn
1	129	Có	Đủ	Có	Đủ	Có	Đủ	Có	Đủ
2	172	Có	Đủ	Có	75-152	Có	75-152	Không	—
3	200	Có	Đủ	Có	104-182	Có	104-182	Không	—
4	91	Có	Đủ	Có	Đủ	Có	Đủ	Có	Đủ
5	139	Có	Đủ	Có	Đủ	Có	Đủ	Có	Đủ
6	135	Có	Đủ	Có	Đủ	Có	Đủ	Có	Đủ
7	151	Có	Đủ	Có	Đủ	Có	Đủ	Có	Đủ
8	124	Có	Đủ	Có	Đủ	Có	Đủ	Có	Đủ
9	111	Có	Đủ	Có	Đủ	Có	Đủ	Có	Đủ
10	122	Có	Đủ	Có	49-113	Có	49-113	Không	—
11	132	Có	Đủ	Có	Đủ	Có	Đủ	Có	Đủ
12	260	Có	Đủ	Có	65-260	Có	65-260	Có	66-236
13	122	Có	Đủ	Có	1-117	Có	1-117	Có	Đủ
14	113	Có	Đủ	Có	51-113	Có	51-113	Không	—
15	125	Có	Đủ	Có	62-123	Có	62-123	Không	—
16	155	Có	Đủ	Có	32-143	Có	32-143	Không	—
17	119	Có	Đủ	Có	Đủ	Có	Đủ	Có	Đủ
18	74	Có	Đủ	Có	25-69	Có	25-69	Không	—
19	124	Có	Đủ	Có	Đủ	Có	Đủ	Không	—
20	225	Có	Đủ	Có	Đủ	Có	Đủ	Có	Đủ

4.2.3 Giá trị nhãn không nhất quán

Vấn đề tiếp theo của tập dữ liệu là giá trị nhãn phân đoạn không nhất quán. Đây là vấn đề nghiêm trọng mà nếu không xử lý trước khi sử dụng thì việc huấn luyện sẽ không thể thành công.

Trong hầu hết các nhãn phân đoạn có trong bộ dữ liệu, các giá trị background và foreground được gán số lần lượt là 0 và 255. Tuy nhiên, rất nhiều nhãn phân đoạn ở các cơ quan ở các bệnh nhân khác nhau được đánh giá trị khác nhau. Figure 4.4 là ví dụ việc đọc lên khối nhãn phân đoạn tĩnh mạch chủ của bệnh nhân số 1 và in ra các giá trị riêng biệt, kết quả là tập giá trị chứa 0 và 255. Thực hiện thao tác tương tự trên bệnh nhân số 3, ta có kết quả là tập giá trị chứa 0 và 1 (Figure 4.5).

```

1 path = "3Dircadb1/3Dircadb1.1/MASKS_DICOM/venoussystem/"
2
3 # Load all dicom files to numpy array.
4 volume = []
5 for i in range(len(os.listdir(path))):
6     dicom = pydicom.dcmread(path + "image_" + str(i))
7     volume.append(dicom.pixel_array)
8 volume = numpy.asarray(volume)
9
10 # Print unique value in volume.
11 numpy.unique(volume)

array([ 0, 255], dtype=uint8)

```

Figure 4.4: Giá trị nhãn phân đoạn tĩnh mạch chủ trên bệnh nhân số 1

```

1 path = "3Dircadb1/3Dircadb1.3/MASKS_DICOM/venoussystem/"
2
3 # Load all dicom files to numpy array.
4 volume = []
5 for i in range(len(os.listdir(path))):
6     dicom = pydicom.dcmread(path + "image_" + str(i))
7     volume.append(dicom.pixel_array)
8 volume = numpy.asarray(volume)
9
10 # Print unique value in volume.
11 numpy.unique(volume)

array([ 0, 1], dtype=uint8)

```

Figure 4.5: Giá trị nhãn phân đoạn tĩnh mạch chủ trên bệnh nhân số 3

Tables 4.2 tổng hợp các giá trị background và foreground cho từng nhãn phân đoạn ở các cơ quan gan, tĩnh mạch chủ, tĩnh mạch cửa và động mạch. Từ đây, chúng tôi tiến hành chuẩn hoá giá trị nhãn phân đoạn về chung một bộ giá trị với background và foreground lần lượt là 0 và 1.

Tables 4.2: Giá trị nhãn phân đoạn của cơ quan gan và hệ thống mạch máu.

STT	Gan		Tĩnh mạch chủ		Tĩnh mạch cửa		Động mạch	
	BG	FG	BG	FG	BG	FG	BG	FG
1	0	255	0	255	0	255	0	255
2	0	1	0	255	0	255	—	—
3	0	255	0	1	0	1	—	—
4	0	255	0	255	0	255	0	255
5	0	255	0	255	0	255	0	255
6	0	255	0	255	0	255	0	255
7	0	255	0	255	0	255	0	255
8	0	255	0	255	0	255	0	255
9	0	255	0	255	0	255	0	255
10	0	255	0	255	0	255	—	—
11	0	255	0	255	0	255	0	1, 255
12	0	1, 255	0	1	0	1	0	1
13	0	255	0	255	0	255	0	255
14	0	255	0	255	0	1, 255	—	—
15	0	255	0	255	0	255	—	—
16	0	255	0	255	0	255	—	—
17	0	255	0	255	0	255	0	255
18	0	1	0	255	0	255	—	—
19	0	255	0	255	0	255	—	—
20	0	1	0	1, 255	0	255	0	1, 255

5

PROPOSED APPROACH

This chapter displays the problems as well as the malicious request validator's input and output. Then we offer the design and architecture of the problem's solutions.

Table of Contents

5.1 API security problems	28
5.2 Designs	34

WAFs, as previously indicated, are frequently used, however they suffer from high FP. Our aim is to enhance the accuracy of WAFs, using the help of machine learning rather than the rule-based approaches.

Because WAFs only cover the Application Layer, the network requests are the model's input (mostly HTTP requests). Our methodology produces the same result as WAFs: whether the request is malicious or not.

With its nature, WAF is obligated to have fast processing speed in deciding whether an incoming request is reliable or not. For user experiences, we can't examine the request's veracity for minutes before granting or denying access. Our system's time constraint must be in milliseconds.

5.1 API security problems

Here are the most common vulnerabilities¹ :

- **Lack of rate limiting, DoS and brute force attacks on APIs**

- *Principle and functioning of DoS attacks*

An assault known as **a denial of service (DoS)** aims to render services unavailable. In fact, a DoS attack functions by depleting a resource that an API requires to respond to valid requests. By overloading an API with erroneous requests, its resources are only able to reply to the ones that were submitted.

The objective of DoS attacks is not to alter, delete or steal data. The aim is simply to damage the operation of a web service or the reputation of a company offering such services.

It is obvious that slowing down or even blocking their services for a few minutes could lead to significant financial losses and alter users' trust. It is therefore necessary to find solutions to protect against this, including request verification, traffic monitoring, the implementation of rules and rate limiting, etc. Similarly, during an API or web application pentest, DoS tests should be included to assess the resistance of the services to this type of attack.

- *Brute force attacks on APIs*

In a brute force attack, an attacker use tools to send a steady stream of requests to an application or API in an effort to try every combination of a parameter in the hopes of making the right guesses. The goals can vary: brute-forcing an authentication form to steal an account, brute-forcing a login to retrieve private information, etc.

This is a “trivial” attack method, easy to perform, but still very effective and widely used by attackers.

- *Implement rate limiting mechanisms to counter DoS and brute force attacks*

Securing APIs against DoS or brute force attacks requires the implementation

¹ Vaadata. *How to strengthen the security of your APIs to counter the most common attacks?*. April 2022. <https://www.vaadata.com/blog/how-to-strengthen-the-security-of-your-apis-to-counter-the-most-common-attacks/>

of rate limiting mechanisms. These mechanisms protect APIs and other services from excessive and abusive use, in order to ensure their availability.

Rate restriction works on a pretty straightforward principle. It entails making requests in advance of when one or more clients—systems—might utilize more than their "fair" share of a resource. Additionally, one can lessen the danger of DoS or brute force attacks by restricting the number of requests that a specific user is permitted to send in a given window of time.

For example, after a user has been authenticated, your API or application may apply quotas that restrict what they are allowed to do, including the limit of requests they can send. For example, you can limit each user to a certain number of API requests per hour, to prevent them from flooding the system with too many requests.

Similar to this, you can set limitations before authenticating a user to lower the overall number of requests, or just those coming from a specific IP address or time period. Therefore, if rate limitation is enabled, your API will monitor the volume of requests and reject those that are greater than the permitted threshold. Additionally, rules can be applied to totally shut off connections when the limit is reached or to sluggish down request processing. This action is referred to as "throttling."

In short, Rate Limiting prevents resource depletion by managing rules and quotas. There are different techniques for applying rate limiting, each with its own specificities: Token bucket, Leaky bucket, Fixed window and Sliding window.

- **Lack of user input validation and injection attacks**

- *Code injections*

Code injection is one of the most common types of injection attacks. If attackers know the programming language used by an application or API, they can inject code through text input fields to force the web server to execute the desired instructions.

- *SQL injections (SQLi)*

Injections represent a significant part of the vulnerabilities encountered in applications and APIs. The best known and most dangerous is SQL injection.

In an SQL injection attack, an attacker injects data to manipulate SQL commands, thereby interacting with the database through unintended queries. These flaws can lead to theft, deletion or manipulation of stored data. Worse still: if the rights are too permissive, this can even lead to a compromise of the server.

Let's look at this in more detail with a concrete example:

One can imagine an API endpoint that returns the information of a country based on its "CountryCode".

```
URL: http://localhost:8042/?CC=FR
```

```
[  
 {  
   "name": "France",  
   "code": "FR"  
 }  
]
```

Figure 5.1: Returned information

To perform the desired action, the query interacts with a database. Below is the SQL query that the server must run on the database.

```
SELECT name_fr as name , alpha2 as code FROM Country WHERE alpha2= "FR"
```

Figure 5.2: Compulsory query

Example of a vulnerable PHP code:

```
<?php

try{
    $db = new PDO('sqlite:database/base.sqlite3');
    $db->setAttribute(PDO::ATTR_DEFAULT_FETCH_MODE, PDO::FETCH_ASSOC);
    $db->setAttribute(PDO::ATTR_ERRMODE, PDO::ERRMODE_EXCEPTION); // 
ERRMODE_WARNING | ERRMODE_EXCEPTION | ERRMODE_SILENT
} catch(Exception $e) {
    echo "Unable to access the SQLite database:". $e->getMessage();
    die();
}

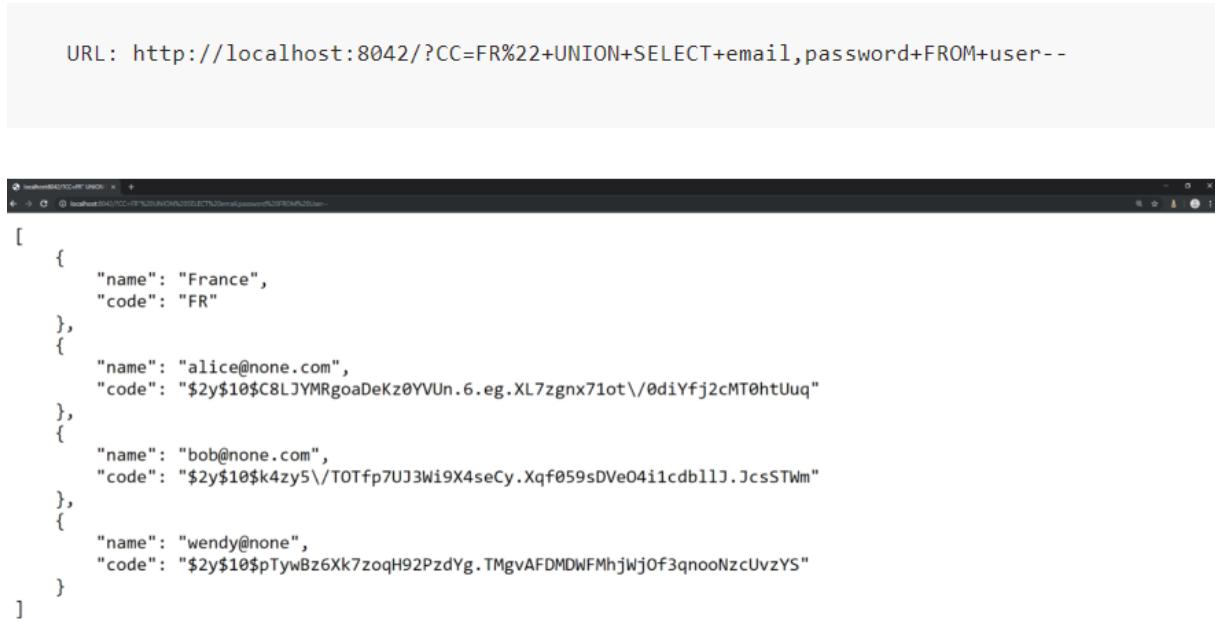
$recipesStatement = $db->prepare('SELECT name_fr as name ,alpha2 as code FROM
Country where alpha2 = ''.'.$_GET['CC'].'''');
$recipesStatement->execute();
$recipes = $recipesStatement->fetchAll();

header('Content-Type: application/json; charset=utf-8');
echo json_encode($recipes, JSON_PRETTY_PRINT);
```

Figure 5.3: Vulnerable PHP Code

As we can see, the CC parameter that is controlled by the user is directly concatenated to the query.

Now, let's assume that on this same database, there is a 'users' table that contains the email addresses and passwords of the users registered on the application. Let's look at what happens if an attacker makes the following query:



The screenshot shows a browser window with the URL `http://localhost:8042/?CC=FR%22+UNION+SELECT+email,password+FROM+user--`. The page displays a JSON array containing four objects. Each object has two properties: 'name' and 'code'. The first object represents France ('name': 'France', 'code': 'FR'). The subsequent three objects represent users ('name': 'alice@none.com', 'code': '...'), ('name': 'bob@none.com', 'code': '...'), and ('name': 'wendy@none', 'code': '...'). The 'code' values for the users are extremely long and complex, likely generated by a hashing function.

```
[
  {
    "name": "France",
    "code": "FR"
  },
  {
    "name": "alice@none.com",
    "code": "$2y$10$c8LjYMRgoaDeKz0YVUn.6,eg.XL7zgnx71ot/0diYfj2cMT0htUuq"
  },
  {
    "name": "bob@none.com",
    "code": "$2y$10$kt4zy5/0Tfp7UJ3Wi9X4seCy.Xqf059sDVe04i1cdb11J.JcsSTWm"
  },
  {
    "name": "wendy@none",
    "code": "$2y$10$pTywBz6Xk7zoqH92PzdYg.TMgvAFDMDFMhjWj0f3qnooNzcUvzYS"
  }
]
```

Figure 5.4: The hacker's query and result

In addition to the country name, this query retrieves all users and their password hashes.

This SQL injection flaw was discovered. A flaw that enables an attacker to "pervert" the SQL query the program generates. The attacker may be able to view or even change the database's data with this behavior. Evidently, if an API or online application penetration test were conducted, this significant vulnerability would be discovered and disclosed.

- *Validate user input to prevent injection attacks*

The strongest defense against SQL and code injections is validating user input. In theory, it should be recognized that data received by an application or API cannot be deemed "always" safe. In order to prevent such vulnerabilities, it is necessary to put in place methods to verify that user input meets the anticipated parameters.

The most effective method of protecting against SQL injections is the use of prepared statements, which separate the SQL commands from the data sent by a user.

The fix is to use prepared queries. On the PHP documentation we can see the following information:

- * The query can be run multiple times with the same or different parameters after only one analysis (or preparation). The database will parse, construct, and optimize its plan to execute the query once it is prepared. If you have to repeat the same query multiple times with various parameters, this procedure can be very time-consuming for sophisticated queries, which will slow down your apps. You can avoid repeating the cycle of analysis, compilation, and optimization by using prepared statements. In short, prepared queries execute more quickly and with fewer resources.
- * You don't need to surround query parameters in quotes; the driver takes care of it. If your application only employs prepared statements, you can

be certain that SQL injection is not a possibility (however, if you build other parts of the statement based on user input, you are still taking a risk).

Thus, the following code is correctly protected against SQL injections:

```
<?php

try{
    $db = new PDO('sqlite:database/base.sqlite3');
    $db->setAttribute(PDO::ATTR_DEFAULT_FETCH_MODE, PDO::FETCH_ASSOC);
    $db->setAttribute(PDO::ATTR_ERRMODE, PDO::ERRMODE_EXCEPTION); // 
    ERRMODE_WARNING | ERRMODE_EXCEPTION | ERRMODE_SILENT
} catch(Exception $e) {
    echo "Unable to access the SQLite database:". $e->getMessage();
    die();
}

$recipesStatement = $db->prepare('SELECT name_fr as name ,alpha2 as code FROM
Country where alpha2 = :CC');
$recipesStatement->execute([
    'CC' => $_GET['CC']
]);

$recipesStatement->execute();
$recipes = $recipesStatement->fetchAll();
header('Content-Type: application/json; charset=utf-8');
echo json_encode($recipes, JSON_PRETTY_PRINT);
```

Figure 5.5: The code is protected against SQL injections

As we can see, the difference with the vulnerable code is that the parameters coming from a user are no longer concatenated with the query, but directly provided at query execution. This also shows that a prepared statement can still be vulnerable to SQL injection if the data is concatenated. Because in the vulnerable example it was already a prepared statement that was used.

- **Lack of data encryption and Man In The Middle attacks**

- *Man In The Middle attacks*

An attack known as a "Man in the Middle" occurs when a hostile person intrudes on a communication or data transfer taking place between a client and a server, a server and a server, or a client and a client. Its goals might vary, from merely intercepting sensitive data—such as passwords, financial information, personal information, and sensitive documents—to manipulating communication in order to, for instance, implant malware.

This type of attack is possible if and only if the communications are not encrypted. It is therefore quite easy to protect against it.

- *Encrypting data with TLS to counter Man In The Middle attacks*

One of the most fundamental components of guaranteeing the security of an API or service is encryption. Indeed, the encryption protocol TLS (the successor to SSL) ensures safe communications over a computer network. Connections

between a client and a server that are TLS-secured include one or more of the following characteristics:

- * The connection is private (i.e. secure) because the data transmitted is encrypted.
- * The encryption keys are uniquely generated for each connection and are based on a shared secret negotiated at the beginning of the session.
- * The connection guarantees integrity because each transmitted message includes a signature verification of the integrity of the message, thus avoiding any undetected loss or alteration of the data during transmission.

The use of this encryption protocol therefore reduces or even eliminates the risks of Man In The Middle attacks. Furthermore, to reinforce security, we recommend implementing the HSTS (http Strict Transport Security) header on your servers in order to force a browser to use HTTPS secure connections. Without this setting, you run the risk of users accessing your domain without the HTTPS protocol, which can lead to a breach in communications.

5.2 Designs

To summary, numerous strategies for identifying malicious URLs have been proposed. Most of these solutions utilize supervised-based machine learning techniques for classification. This technique can detect irregularities between requests, but it cannot extract these abnormalities into human-readable form in order to reconstruct the WAF. Relying solely on machine learning eventhough brought higher accuracy like in [7], WAFs need to be time-efficient. We can not compromise accuracy for speed, hence these methods don't work for WAFs.

We come to the approach is to categorize the incoming request and analyze its structure, with moderate reliability, then combining the result of these two process to achieve the high precision yet does not expending an excessive amount of time

5.2.1 Architecture

5.2.1.1 Data exploration and Sanitization

Our goal is to classify request supplied as inputs in order to determine whether they are harmful or inoffensive. The sample of request data consists of different categories including:

- *Plain text*: request that contains data but does not trigger any machine execution, typically in the form of HTML, JSON, XML, and CSV, etc.
- *Client-side script*: request in form of programming languages or scripts that can be executed on client's machine

- *Server-side script*: request in the form of programming languages (mostly back-end programming languages like PHP, JAVA, PYTHON, and so on) that can change the behavior of the application or web
- *Shell script*: request in the form of shell scripts that can run jobs on the server or change the server's or operating system's behavior
- *SQL script*: request that in form of SQL, can be used to query data from the database.

The textual data acquired in the previous phase is going to be cleaned and standardized. To minimize feature complexity and improve classification performance, the request was preprocessed by eliminating symbols and punctuation. The text data collected will be transformed to lower case and then normalized. The goals of the normalization procedure were dual. First, the text must be converted from unstructured data to a structured word vector. Second, by deleting unneeded words and decreasing the number of words by roots them to their originals, we may reduce the scarcity of feature vectors. However, malicious attacker usually use methods like slight modification to make an imposter URL or request looks legitimate, we skip severals step in conventional Natural Language Processing:

- Removing stop word
- Stemming (the process of reducing infected words to their stem)
- Lemmatization (returning to the base form of the words...)

5.2.1.2 Model Selection and applying Natural language processing

Because four of the five categories are structured languages, we must first create a special tokenizer to map the request into a defined set of keywords (client-side script, server- side script, and SQL script). To convert the words (the tokens) to their numerical equivalents, a corpus containing a list of unique tokens based on their frequency of occurrence in each class was created. To put it in more formal mathematical terms, the TF-IDF score for the word t in the document d from the document set D is calculated as follows:

$$tf\ idf(t, d, D) = tf(t, d) \cdot idf(t, D) \quad (5.1)$$

Where:

$$tf(t, d) = 1 + \log(freq(t, d)) \quad (5.2)$$

$$idf(t, D) = \log\left(\frac{N}{count(d \in D : t \in d)}\right) \quad (5.3)$$

The statistical-based text representation, TF-IDF, was then computed using the following equation:

$$tf_idf = tf \cdot \log\left(\frac{N}{df}\right) \quad (5.4)$$

where tf is the term frequency of the word in a specific instance, df is the document frequency for the word, N is the number of samples in the dataset. The term frequency tf is the number of times a word has occurred in the sample while the inverse document frequency idf refers to the inverse number of documents where the word has occurred. The higher the tf_idf of a word in a document, the more relevant the document. The output of this phase was three numerical vectors for each sample.

5.2.1.3 Distinguishing request

We present a hypothesis: all normal server requests fall into the same category. A static web request, for example, may only contain plain text, whereas API server requests are mostly in JSON format, incoming database queries are SQL, and online compilers use programming language-format requests. A malicious request must be classified differently than normal requests, such as a script request to a static web server or API server, which can be classified as code injection ¹ or command injection ². SQL requests to an API server can be classified as SQLi, and script requests to a database can be classified as command injection or stored XSS attacks. We can determine whether a request is 'normal' to a server by comparing the types of incoming sketchy requests to the average categories of normal requests. If the similarity is low, we can consider that the origin of the incoming request is unusual. For example, if a typical request to a secured server is in JSON format (plain text), but a suspicious request is in JavaScript (client-side script), we can conclude that the incoming request is malicious. If a suspicious request is in XML format (plain text), we can safely assume that the user made a mistake and the alert was false.

From the hypothesis, We can create a CNN model to detect the type of request. Then we compare the incoming requests with the "normal" corresponding category and decide whether the requests is malicious or not using Logistic regression.

Logistic Regression is used when the dependent variable(target) is categorical.

For example, in our case:

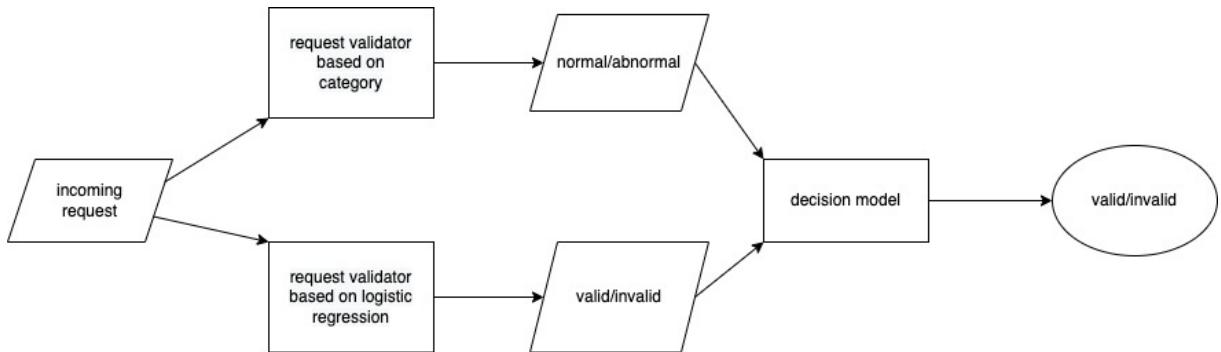
- To predict whether a request is malicious or not.

Our architecture is described in the below figure (Figure 13). Suggest a reasonable decision model for the combined result: When two prediction is the same, the result is straight forward. When the logistic regression model decided that the request is malicious, but the CNN model predicted the request is nornal, the CNN is favored. Otherwise, the Regression model classified the request as valid but the CNN predicted as abnormal, we'll favor the Regression. The desicion model can be expressed in a decision table:

The CNN validator will run for a set period of time (usually one or two weeks) to collect the familiar category of incoming requests and assign a threshold (which can be the mean or maximum (if we are optimistic) distance between each request vector in the observing stage and the sum vector). The same will also happened with the Regression model. After that training phase, the module will run on 'active phase', parallel with the WAF.

1 Code injection is the exploitation of a computer bug that is caused by processing invalid data. The Injection is used by an attacker to introduce code into a vulnerable computer program and change the course of execution.

2 Command injection is an attack in which the goal is the execution of arbitrary commands on the host operating system via a vulnerable application.

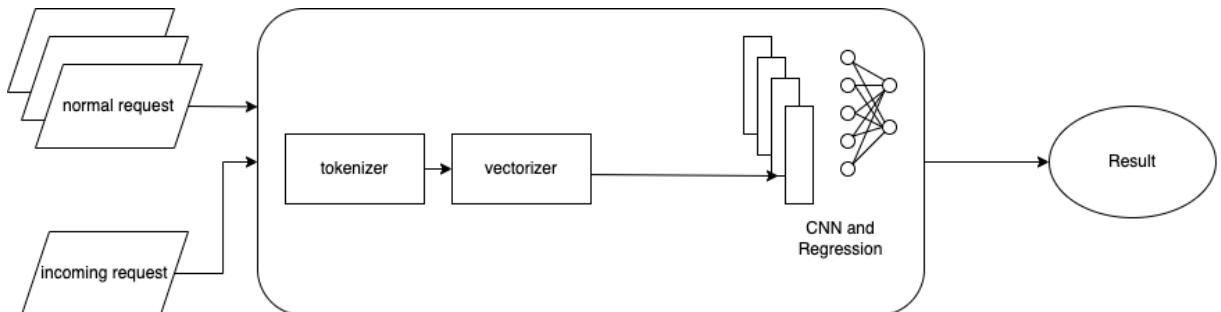
**Figure 5.6:** Malicious request validator architecture

	Regression	CNN	Result
Valid	Normal	Valid	
Valid	Abnormal	Valid	
Invalid	Normal	Valid	
Invalid	Abnormal	Invalid	

Tables 5.1: Table 4.1

- To predict whether a request is malicious or not.

Our architecture is described in Figure 14 below, suggest a reasonable decision model for the combination of CNN and the Regression model

**Figure 5.7:** Decision model for the combination of CNN and the Regression model

The request is routed through the module, which predicts the category. The category of suspicious request is then compared with the common category. Then the Regression model will determined whether the request is good or bad, combining with the normal or abnormal status to decide the result.

6

IMPLEMENTATION

This chapter shows the implementation and deployment processes, including system frameworks and system specifications.

Table of Contents

6.1	Tiền xử lý dữ liệu	40
6.2	Xây dựng mã nguồn hệ thống	44
6.3	Đặc tả phần cứng	45
6.4	Hậu xử lý kết quả phân đoạn	45
6.5	Trực quan hóa kết quả thí nghiệm	46

6.1 Tiề̂n xử lý dữ liệu

Trong mục này, chúng tôi trình bày các công việc chúng tôi sẽ thực hiện trên tập dữ liệu trước lúc sử dụng cho quá trình huấn luyện bao gồm nội suy dữ liệu, trích xuất thành phần gan và biến đổi cường độ sáng điểm ảnh.

6.1.1 Nội suy dữ liệu

Khi xem xét thông số kích thước điểm ảnh trong khối ảnh CT (cột thứ ba Tables A.1), chúng tôi nhận thấy kích thước điểm ảnh giữa các bộ ảnh CT là khác nhau, đặc biệt khoảng cách giữa hai lớp ảnh CT ở một số bộ ảnh rất lớn. Điều này không tốt cho việc học của lớp convolution bởi tính chất khai thác thông tin về cấu trúc không gian của nó. Chúng tôi thực hiện nội suy dữ liệu như mô phỏng ở Figure 6.1. Ở đây chúng tôi sử dụng nội suy bậc ba với hàm `zoom` trong gói `ndimage` của thư viện `scipy` trong Python với kích thước đầu ra của mỗi điểm ảnh là $1\text{mm}(W) \times 1\text{mm}(H) \times 1\text{mm}(D)$.

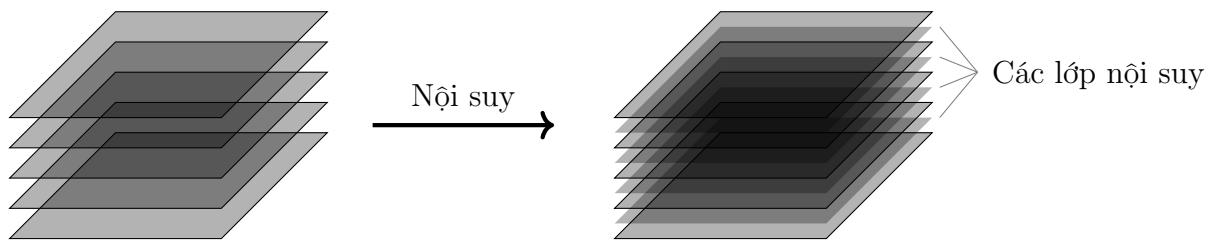


Figure 6.1: Nội suy được sử dụng để lấp các lớp bị thiếu trong khối ảnh CT (Nguồn [11]).

6.1.2 Trích xuất thành phần gan

Từ kết quả khảo sát tình trạng bộ dữ liệu (Tables 4.1), chúng tôi nhận thấy việc sử dụng khối ảnh CT gốc cho việc huấn luyện là bất khả thi vì nhãn phân đoạn cho động mạch ở một số bệnh nhân không được cung cấp. Tuy nhiên, phần động mạch trong cơ quan gan là không đáng kể và hầu hết động mạch đều nằm bên ngoài gan, chúng tôi đề xuất thực hiện trích xuất thành phần gan và chỉ sử dụng thành phần này để huấn luyện hệ thống. Figure 6.2 mô tả việc trích xuất thành phần gan trong khối ảnh CT nhờ sử dụng nhãn phân đoạn gan được cung cấp.

Từ đây trở đi, hệ thống của chúng tôi sẽ đặt trên một giả thiết rằng nhãn phân đoạn cơ quan gan đã biết trước và hệ thống chỉ phân đoạn hệ thống mạch máu bên trong cơ quan gan.

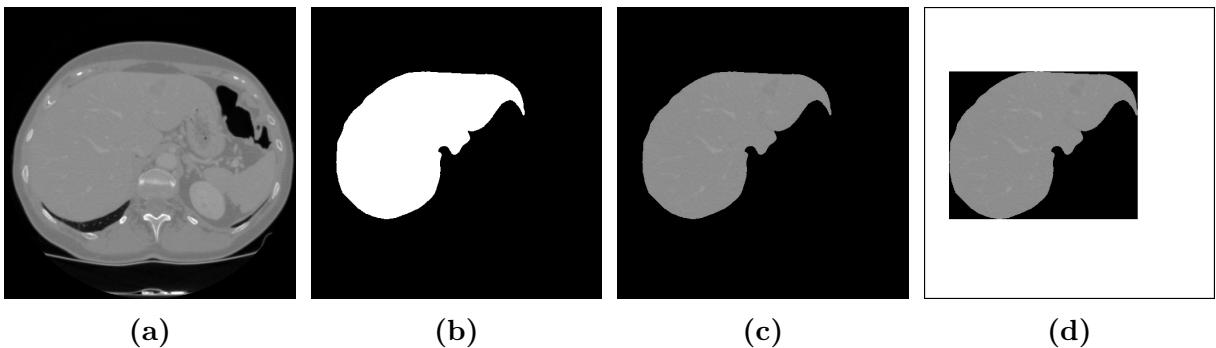


Figure 6.2: Mô phỏng quá trình trích xuất thành phần gan trong khối ảnh CT. (a) ảnh CT gốc. (b) nhãn phân đoạn gan được sử dụng để trích xuất thành phần gan trong ảnh CT. (c) phần ảnh CT chứa gan sau trích xuất. (d) thu giảm kích thước dữ liệu bằng cách giữ lại khối dữ liệu nhỏ nhất chứa gan.

6.1.3 Biến đổi cường độ sáng

Chúng tôi thực hiện biến đổi cường độ sáng điểm ảnh với mong muốn làm rõ hình ảnh mạch máu trên dữ liệu đầu vào, giúp quá trình học dễ dàng hơn. Chúng tôi tiến hành phân tích phân phối mức sáng của các điểm ảnh trong cơ quan gan. Figure 6.3 là biểu đồ phân phối mức sáng nền gan và mạch máu gan của hai đại diện là bệnh nhân số 5 và bệnh nhân số 8. Từ biểu đồ chúng ta thấy phân phối mức sáng trên các bệnh nhân khác nhau có sự chênh lệch. Tuy nhiên, phân phối này có dạng phân phối chuẩn. Chúng tôi đề xuất thực hiện chuẩn hoá để đưa dữ liệu về cùng một phân phối. Chúng tôi thực hiện tính các giá trị ngưỡng bao gồm

$$\text{Left limit} = \mu - \alpha\sigma \quad (6.1)$$

và

$$\text{Right limit} = \mu + \beta\sigma. \quad (6.2)$$

Trong đó, *Left limit* là ngưỡng dưới, *Right limit* là ngưỡng trên, μ là giá trị trung bình và σ là độ lệch chuẩn của các điểm ảnh trong khối cơ quan gan. Hai giá trị α và β chúng tôi đề xuất lần lượt là 3 và 3.5. Giá trị β cao hơn α do độ sáng của các điểm ảnh thuộc mạch máu hầu hết cao hơn độ sáng nền gan, mục đích giữ lại các điểm ảnh thuộc mạch máu có độ sáng cao. Chúng tôi áp dụng các giá trị ngưỡng trên dữ liệu đầu vào. Gọi I^3 là không gian khối dữ liệu, công thức áp dụng ngưỡng vào khối dữ liệu như sau

$$P_{1x} = \begin{cases} \text{Left limit}, & \text{nếu } P_{0x} < \text{Left limit} \\ \text{Right limit}, & \text{nếu } P_{0x} > \text{Right limit} \\ P_{0x}, & \text{còn lại,} \end{cases} \quad (6.3)$$

trong đó, P_{0x} là giá trị điểm ảnh đầu vào và P_{1x} là giá trị điểm ảnh đầu ra tại toạ độ x với $x \in I^3$. Sau đó, chúng tôi chuẩn hoá miền giá trị dữ liệu về khoảng 0 đến 1 theo công thức sau

$$P_{2x} = \frac{P_{1x} - \text{Left limit}}{\text{Right limit} - \text{Left limit}}, \quad (6.4)$$

trong đó, P_{1x} là giá trị đầu ra trong Formula 6.3 và P_{2x} là giá trị sau chuẩn hoá.

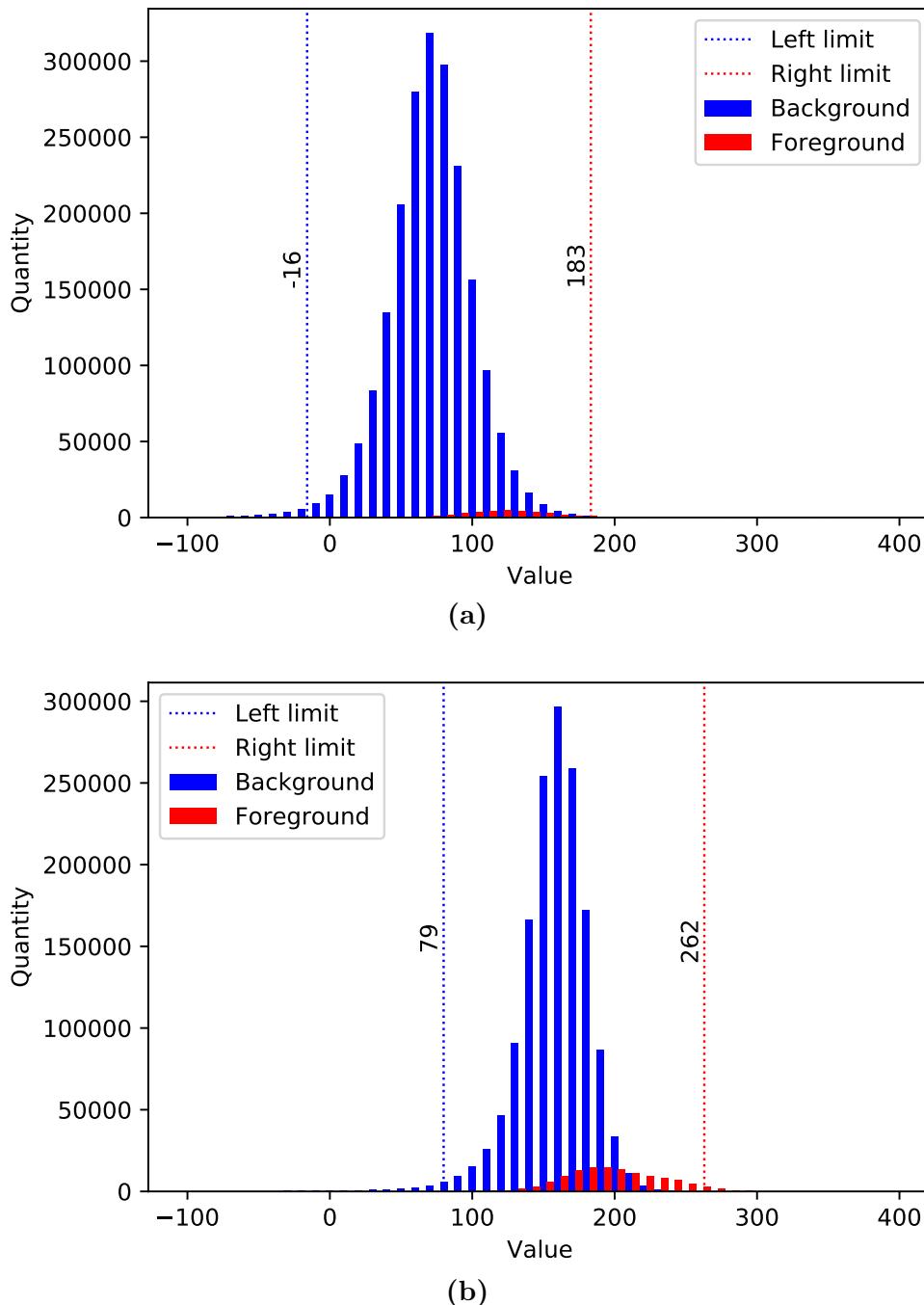


Figure 6.3: Phân phối mức sáng của các điểm ảnh thuộc nền cơ quan gan (Background) và mạch máu gan (Foreground) cùng ngưỡng giới hạn trái (Left limit) và phải (Right limit). **(a)** bệnh nhân số 5. **(b)** bệnh nhân số 8.

Để hình ảnh mạch máu trở nên rõ hơn, chúng tôi tiến hành biến đổi giá trị mức sáng của các điểm ảnh theo công thức sau

$$P_{3x} = P_{2x}^2. \quad (6.5)$$

Figure 6.4 mô tả kết quả chuẩn hóa dữ liệu sau mỗi giai đoạn. So sánh kết quả trước và sau chuẩn hóa, chúng ta có thể thấy hình ảnh mạch máu đã trở nên rõ ràng hơn rất nhiều. Đây là cơ sở quan trọng giúp công tác huấn luyện đạt hiệu quả tốt hơn.

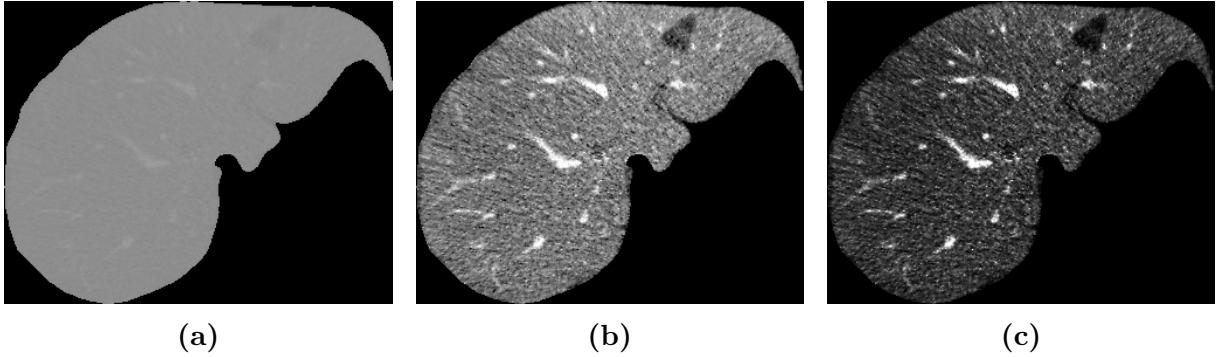


Figure 6.4: Biến đổi cường độ sáng điểm ảnh để làm rõ mạch máu. (a) ảnh trước khi biến đổi. (b) ảnh sau khi thực hiện đặt ngưỡng trên và dưới. (c) biến đổi cường độ sáng bằng hàm bình phương giá trị điểm ảnh.

6.1.4 Làm giàu dữ liệu

Khi huấn luyện một mạng học sâu, thực chất chúng ta đang học một hàm biến đổi có thể ánh xạ từ dữ liệu đầu vào tới nhãn tương ứng. Mô hình càng lớn, số lượng tham số cần học càng nhiều và quá trình huấn luyện đòi hỏi cần nhiều dữ liệu để đạt hiệu quả cao. Do đó, đối với những tập dữ liệu nhỏ, bước làm giàu dữ liệu có vai trò quan trọng trong việc cải thiện hiệu năng mô hình.

Có nhiều cách làm giàu dữ liệu, ví dụ như phép lật hoặc xoay hình. Tuy nhiên, đối với ảnh y khoa như ảnh CT, những phép biến đổi này không có nhiều ý nghĩa vì khi chụp ảnh CT, bệnh nhân được yêu cầu nằm ở một vị trí cố định. Việc xoay, lật ảnh CT sẽ tạo ra các ảnh chụp với các cơ quan bị đảo ngược. Điều này không xảy ra trong thực tế.

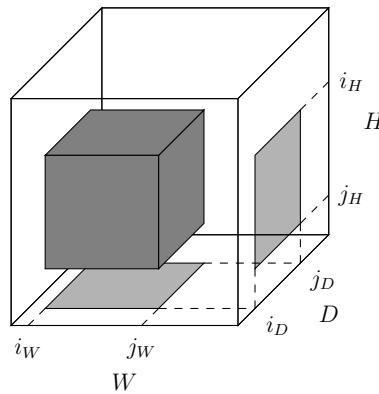


Figure 6.5: Làm giàu dữ liệu bằng phép trích xuất khối ngẫu nhiên.

Trong luận văn này, chúng tôi làm giàu dữ liệu bằng cách thực hiện cắt ngẫu nhiên một khối dữ liệu từ khối dữ liệu ban đầu như Figure 6.5. Kích thước của khối dữ liệu đầu ra là ngẫu nhiên. Tuy nhiên, với mong muốn giữ lại được nhiều thông tin khi đưa vào huấn luyện, chúng tôi đặt ràng buộc kích thước cắt cho khối dữ liệu. Gọi S_H, S_W, S_D lần lượt là kích thước chiều cao, chiều rộng và chiều sâu của khối dữ liệu ban đầu; i_k, j_k lần lượt là chỉ số bắt đầu và kết thúc của khối dữ liệu đầu ra trên khối dữ liệu đầu vào theo chiều k , với $k \in \{H, W, D\}$. Chúng tôi giới hạn i_k trong khoảng $[0, \lfloor 0.1S_k \rfloor]$ và j_k trong khoảng $[\lfloor 0.9S_k \rfloor, S_k]$. Sau đó, chúng tôi thực hiện chia không chồng lấp khối dữ liệu có được thành các khối có kích thước $112 \times 112 \times 112$ và lần lượt đưa vào huấn luyện.

6.2 Xây dựng mã nguồn hệ thống

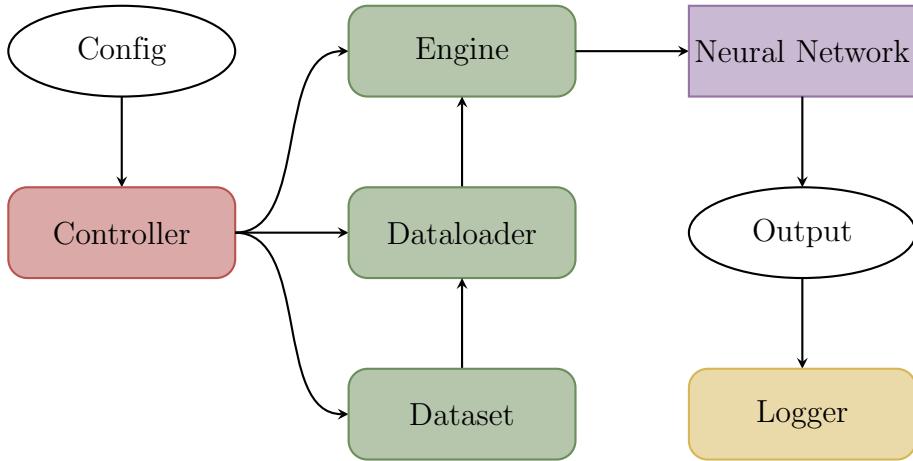


Figure 6.6: Kiến trúc của bộ mã nguồn *Insight Deep Learning* (Source: [12]).

Để thuận tiện hơn trong quá trình thí nghiệm cũng như giúp hệ thống có thể dễ dàng phát triển và bảo trì trong tương lai, chúng tôi sử dụng bộ mã nguồn *Insight Deep Learning* của GVLab được hiện thực bằng PyTorch với những ưu điểm sau đây:

- Bộ mã được tổ chức thành các khái niệm rõ ràng, hợp lý giúp dễ dàng trong quá trình sử dụng, mở rộng và bảo trì.
- Hỗ trợ huấn luyện song song trên nhiều GPU.
- Hỗ trợ giám sát thời gian thực trong quá trình huấn luyện với nhiều dạng khác nhau như biểu đồ, chữ,...
- Trong trường hợp quá trình huấn luyện bị dừng đột ngột vì lý do không mong muốn, hệ thống cho phép phục hồi quá trình huấn luyện nhờ khả năng sao lưu và khôi phục trạng thái trong quá trình huấn luyện.
- Các siêu tham số và các cài đặt cho quá trình huấn luyện được quản lý trong một tệp YAML¹. Nhờ vậy, khi cần thay đổi thông số cài đặt, lập trình viên chỉ cần chỉnh sửa thông tin trong tệp này.
- Sử dụng chung một bộ mã nguồn với cùng một ngôn ngữ, cấu trúc cho phép các thành viên của GVLab hợp tác với nhau thuận lợi hơn cũng như quá trình kế thừa và phát triển mã nguồn dễ dàng hơn.

Kiến trúc của bộ mã được mô tả trong Figure 6.6. Controller là khái niệm được thực thi đầu tiên với các thông số lấy từ khái niệm Config. Controller có nhiệm vụ điều khiển các thành phần Dataset, DataLoader, Engine làm việc với nhau. Dataset xử lý tất cả các tác vụ liên quan đến dữ liệu như đọc tệp, làm giàu dữ liệu. DataLoader quản lý số lượng mẫu và cách mà mỗi mẫu được đưa vào huấn luyện. Engine sẽ lấy dữ liệu từ DataLoader để đưa vào Neural Network. Phụ thuộc vào loại Engine, đầu ra của Neural Network sẽ được dùng để tính giá trị lỗi, đưa ra kết quả dự đoán hoặc để tính toán các độ đo ghi vào Logger.

¹ YAML là một chuẩn tuân tự hóa dữ liệu cho nhiều ngôn ngữ lập trình dưới dạng chữ, giúp con người dễ dàng đọc, hiểu.

Trong luận văn này, chúng tôi kế thừa và phát triển bộ mã để phục vụ cho quá trình xây dựng hệ thống. Những bổ sung chính mà chúng tôi thực hiện trên bộ mã liên quan đến các mô hình trong các công trình liên quan chúng tôi tham khảo và khả năng làm việc với dữ liệu 3D.

6.3 Đặc tả phần cứng

Huấn luyện mạng học sâu với một lượng lớn dữ liệu đòi hỏi sức mạnh tính toán cao để công tác huấn luyện diễn ra nhanh chóng và hiệu quả. Để đáp ứng yêu cầu đó, trong luận văn này, chúng tôi sử dụng hệ thống máy tính hiệu năng cao (HPCC) được hỗ trợ bởi GVLab nhằm mục đích huấn luyện hệ thống. Tables 6.1 mô tả chi tiết thông số cấu hình hệ thống mà chúng tôi đã sử dụng.

Tables 6.1: Thông tin phần cứng hệ thống máy tính.

STT	Phần cứng	Cấu hình
1	CPU	Intel(R) Xeon(R) CPU E5-2640 v3 @ 2.60GHz 16 cores
2	RAM	128 GB
3	GPU	NVIDIA Tesla P100 16GB

6.4 Hậu xử lý kết quả phân đoạn

Hệ thống mạch máu là một cấu trúc khép kín. Nếu coi hệ thống này là một đối tượng trong không gian thì nó sẽ gói gọn trong một thành phần liên thông duy nhất. Đối với hệ thống mạch máu trong khối ảnh CT của một phần cơ thể, cụ thể với giả thiết được nêu ra trong Sector 6.1.2, hệ thống này sẽ bao gồm một số lượng hữu hạn ¹ các nhánh mạch máu mà mỗi nhánh này là một thành phần liên thông.

Hệ thống mạch máu có được từ kết quả phân đoạn không thể tránh khỏi có những điểm ảnh bị phân đoạn sai. Đặc biệt, những điểm ảnh rời rạc không thuộc mạch máu bị phân đoạn sai sẽ làm xấu đi kết quả thí nghiệm. Nhằm cải thiện kết quả phân đoạn, chúng tôi đề xuất thực hiện xác định các thành phần liên thông trong cây mạch máu được sinh ra. Sau đó, lần lượt loại bỏ những thành phần liên thông có tổng số lượng điểm ảnh là nhỏ nhất sao cho tổng số lượng điểm ảnh bị loại bỏ không vượt quá 10% tổng số điểm ảnh được phân đoạn là mạch máu.

Việc xác định các thành phần liên thông và loại bỏ các thành phần liên thông nhỏ được chúng tôi thực hiện thông qua các hàm đã được hiện thực sẵn trong ngôn ngữ Python lần lượt là hàm `label` trong gói `ndimage.measurements` của thư viện `scipy` và hàm `histogram` và `argsort` của thư viện `numpy`.

¹ Có khoảng 6 nhánh mạch máu trong cơ quan gan với 3 nhánh xuất phát từ tĩnh mạch chủ và 3 nhánh còn lại xuất bạn từ tĩnh mạch cửa.

6.5 Trực quan hoá kết quả thí nghiệm

Sau khi có kết quả phân đoạn mạch máu cũng như đường chính giữa và điểm phân nhánh mạch máu. Chúng tôi thực hiện trực quan hoá kết quả thông qua hai bước.

Bước thứ nhất, chúng tôi tiến hành trích xuất lưới bề mặt đối tượng cần trực quan bằng thư viện VTK. Bước này nhằm giảm thiểu khối lượng tính toán trên GPU lúc hiển thị, bởi nếu sử dụng ngay kết quả đầu ra của hệ thống là mảng ba chiều thì khối lượng tính toán cần thực hiện trên GPU là rất lớn. Từ đó, cải thiện chất lượng hiển thị nhờ tăng số lượng khung hình trên giây và việc trực quan sẽ mượt mà hơn.

Bước thứ hai, hiển thị lưới bề mặt của đối tượng trên ứng dụng Slicer. Với ứng dụng Slicer, chúng tôi có thể hiển thị chồng kết quả phân đoạn và nhãn phân đoạn tương ứng lên nhau, từ đó có thể dễ dàng so sánh sự sai khác của kết quả phân đoạn. Figure 6.7 là ví dụ về việc sử dụng ứng dụng Slicer để trực quan hệ thống tĩnh mạch của một bệnh nhân.

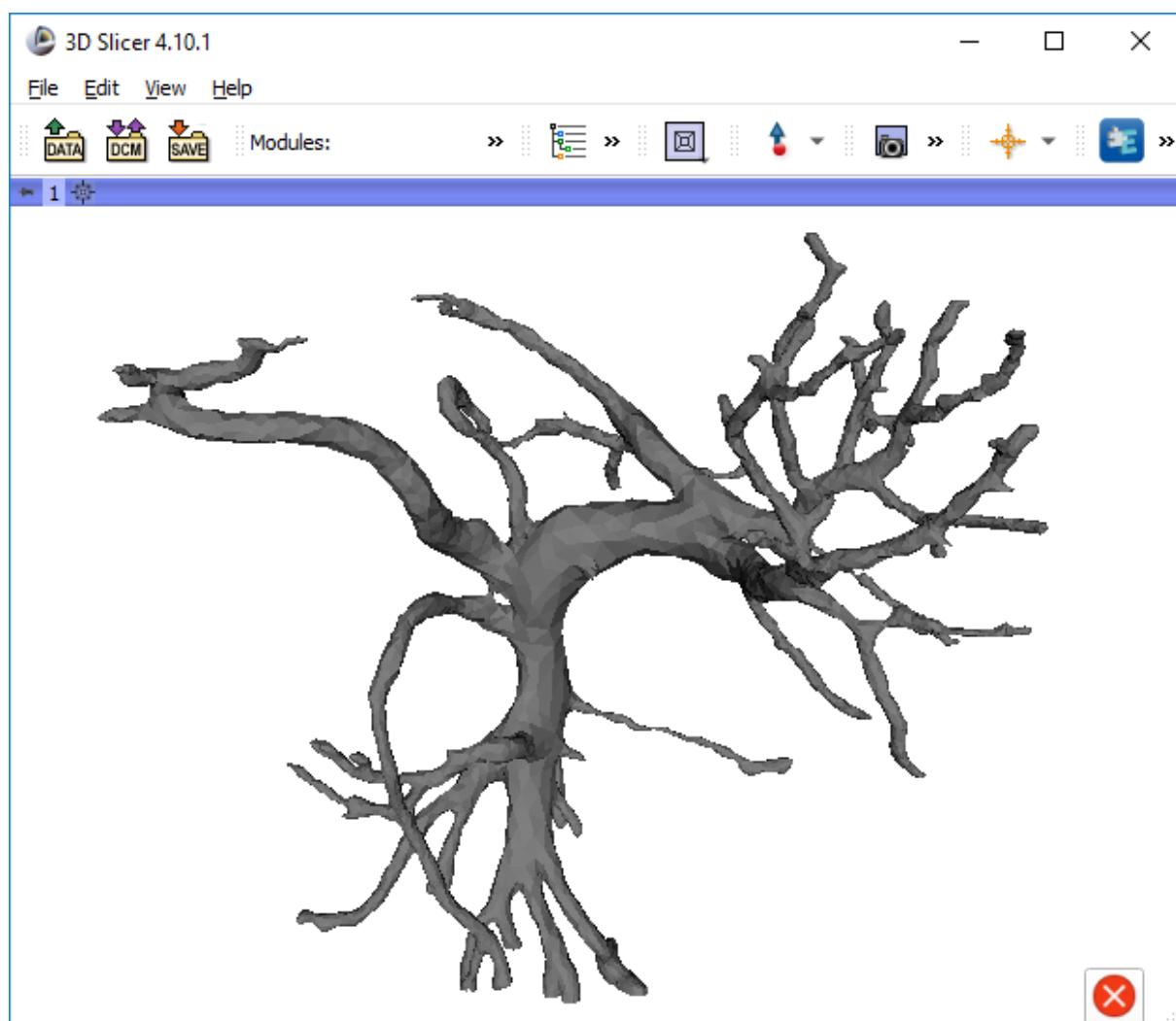


Figure 6.7: Trực quan hoá kết quả trên ứng dụng Slicer.

7

EXPERIMENTS

This chapter discusses the division of the data set before training, assessment techniques, and experimental results. Compare the outcomes of the experiments we suggest to the reference experiments from relevant works.

Table of Contents

7.1 Chuẩn bị dữ liệu	48
7.2 Phương pháp đánh giá	48
7.3 Kết quả thí nghiệm	50

7.1 Chuẩn bị dữ liệu

Đối với các giải thuật học máy nói chung cũng như mạng học sâu nói riêng, việc chuẩn bị dữ liệu là vô cùng quan trọng. Nếu sử dụng mô hình đã huấn luyện để dự đoán trên tập dữ liệu không cùng phân bố với tập dữ liệu được sử dụng để huấn luyện mô hình trước đó thì kết quả dự đoán không thể chính xác. Do đó, phân chia dữ liệu sao cho phân bố trên các tập đồng đều với nhau là một trong những yếu tố quan trọng quyết định tới mức hiệu quả của mạng học sâu.

Sau khi khảo sát các ảnh chụp CT của 20 bệnh nhân, chúng tôi nhận thấy sự xuất hiện các khối u có ảnh hưởng đến phổi mức sáng của các điểm ảnh trong cơ quan gan. Do đó, chúng tôi chia các bệnh nhân ra 03 nhóm: bệnh nhân không có khối u trong gan, bệnh nhân có một khối u trong gan và bệnh nhân có nhiều khối u trong gan. Chúng tôi chia tập dữ liệu thành ba tập dữ liệu bao gồm tập huấn luyện, tập kiểm thử và tập kiểm tra sao cho mỗi tập đều có bệnh nhân không có u, bệnh nhân có một khối u và bệnh nhân có nhiều khối u. Chi tiết các tập dữ liệu được chúng tôi trình bày trong Tables 7.1.

Tables 7.1: Bảng phân chia các bệnh nhân thành các tập dữ liệu.

STT	Tập	Số lượng khối u gan			Tổng
		0 khối u	1 khối u	Nhiều khối u	
1	Huấn luyện	5, 7, 11	2, 3, 9, 12	1, 4, 8, 10, 15, 17, 19	14
2	Kiểm thử	14	16	6	3
3	Kiểm tra	20	18	13	3

7.2 Phương pháp đánh giá

Quá trình đánh giá một mô hình phân đoạn hình ảnh y khoa liên quan đến việc đánh giá độ chính xác của kết quả dự đoán. Hai hoạt động cơ bản cần được tiến hành để đánh giá một cách toàn diện một mô hình là đánh giá định tính và đánh giá định lượng. Đánh giá định lượng cho biết một cách tổng quát độ tốt của mô hình, trong khi đó đánh giá định tính cho biết mức độ ổn định của mô hình trong quá trình làm việc (trong trường hợp xấu nhất, trung bình và tốt nhất). Trong phần này, chúng tôi trình bày 4 độ đo có liên quan bao gồm Precision, Recall, IoU và Dice. Trong đó, IoU và Dice được chúng tôi sử dụng để đánh giá kết quả dự đoán của các mô hình.

Ta quy ước,

- TP (true positive): số lượng điểm ảnh foreground được dự đoán đúng,
- TN (true negative): số lượng điểm ảnh background được dự đoán đúng,
- FP (false positive): số lượng điểm ảnh background bị dự đoán sai,
- FN (false negative): số lượng điểm ảnh foreground bị dự đoán sai.

Precision trả lời cho câu hỏi “Số dự đoán thực sự chính xác chiếm bao nhiêu phần trong số các dự đoán foreground?”. Công thức Precision được định nghĩa trong Formula 7.1. Precision càng cao, kết quả dự đoán càng tốt.

$$Precision = \frac{TP}{TP + FP} \quad (7.1)$$

Recall trả lời cho câu hỏi “Số dự đoán thực sự chính xác chiếm bao nhiêu phần trong số các mẫu foreground”. Công thức Recall được định nghĩa trong Formula 7.2. Recall càng cao, kết quả dự đoán càng tốt.

$$Recall = \frac{TP}{TP + FN} \quad (7.2)$$

Dice là độ đo cân bằng giữa Precision và Recall. Trong trường hợp mô hình dự đoán chính xác một lượng nhỏ mẫu thuộc foreground, giá trị Precision sẽ cao, tuy nhiên số lượng false negative sẽ cao theo, nghĩa là giá trị Recall thấp. Trong trường hợp mô hình dự đoán một lượng lớn mẫu thuộc foreground, giá trị Recall sẽ cao, tuy nhiên số lượng false positive sẽ cao theo, tức là giá trị Precision thấp. Dice được tính toán bằng cách sử dụng đồng thời Precision và Recall. Khi giá trị Precision và giá trị Recall cùng cao, giá trị Dice sẽ cao, ngược lại, khi một trong hai giá trị Precision và Recall thấp, giá trị Dice sẽ thấp. Do đó, Dice phù hợp cho các bài toán phân loại có sự mất cân bằng về nhãn. Công thức Dice được định nghĩa trong Formula 7.2.

$$Dice = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (7.3)$$

IoU cho biết mức độ tương tự giữa hai mẫu X và Y . IoU được định nghĩa trong Formula 7.4.

$$IoU = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|} \quad (7.4)$$

7.3 Kết quả thí nghiệm

Trước khi tiến hành các thí nghiệm để đánh giá hệ thống, chúng tôi thực hiện kiểm chứng tính hiệu quả của đề xuất cắt giảm độ sâu trong mô hình U-Net, để từ đó chọn ra mô hình tốt hơn phục vụ cho các thí nghiệm về sau. Tables 7.2 đặc tả thông số chi tiết của các thí nghiệm dùng để so sánh tính hiệu quả trước và sau điều chỉnh số tầng ở mô hình U-Net sử dụng convolution 3D.

Tables 7.2: Thông số các thí nghiệm so sánh tính hiệu quả trước và sau điều chỉnh số tầng ở mô hình U-Net sử dụng convolution 3D.

STT	Kiến trúc	Số tầng	BatchNorm	Hàm lỗi	Learning rate	Momentum
1	U-Net	5	Không	CrossEntropy	0.0001	0.90
2	U-Net*	3	Không	CrossEntropy	0.0001	0.90

Tables 7.3 là kết quả thu được khi so sánh hai mô hình. Từ kết quả này chúng ta thấy rằng, chất lượng phân đoạn hệ thống mạch máu trên tập kiểm tra trước và sau hậu xử lý (cột **Tập kiểm tra** và **Tập kiểm tra***) ở mô hình U-Net* không có nhiều sự khác biệt so với mô hình U-Net. Chúng tôi, việc cắt giảm hai tầng 4 và 5 trong mô hình U-Net là hợp lý vì chúng không có nhiều đóng góp trong quá trình học. Hơn nữa, việc cắt giảm số tầng giúp cho mô hình trở nên nhẹ hơn, thời gian inference¹ trung bình cho một tập ảnh CT ở mô hình U-Net* là 4.46 giây, nhanh hơn gần 1 giây so với mô hình U-Net là 5.40 giây. Chênh lệch này tuy nhỏ nhưng sẽ có ý nghĩa rất lớn khi áp dụng mô hình vào thực tiễn với khối lượng dữ liệu khổng lồ. Thời gian dành cho công tác chẩn đoán càng được rút ngắn bao nhiêu, cơ hội chữa trị thành công cho bệnh nhân càng lớn bấy nhiêu. Vì vậy, chúng tôi lựa chọn sử dụng mô hình U-Net* trong các thí nghiệm về sau để đánh giá hệ thống.

Tables 7.3: Kết quả so sánh tính hiệu quả trước và sau điều chỉnh số tầng ở mô hình U-Net sử dụng convolution 3D.

STT	Kiến trúc	Thời gian inference (s)	Tập kiểm tra		Tập kiểm tra*	
			IoU	Dice	IoU	Dice
1	U-Net	5.40	0.365	0.533	0.379	0.548
2	U-Net*	4.46	0.380	0.550	0.375	0.545

¹ Inference trong lĩnh vực học sâu là thuật ngữ dùng để chỉ giai đoạn mô hình sau huấn luyện được sử dụng để dự đoán các mẫu dữ liệu thử nghiệm. Không giống quá trình huấn luyện, inference không thực hiện lan truyền ngược để tính toán lỗi và cập nhật trọng số.

Dầu tiên, chúng tôi tiến hành 2 thí nghiệm tham khảo từ các công trình có liên quan. Tables 7.4 mô tả chi tiết các thí nghiệm và các siêu tham số được sử dụng.

Tables 7.4: Thông số các thí nghiệm tham khảo.

STT	Thí nghiệm	Kiến trúc	BatchNorm	Hàm lõi	Learning rate	Momentum
1	Thí nghiệm 1	DeepVesselNet	Không	CrossEntropy	0.0001	0.90
2	Thí nghiệm 2	U-Net*	Không	CrossEntropy	0.0001	0.90

Tables 7.5 là kết quả huấn luyện của các thí nghiệm được kiểm tra trên các tập dữ liệu. Từ kết quả thu được, chúng ta thấy rằng mô hình U-Net hoạt động hiệu quả hơn mô hình DeepVesselNet với kết quả tốt nhất trên hai giá trị đo IoU và Dice lần lượt là 0.380 và 0.550 trên tập kiểm tra. Tuy nhiên, các giá trị này trở nên xấu đi sau bước hậu xử lý. Điều này có thể được giải thích rằng, mô hình phân đoạn cho kết quả hệ thống mạch máu rời rạc, những điểm dự đoán đúng mạch máu đã bị loại bỏ vì thể tích thành phần liên thông của nó quá nhỏ.

Tables 7.5: Kết quả các thí nghiệm tham khảo.

STT	Thí nghiệm	Tập huấn luyện		Tập kiểm thử		Tập kiểm tra		Tập kiểm tra*	
		IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice
1	Thí nghiệm 1	0.364	0.530	0.459	0.627	0.365	0.533	0.362	0.531
2	Thí nghiệm 2	0.392	0.557	0.518	0.681	0.380	0.550	0.375	0.545

Chi tiết kết quả phân đoạn cho từng bệnh nhân trong tập kiểm tra được thể hiện trong Tables 7.6. Trong đó, giá trị IoU và Dice trong trường hợp tốt nhất và xấu nhất lần lượt là 0.404, 0.575 và 0.344, 0.512. Figure 7.1, Figure 7.2 và Figure 7.3 là hình ảnh trực quan kết quả thí nghiệm 2 cho trường hợp tốt nhất. Figure 7.4, Figure 7.5 và Figure 7.6 là hình ảnh trực quan kết quả thí nghiệm 2 cho trường hợp xấu nhất.

Tables 7.6: Kết quả thí nghiệm 2 của từng bệnh nhân trong tập kiểm tra.

STT	Bệnh nhân	IoU	Dice
1	Bệnh nhân 13	0.404	0.575
2	Bệnh nhân 18	0.344	0.512
3	Bệnh nhân 20	0.378	0.549

Tiếp theo, chúng tôi tiến hành 6 thí nghiệm do chúng tôi đề xuất. Tables 7.7 mô tả chi tiết các thí nghiệm và các siêu tham số được sử dụng.

Tables 7.7: Thông số các thí nghiệm đề xuất.

STT	Thí nghiệm	Kiến trúc	BatchNorm	Hàm lỗi	Learning rate	Momentum
1	Thí nghiệm 3	DeepVesselNet	Có	CrossEntropy	0.001	0.90
2	Thí nghiệm 4	DeepVesselNet	Không	Dice	0.0001	0.90
3	Thí nghiệm 5	DeepVesselNet	Có	Dice	0.001	0.90
4	Thí nghiệm 6	U-Net*	Có	CrossEntropy	0.001	0.90
5	Thí nghiệm 7	U-Net*	Không	Dice	0.0001	0.90
6	Thí nghiệm 8	U-Net*	Có	Dice	0.001	0.90

Tables 7.8 là kết quả huấn luyện cho 6 thí nghiệm này. Với mỗi thí nghiệm, chúng tôi đánh giá trên tất cả các tập dữ liệu cũng như kết quả sau bước hậu xử lý. Từ số liệu thu được, chúng ta thấy được rằng, các thí nghiệm có sử dụng lớp batchnorm và hàm lỗi dice (thí nghiệm 5 và thí nghiệm 8) cho kết quả tốt hơn hẳn với kết quả tốt nhất trên các độ đo IoU, Dice lần lượt là 0.384, 0.552 trên thí nghiệm sử dụng mô hình U-Net. Đồng thời, ở các thí nghiệm này, việc thực hiện hậu xử lý cho thấy sự hiệu quả, các giá trị IoU, Dice được cải thiện với giá trị tương ứng là 0.400, 0.569. So với kết quả trong các thí nghiệm tham khảo từ các công trình liên quan, kết quả thí nghiệm do chúng tôi đề xuất cải thiện kết quả của hệ thống 4%.

Tables 7.8: Kết quả các thí nghiệm đề xuất.

STT	Thí nghiệm	Tập huấn luyện		Tập kiểm thử		Tập kiểm tra		Tập kiểm tra*	
		IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice
1	Thí nghiệm 3	0.347	0.508	0.457	0.624	0.256	0.402	0.261	0.409
2	Thí nghiệm 4	0.417	0.584	0.555	0.713	0.351	0.517	0.371	0.538
3	Thí nghiệm 5	0.411	0.578	0.531	0.692	0.373	0.541	0.381	0.550
4	Thí nghiệm 6	0.328	0.483	0.402	0.565	0.377	0.545	0.371	0.537
5	Thí nghiệm 7	0.406	0.572	0.558	0.715	0.352	0.519	0.360	0.527
6	Thí nghiệm 8	0.439	0.606	0.539	0.699	0.384	0.552	0.400	0.569

Chi tiết kết quả phân đoạn cho từng bệnh nhân trong tập kiểm tra được thể hiện trong Tables 7.9. Trong đó, giá trị IoU và Dice trong trường hợp tốt nhất và xấu nhất lần lượt là 0.474, 0.643 và 0.342, 0.510. Figure 7.7, Figure 7.8 và Figure 7.9 là hình ảnh trực quan kết quả thí nghiệm 8 cho trường hợp tốt nhất. Figure 7.10, Figure 7.11 và Figure 7.12 là hình ảnh trực quan kết quả thí nghiệm 8 cho trường hợp xấu nhất.

Tables 7.9: Kết quả thí nghiệm 8 của từng bệnh nhân trong tập kiểm tra.

STT	Bệnh nhân	IoU	Dice
1	Bệnh nhân 13	0.474	0.643
2	Bệnh nhân 18	0.342	0.510
3	Bệnh nhân 20	0.384	0.555

Với mong muốn tiếp tục cải thiện kết quả thí nghiệm và nhờ hiểu được những ưu điểm của mạng DenseNet, ResNet; chúng tôi đề xuất sử dụng ý tưởng của hai mạng này cho mô hình U-Net. Tables 7.10 mô tả chi tiết kiến trúc và hàm lỗi được sử dụng cũng như các siêu tham số cho các thí nghiệm.

Tables 7.10: Thông số các thí nghiệm có sự kết hợp của DenseNet và ResNet.

STT	Thí nghiệm	Kiến trúc	BatchNorm	Hàm lỗi	Learning rate	Momentum
1	Thí nghiệm 9	Dense-U-Net	Có	Dice	0.001	0.90
2	Thí nghiệm 10	Res-U-Net	Có	Dice	0.001	0.90

Tables 7.11 là kết quả huấn luyện hai mô hình Dense-U-Net và Res-U-Net. Từ kết quả thí nghiệm, chúng ta thấy rằng, mô hình kết hợp ý tưởng DenseNet cho kết quả tốt hơn mô hình kết hợp ý tưởng ResNet. Tuy nhiên, kết quả trong các thí nghiệm này không cải thiện hơn so với các thí nghiệm trước.

Tables 7.11: Kết quả các thí nghiệm DenseUNet và ResUNet.

STT	Thí nghiệm	Tập huấn luyện		Tập kiểm thử		Tập kiểm tra		Tập kiểm tra*	
		IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice
1	Thí nghiệm 9	0.417	0.584	0.529	0.691	0.356	0.521	0.382	0.549
2	Thí nghiệm 10	0.422	0.590	0.536	0.697	0.355	0.520	0.377	0.544

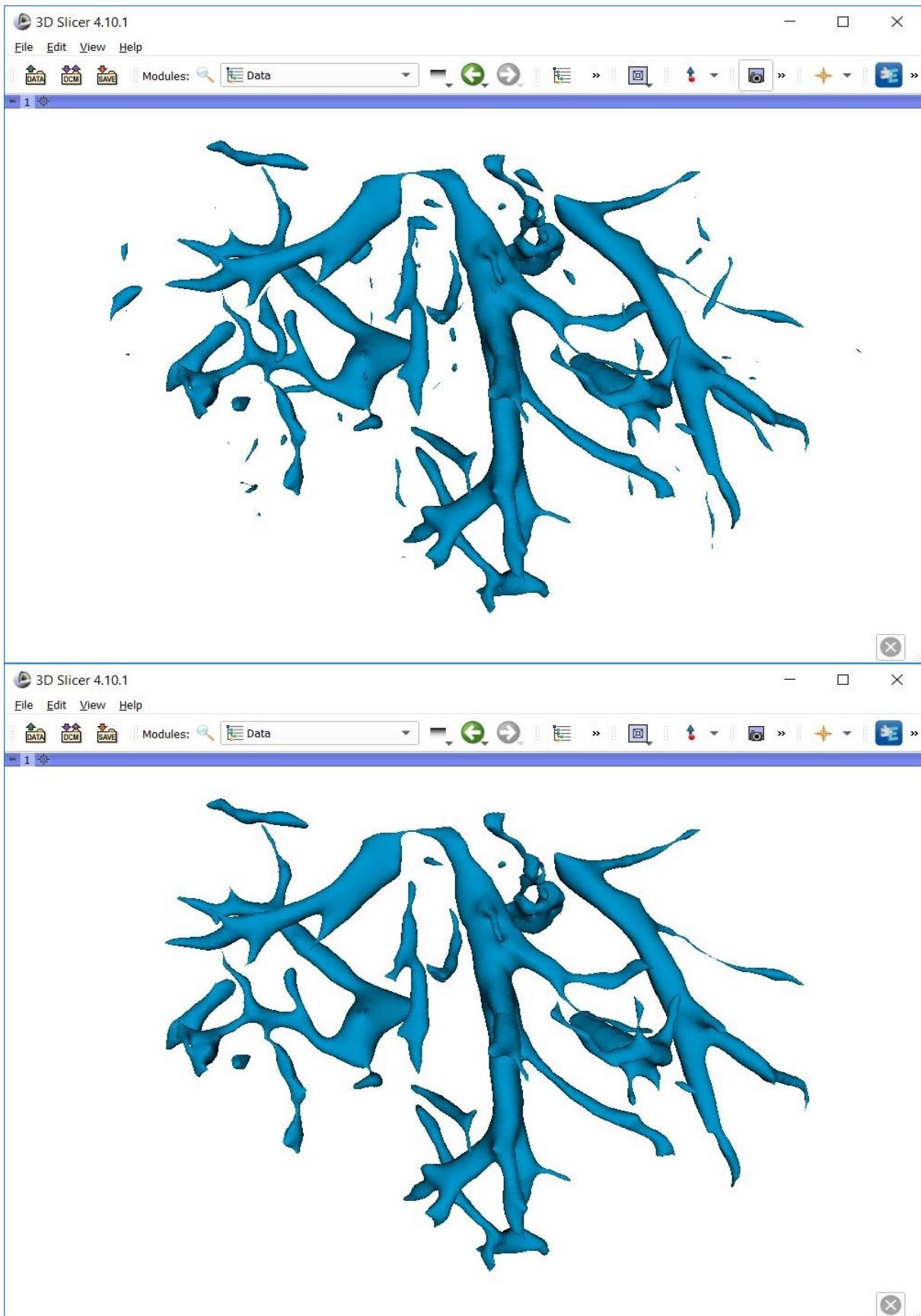


Figure 7.1: Kết quả hệ thống mạch máu của trường hợp tốt nhất trong thí nghiệm 2 trước và sau hậu xử lý. Các thành phần nhỏ và rời rạc đã được loại bỏ.

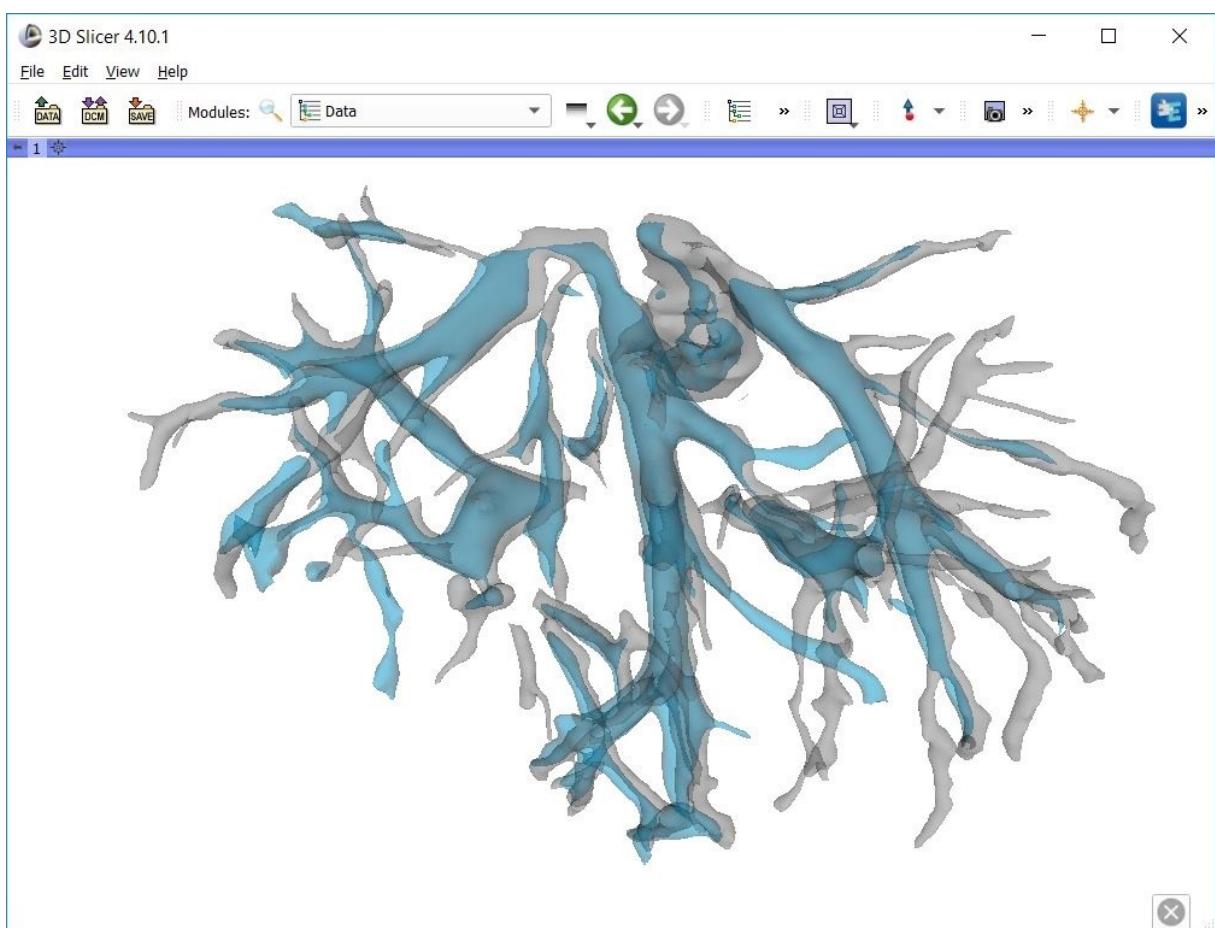


Figure 7.2: Kết quả hệ thống mạch máu (màu xanh) của trường hợp tốt nhất trong thí nghiệm 2 khi so sánh với nhãn phân đoạn (màu xám). Hệ thống đã phân đoạn được các nhánh chính của mạch máu, tuy nhiên, còn rất nhiều nhánh mạch máu nhỏ hệ thống chưa phân đoạn được.

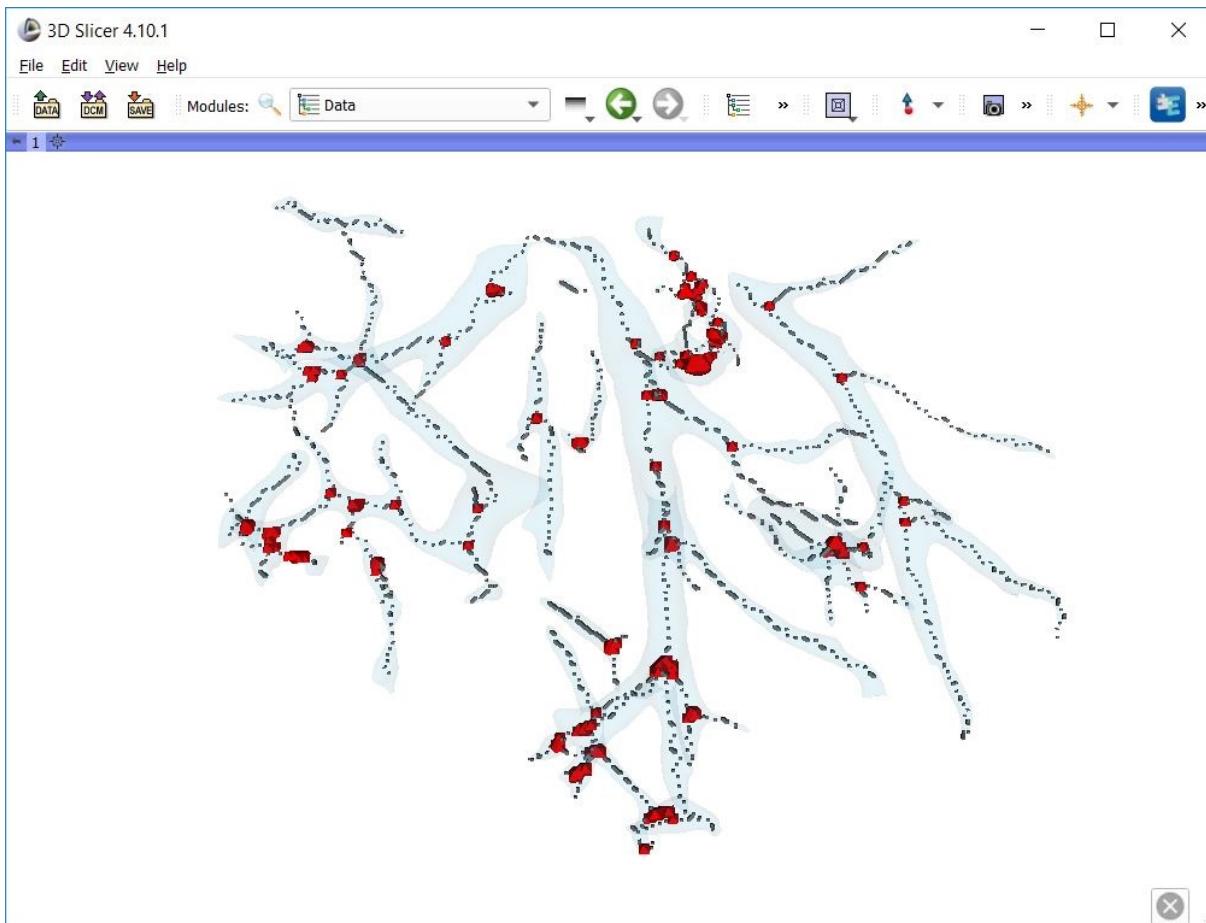


Figure 7.3: Kết quả tìm đường chính giữa (đường màu xám) và điểm phân nhánh (màu đỏ) của trường hợp tốt nhất trong thí nghiệm 2. Hệ thống đã xác định được đường chính giữa và các điểm phân nhánh của mạch máu.

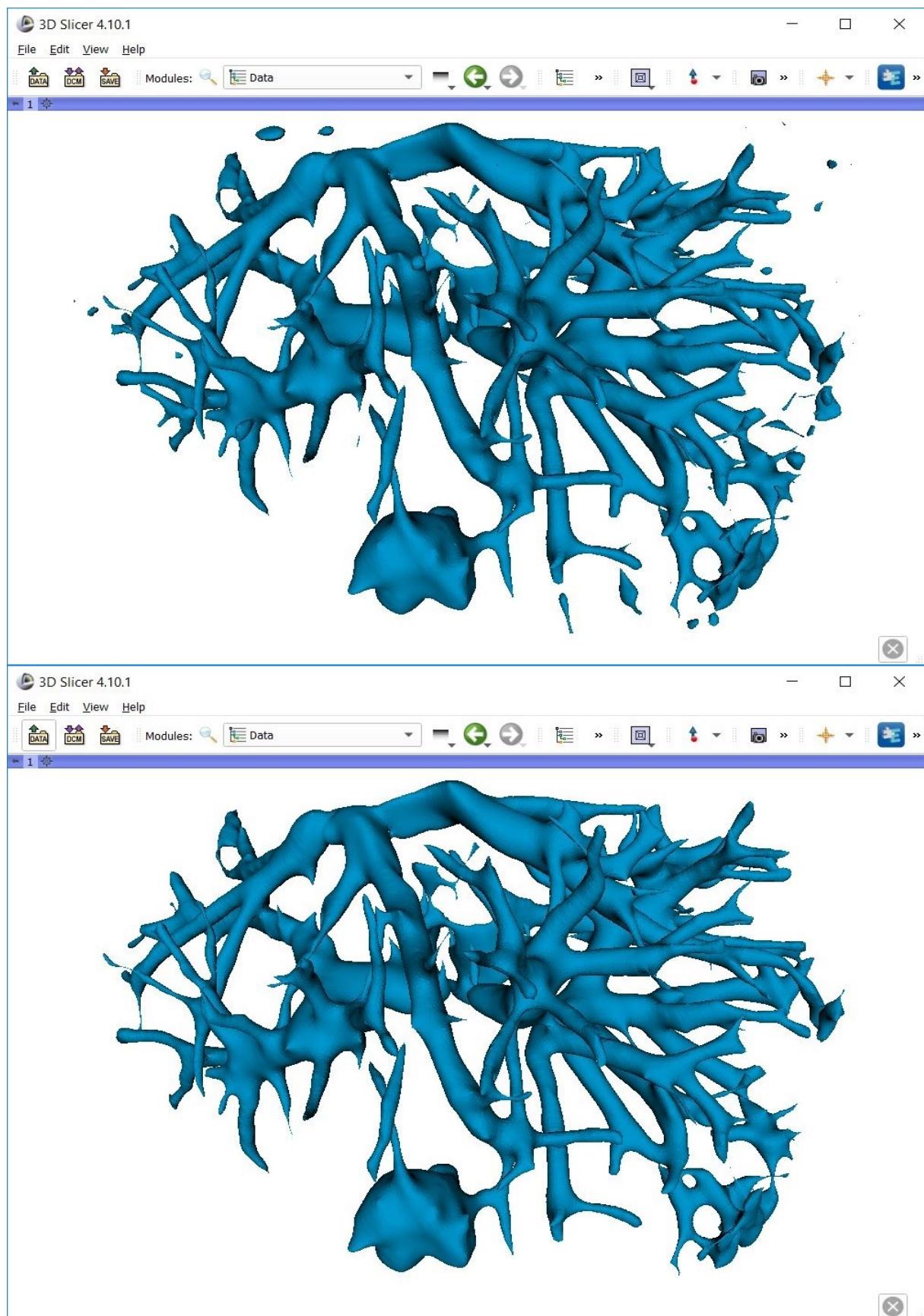


Figure 7.4: Kết quả hệ thống mạch máu của trường hợp xấu nhất trong thí nghiệm 2 trước và sau hậu xử lý. Hệ thống phân đoạn mạch máu quá lớn, bước hậu xử lý không có nhiều hiệu quả.

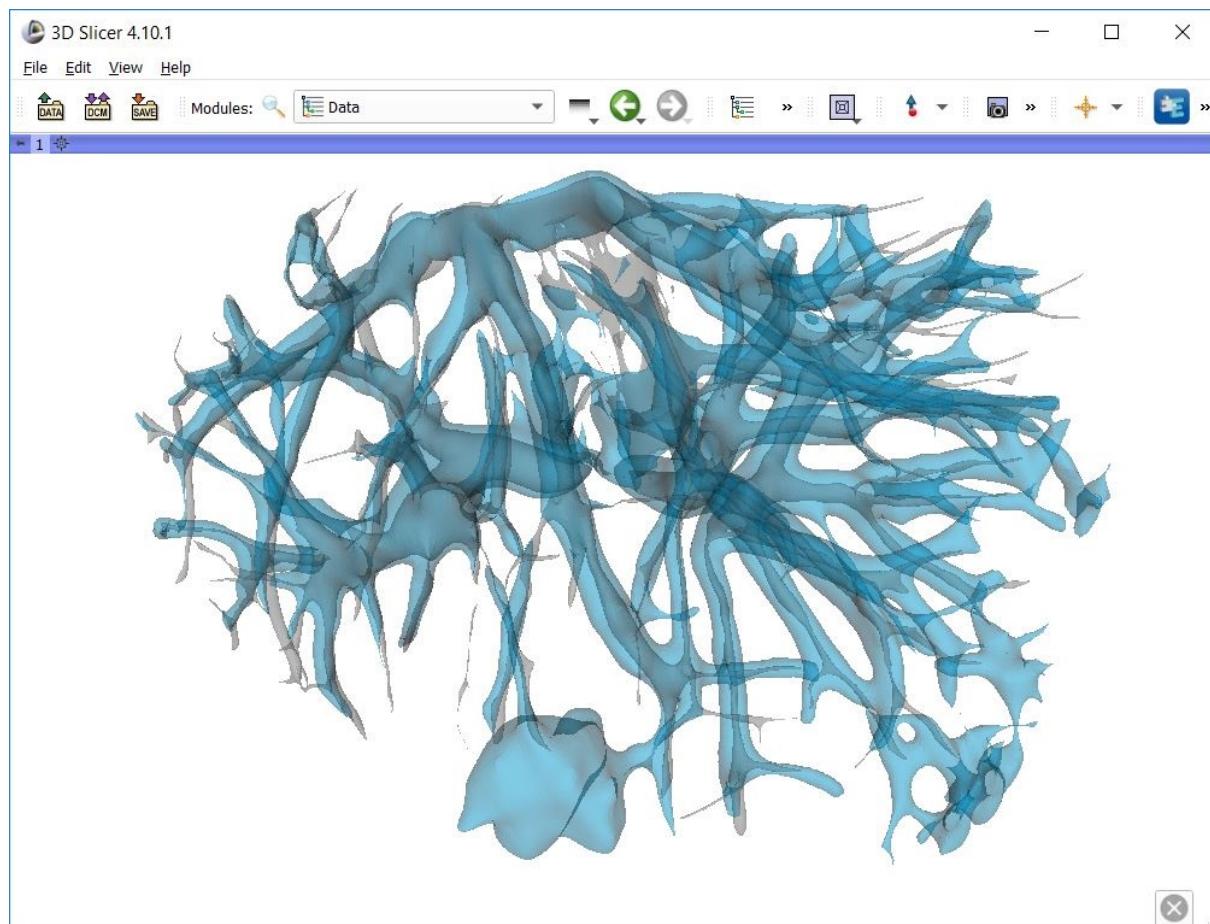


Figure 7.5: Kết quả hệ thống mạch máu (màu xanh) của trường hợp xấu nhất trong thí nghiệm 2 khi so sánh với nhăn phân đoạn (màu xám). Các mạch máu được phân đoạn lớn hơn rất nhiều so với thực tế. Ngoài ra, xuất hiện một khối phân đoạn sai lớn có thể là khối u.

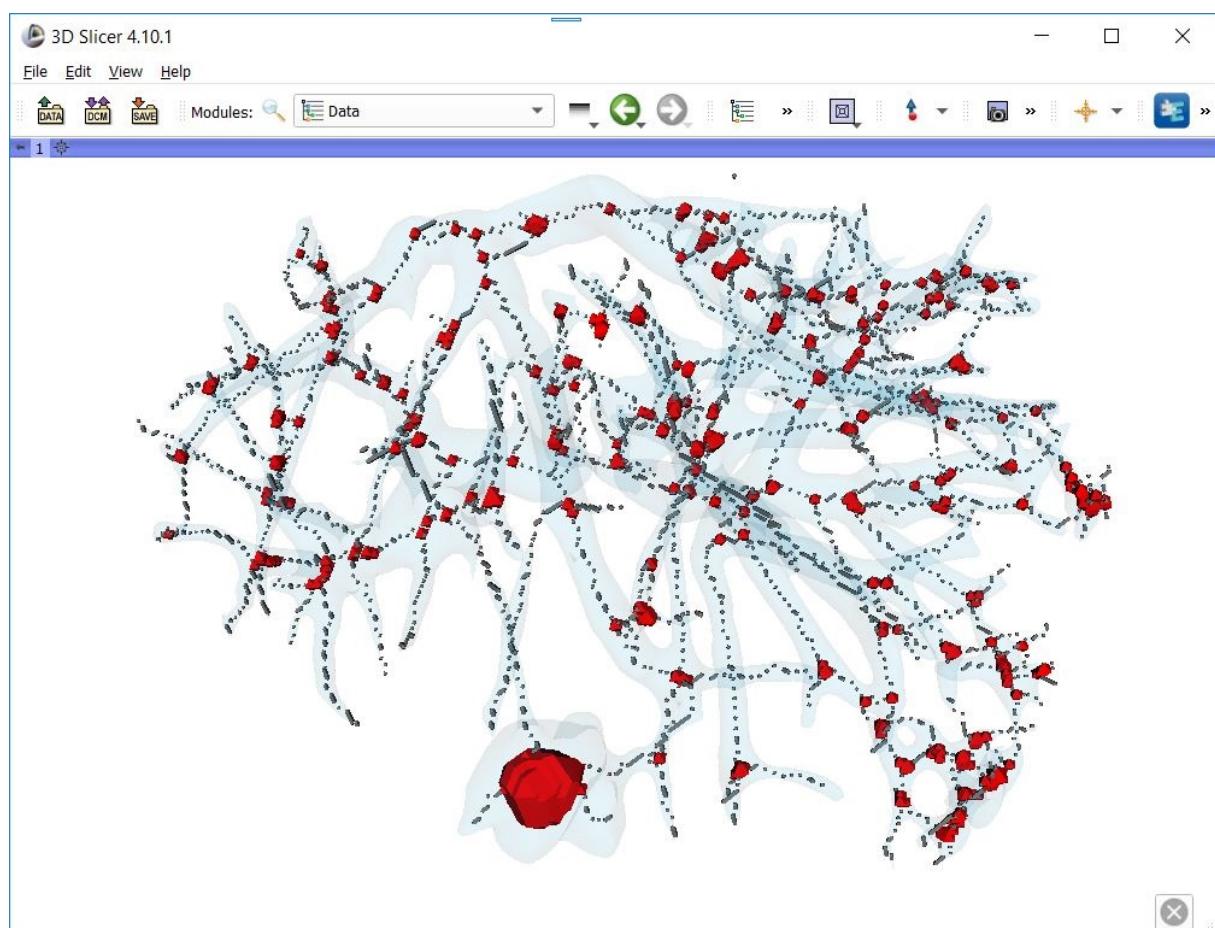


Figure 7.6: Kết quả tìm đường chính giữa (đường màu xám) và điểm phân nhánh (màu đỏ) của trường hợp xấu nhất trong thí nghiệm 2. Hệ thống đã xác định được đường chính giữa và các điểm phân nhánh của mạch máu. Tuy nhiên, xuất hiện rất nhiều điểm phân nhánh sai.

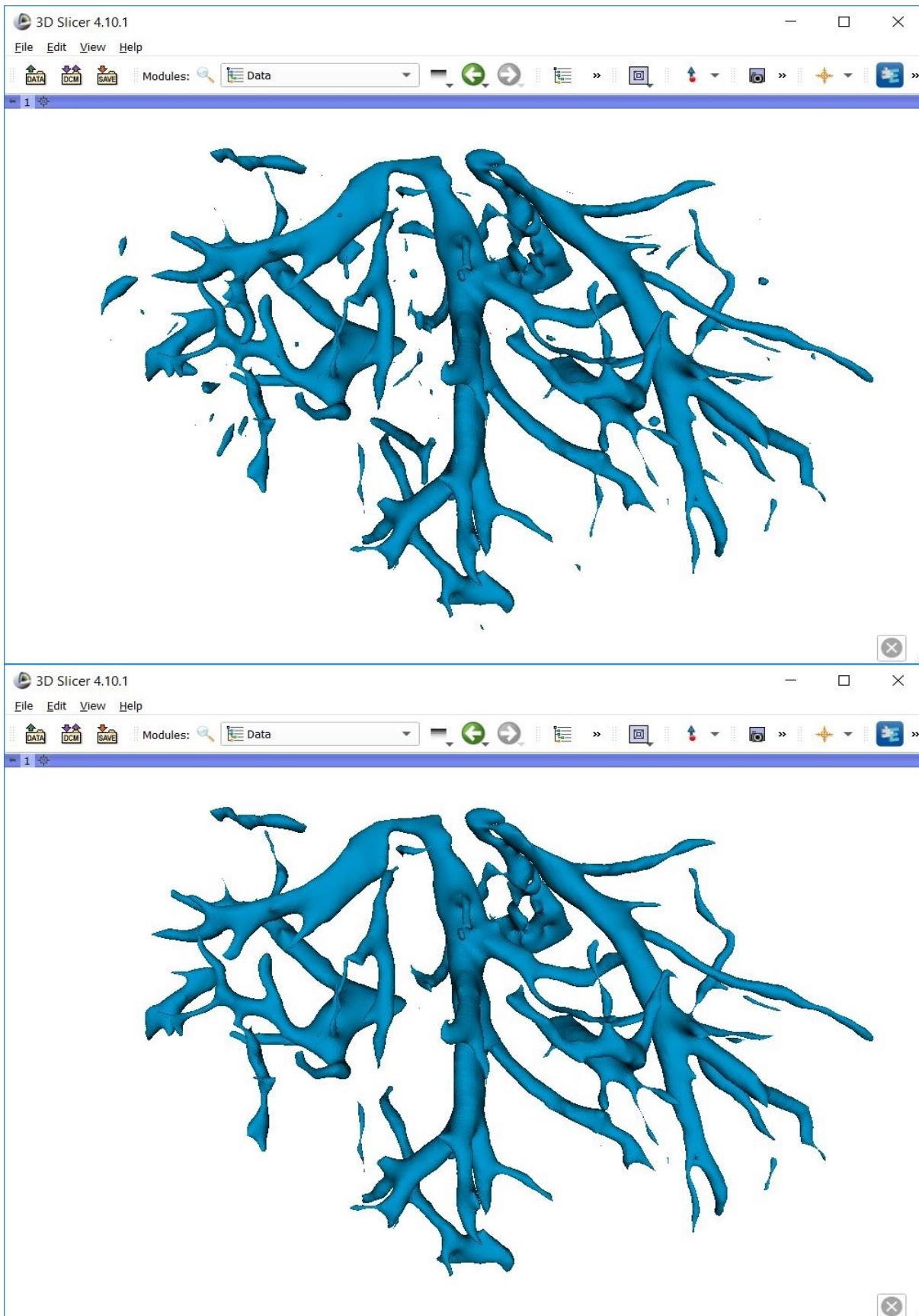


Figure 7.7: Kết quả hệ thống mạch máu của trường hợp tốt nhất trong thí nghiệm 8 trước và sau hậu xử lý. Các thành phần nhỏ và rời rạc hầu như đã được loại bỏ hoàn toàn.

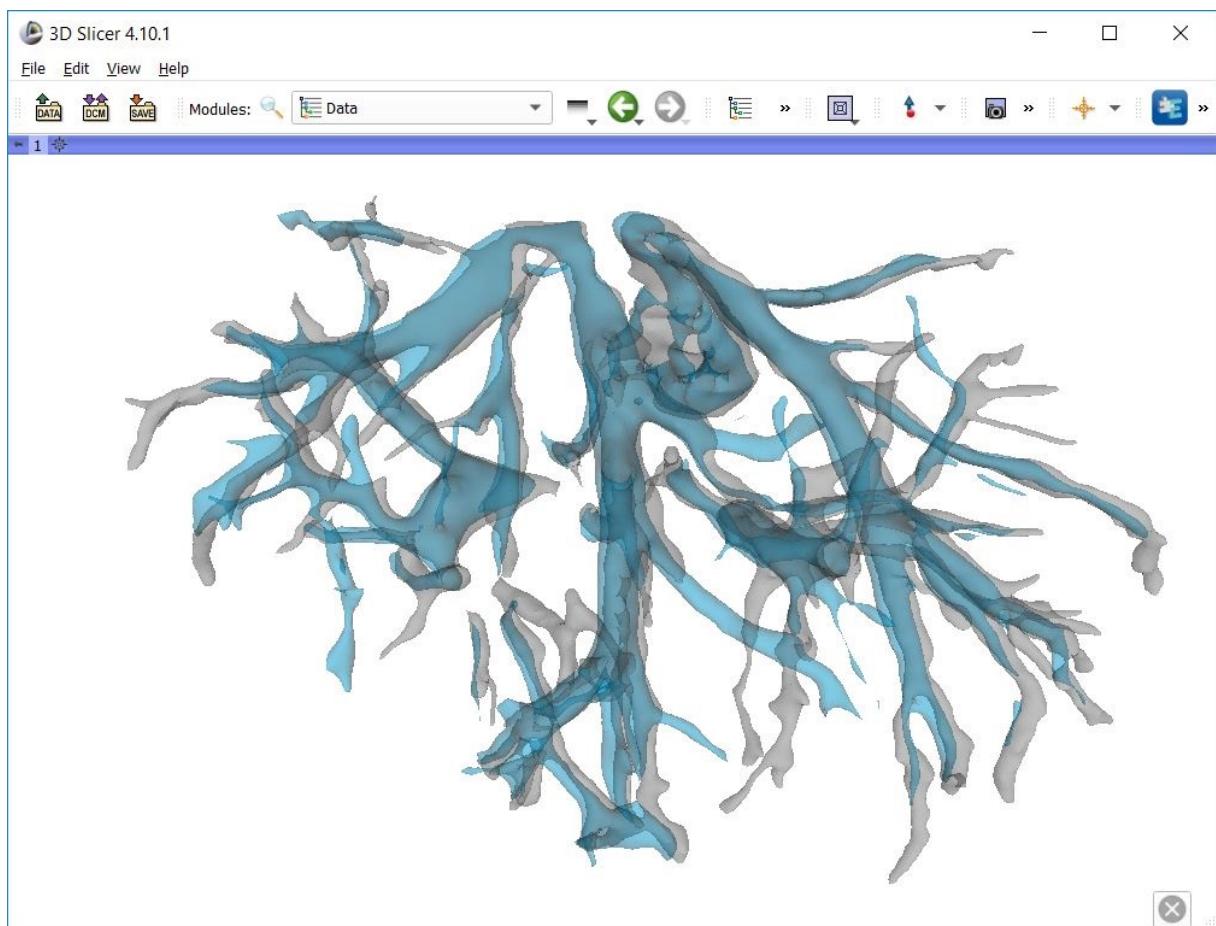


Figure 7.8: Kết quả hệ thống mạch máu (màu xanh) của trường hợp tốt nhất trong thí nghiệm 8 khi so sánh với nhãn phân đoạn (màu xám). Kết quả phân đoạn khá sát với thực tế, chỉ một vài nhánh nhỏ chưa phân đoạn được.

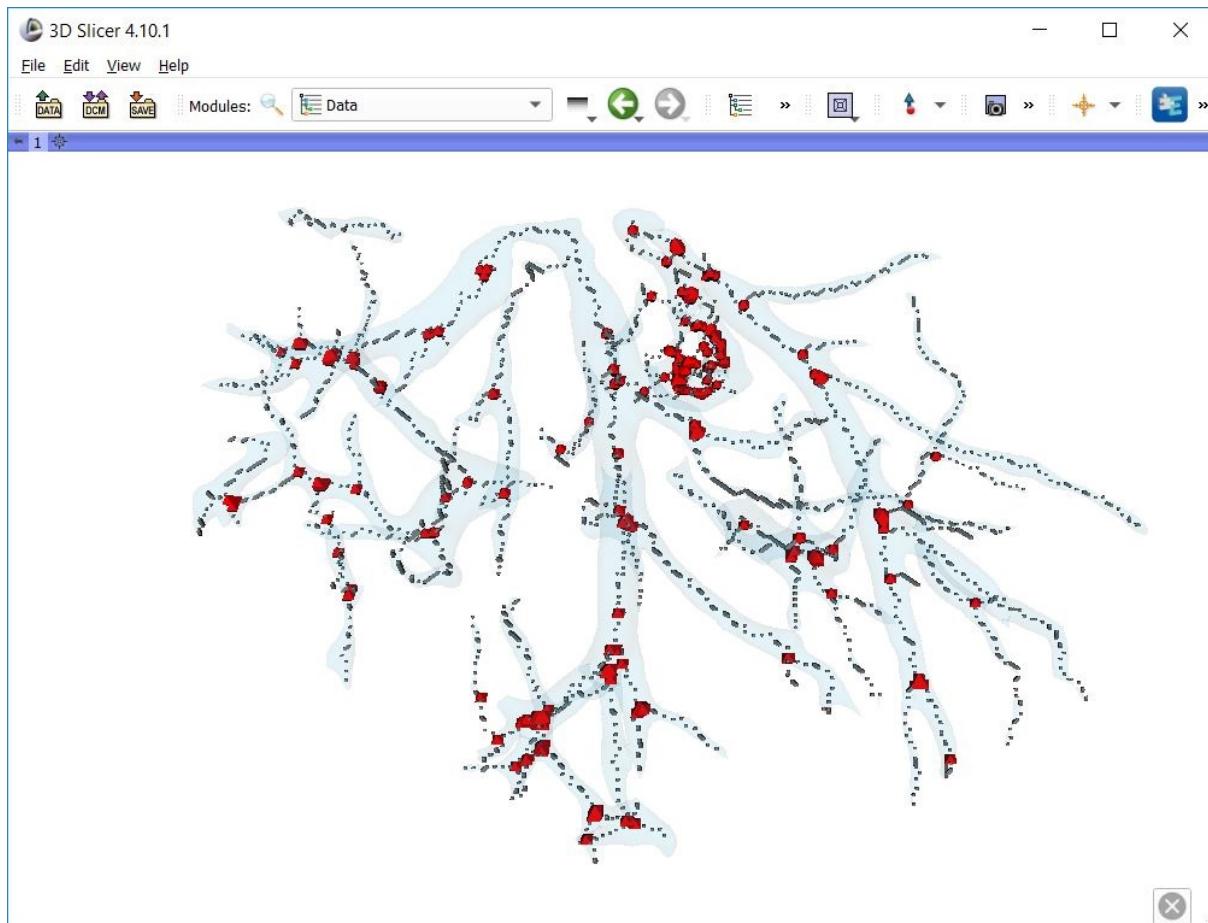


Figure 7.9: Kết quả tìm đường chính giữa (đường màu xám) và điểm phân nhánh (màu đỏ) của trường hợp tốt nhất trong thí nghiệm 8. Hệ thống đã xác định được đường chính giữa và các điểm phân nhánh của mạch máu.

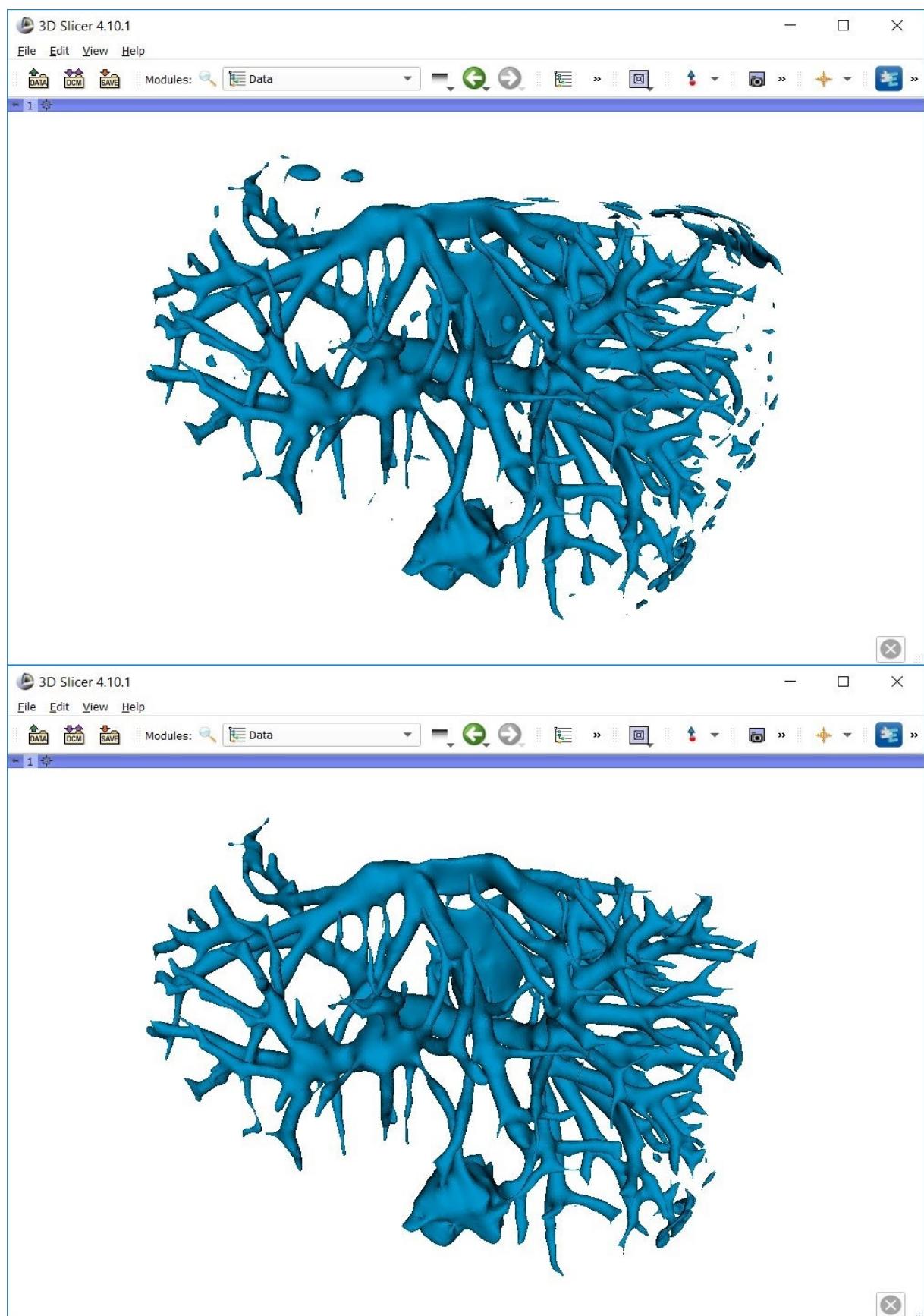


Figure 7.10: Kết quả hệ thống mạch máu của trường hợp xấu nhất trong thí nghiệm 8 trước và sau hậu xử lý. Bước hậu xử lý đã loại bỏ đi rất nhiều vị trí dự đoán sai.

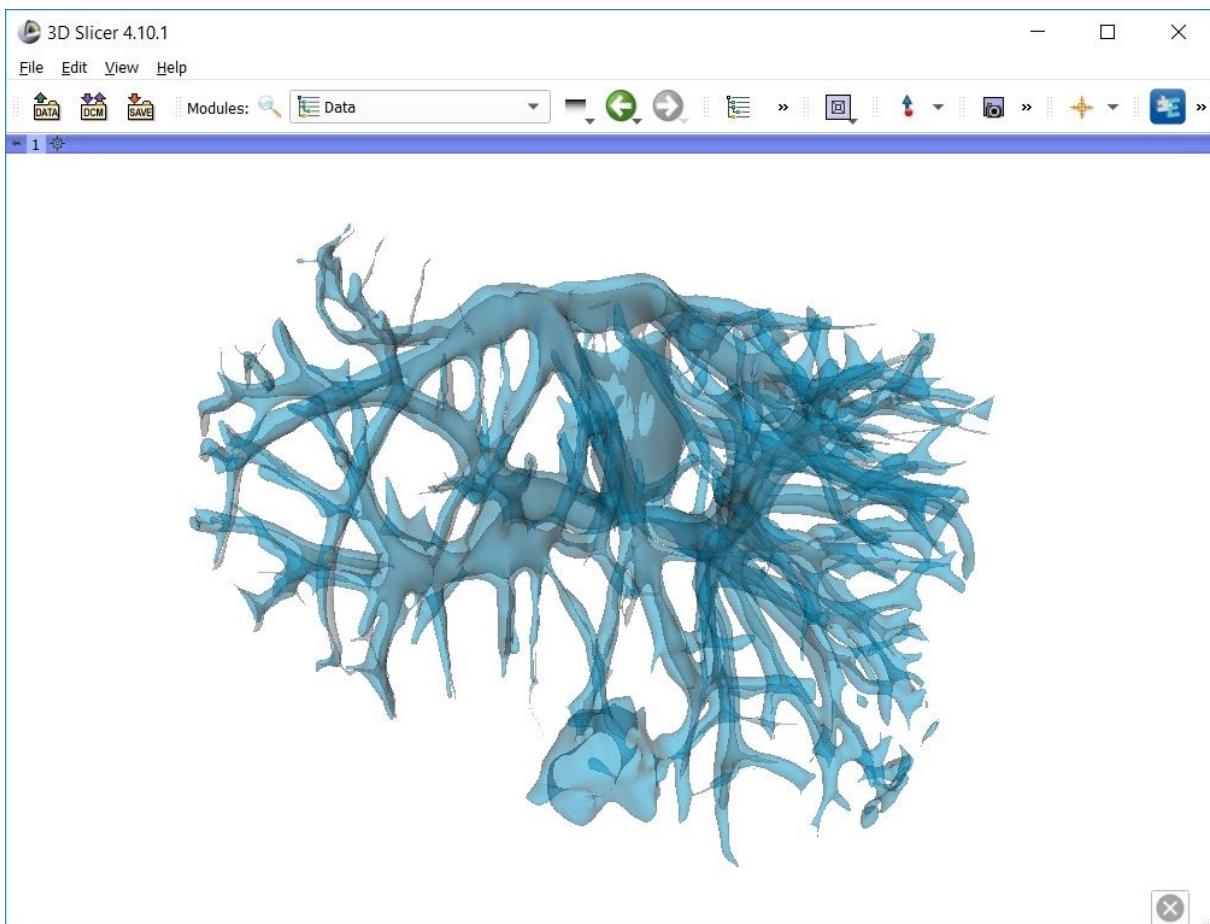


Figure 7.11: Kết quả hệ thống mạch máu (màu xanh) của trường hợp xấu nhất trong thí nghiệm 8 khi so sánh với nhãn phân đoạn (màu xám). Các mạch máu được phân đoạn lớn hơn rất nhiều so với thực tế. Ngoài ra, xuất hiện một khối phân đoạn sai lớn có thể là khối u.

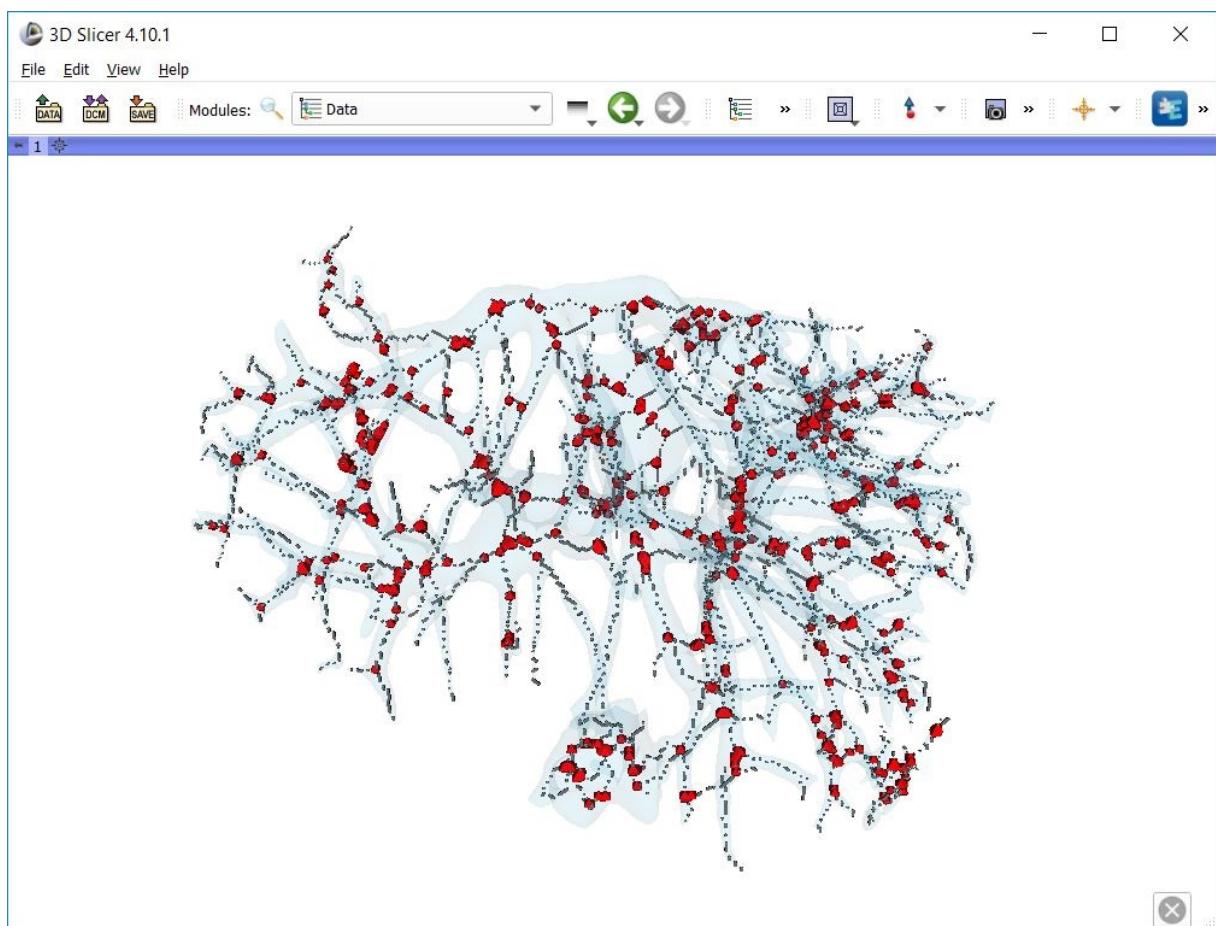


Figure 7.12: Kết quả tìm đường chính giữa (đường màu xám) và điểm phân nhánh (màu đỏ) của trường hợp xấu nhất trong thí nghiệm 8. Hệ thống đã xác định được đường chính giữa và các điểm phân nhánh của mạch máu. Tuy nhiên, xuất hiện rất nhiều đường chính giữa và điểm phân nhánh sai.

8

CONCLUSION

This chapter summarizes the results for the thesis until now. Finally, we want to present a working plan to improve the thesis.

Table of Contents

8.1	Kết quả đạt được	68
8.2	Hạn chế và hướng phát triển	69

8.1 Kết quả đạt được

Kết thúc giai đoạn Luận văn tốt nghiệp, chúng tôi đã được trang bị thêm các kiến thức trong lĩnh vực Xử lý ảnh nói chung, cũng như các phương pháp tiếp cận trong lĩnh vực Phân đoạn hình ảnh y khoa nói riêng. Chúng tôi đã hiện thực được một hệ thống xây dựng hệ thống mạch máu của cơ quan gan từ ảnh chụp CT.

Hệ thống mà chúng tôi xây dựng đã đáp ứng được đầy đủ các yêu cầu trong bài toán đặt ra, bao gồm:

- Phân đoạn được hệ thống mạch máu của cơ quan gan từ ảnh chụp CT sử dụng mạng học sâu. Kết quả phân đoạn trên tập kiểm tra có độ chính xác theo độ đo Dice là 0.569 và kết quả trực quan rất khả quan.
- Xác định được đường chính giữa của mạch máu sau kết quả phân đoạn thông qua đề xuất thực hiện trích xuất khung xương đối tượng bằng phép toán skeleton trong không gian 3D.
- Xác định được điểm phân nhánh của mạch máu thông qua đề xuất xem xét số lượng điểm ảnh liền kề cũng thuộc đường chính giữa mạch máu của từng điểm ảnh trong đường chính giữa mạch máu.

Bên cạnh các kết quả nêu trên, chúng tôi còn có hai đóng góp quan trọng liên quan đến tiền xử lý và hậu xử lý dữ liệu. Những đề xuất này đã giúp kết quả phân đoạn của hệ thống cải thiện đáng kể.

Về tiền xử lý dữ liệu, chúng tôi đề xuất sử dụng phương pháp nội suy để lấp các lớp CT bị thiếu trong khối ảnh CT. Bên cạnh đó, chúng tôi đề xuất thực hiện biến đổi độ sáng điểm ảnh thông qua phương pháp đặt ngưỡng giới hạn và hàm bình phương giá trị điểm ảnh giúp hình ảnh mạch máu hiển thị rõ hơn. Từ đó, kết quả học trứ nên tốt hơn.

Về hậu xử lý dữ liệu, chúng tôi đã cải thiện được kết quả phân đoạn thông qua đề xuất xác định các thành phần liên thông và loại bỏ các thành phần liên thông nhỏ không thuộc mạch máu trong kết quả phân đoạn.

Với những kết quả có được của luận văn này, chúng tôi hy vọng có thể đóng góp một phần vào các công trình nghiên cứu trong lĩnh vực Phân đoạn hình ảnh y khoa về sau. Sớm đưa được các công trình nghiên cứu vào ứng dụng thực tiễn giúp cải thiện chất lượng trong công tác y tế cũng như cuộc sống xã hội.

8.2 Hạn chế và hướng phát triển

Hạn chế. Bên cạnh các kết quả đạt được, chúng tôi cũng gặp các khó khăn trong quá trình hiện thực hệ thống. Nếu các vấn đề này sớm được khắc phục rất có thể kết quả hệ thống sẽ tốt hơn nữa.

- Tập dữ liệu quá nhỏ dẫn tới quá trình huấn luyện không thực sự tốt. Trong nhiều trường hợp hệ thống nhanh chóng rơi vào overfitting trong khi chưa học được toàn diện các đặc trưng của hệ thống mạch máu trong ảnh CT.
- Hệ thống vẫn phải đặt trên giả thiết nhãn phân đoạn cơ quan gan đã được cung cấp do nhãn phân đoạn động mạch ở bên ngoài cơ quan gan không được cung cấp ở một số bệnh nhân trong bộ dữ liệu.
- Hệ thống chưa thực sự dễ sử dụng. Các công đoạn của hệ thống vẫn phải thực hiện một cách riêng lẻ thông qua sự điều khiển, chưa xây dựng được một ứng dụng end-to-end¹.

Hướng phát triển. Không thể nào phủ nhận sự cần thiết của một hệ thống chẩn đoán thông minh trong lĩnh vực y tế, đặc biệt là một hệ thống cho phép cải thiện khả năng nhìn và phán đoán về hình ảnh y khoa. Chính vì vậy, việc mở rộng và phát triển hệ thống mà chúng tôi đã xây dựng mở ra nhiều hướng mới trong tương lai. Do giới hạn về thời gian thực hiện luận văn nên nhiều ý tưởng chúng tôi vẫn chưa thực hiện được.

Thứ nhất, chúng tôi muốn phát triển hệ thống để có thể phân đoạn hệ thống mạch máu trên toàn ảnh CT. Xây dựng hệ thống mạch máu một cách toàn diện không chỉ giúp nhìn thấy hệ thống mạch máu bên trong cơ quan mà còn thấy được sự tương quan của cơ quan gan và các cơ quan khác trong cơ thể.

Thứ hai, chúng tôi muốn hệ thống không chỉ phân đoạn hệ thống mạch máu và còn phân đoạn được các cơ quan khác nhau cũng như các tổn thương bất thường trong cơ thể. Xác định chính xác vị trí tổn thương giúp quá trình chữa trị có thể tiến hành thuận lợi hơn.

Thứ ba, chúng tôi muốn khai thác các thông tin đặc thù trong lĩnh vực y khoa. Tìm hiểu mối liên quan giữa kết quả chẩn đoán và quá trình điều trị. Từ đó, giúp phát triển một hệ thống có khả năng đưa ra phác đồ điều trị từ kết quả dự đoán.

Cuối cùng, chúng tôi muốn xây dựng giao diện người dùng cho hệ thống để đưa hệ thống đến với người dùng. Chúng tôi mong muốn quá trình chẩn đoán thông qua hình ảnh y khoa được thực hiện một cách nhanh chóng và chính xác thay vì phương pháp thủ công trước đây.

¹ End-to-end là khái niệm dùng để chỉ một hệ thống có khả năng tiếp nhận dữ liệu thô ở đầu vào và cho ra kết quả cuối cùng ở đầu ra, mọi công đoạn trung gian được thực hiện một cách tự động.

A

THÔNG SỐ TẬP DỮ LIỆU 3D-IRCADB-01

Tables A.1 trình bày thông số chi tiết về tập dữ liệu 3D-IRCADB-01. Bảng thông số này được lấy từ trang chủ của Viện nghiên cứu chống ung thư đường tiêu hoá IRCAD.

Tables A.1: Thông tin về tập dữ liệu 3D-IRCADB-01 (Nguồn: [9]).

STT	Giới tính	Kích thước voxel (mm)	Kích thước ảnh (pixel)	Kích thước gan (cm)	Bệnh lý gan	Trở ngại phân đoạn
1	Nữ	0,57	512	18,3	7 khối u (III, IV/V, VII, VIII)	Dạ dày
		0,57	512	15,1		Tụy
		1,6	129	14,1		Tá tràng
2	Nữ	0,78	512	20,1	1 khối u (V)	Tụy
		0,78	512	16,9		
		1,6	172	15,7		Tá tràng
3	Nam	0,62	512	16,7	1 khối u (IV)	Kim loại
		0,62	512	14,9		
		1,25	200	15,2		
4	Nam	0,74	512	16,9		
		0,74	512	12,0	7 khối u	Tim
		2	91	17,2		
5	Nam	0,78	512	19,8		Cơ hoành
		0,78	512	18,6	0 khối u	
		1,6	139	19,1		Tá tràng
6	Nam	0,78	512	18,8		
		0,78	512	14,3	20 khối u	Tim
		1,6	135	20,2		
7	Nam	0,78	512	24,9		
		0,78	512	15,2	0 khối u	Lá lách
		1,6	151	16,6		
8	Nữ	0,56	512	23,5		
		0,56	512	17,1	3 khối u (I, II, IV)	
		1,6	124	12,5		–

Tables A.1: Thông tin về tập dữ liệu *3D-IRCADb-01* (Tiếp).

STT	Giới tính	Kích thước voxel (mm)	Kích thước ảnh (pixel)	Kích thước gan (cm)	Bệnh lý gan	Trở ngại phân đoạn
9	Nam	0,87	512	20,6	1 khối u ở V/VIII	Dạ dày Ruột già
		0,87	512	17,0		
		2	111	18,1		
10	Nữ	0,73	512	18,4	8 khối u	Tá tràng Tuy
		0,73	512	15,5		
		1,6	122	14,8		
11	Nam	0,72	512	19,1	0 khối u	Dạ dày Tuy
		0,72	512	14,4		
		1,6	132	16,2		
12	Nữ	0,68	512	19,3	1 khối u VI	–
		0,68	512	17,7		
		1,0	260	11,0		
13	Nam	0,67	512	20,0	20 khối u	Tá tràng Tim
		0,67	512	12,9		
		1,6	122	18,1		
14	Nữ	0,72	512	22,4	0 khối u	Lá lách Tuy
		0,72	512	15,4		
		1,6	113	13,4		
15	Nữ	0,78	512	18,8	2 khối u (II, VIII)	Tuy
		0,78	512	17,7		
		1,6	125	14,7		
16	Nam	0,70	512	20,2	1 khối u (V)	Cơ
		0,70	512	17,7		
		1,6	155	20,2		
17	Nam	0,74	512	19,8	2 khối u (II, VIII)	Dạ dày Tim
		0,74	512	17,4		
		1,6	119	18,9		
18	Nữ	0,74	512	22,5	1 khối u (IV/V)	Thực quản Cơ
		0,74	512	15,1		
		2,5	74	18,6		
19	Nữ	0,70	512	19,5	46 khối u	Dạ dày Tá tràng Tuy
		0,70	512	16,5		
		4	124	14,2		
20	Nữ	0,81	512	20,0	0 khối u	Dạ dày
		0,81	512	16,6		
		2	225	16,8		

B

KẾ HOẠCH THỰC HIỆN LUẬN VĂN

Để có thể hoàn thành luận văn một cách tốt nhất, ngay từ những ngày đầu tiên, chúng tôi đã xây dựng kế hoạch cụ thể cho giai đoạn Đề cương luận văn cũng như giai đoạn Luận văn và đã thực hiện đúng theo kế hoạch đề ra. Đó là một trong những cơ sở quan trọng để chúng tôi có thể thực hiện thành công luận văn này. Chi tiết kế hoạch thực hiện luận văn được chúng tôi mô tả ở Figure B.1.

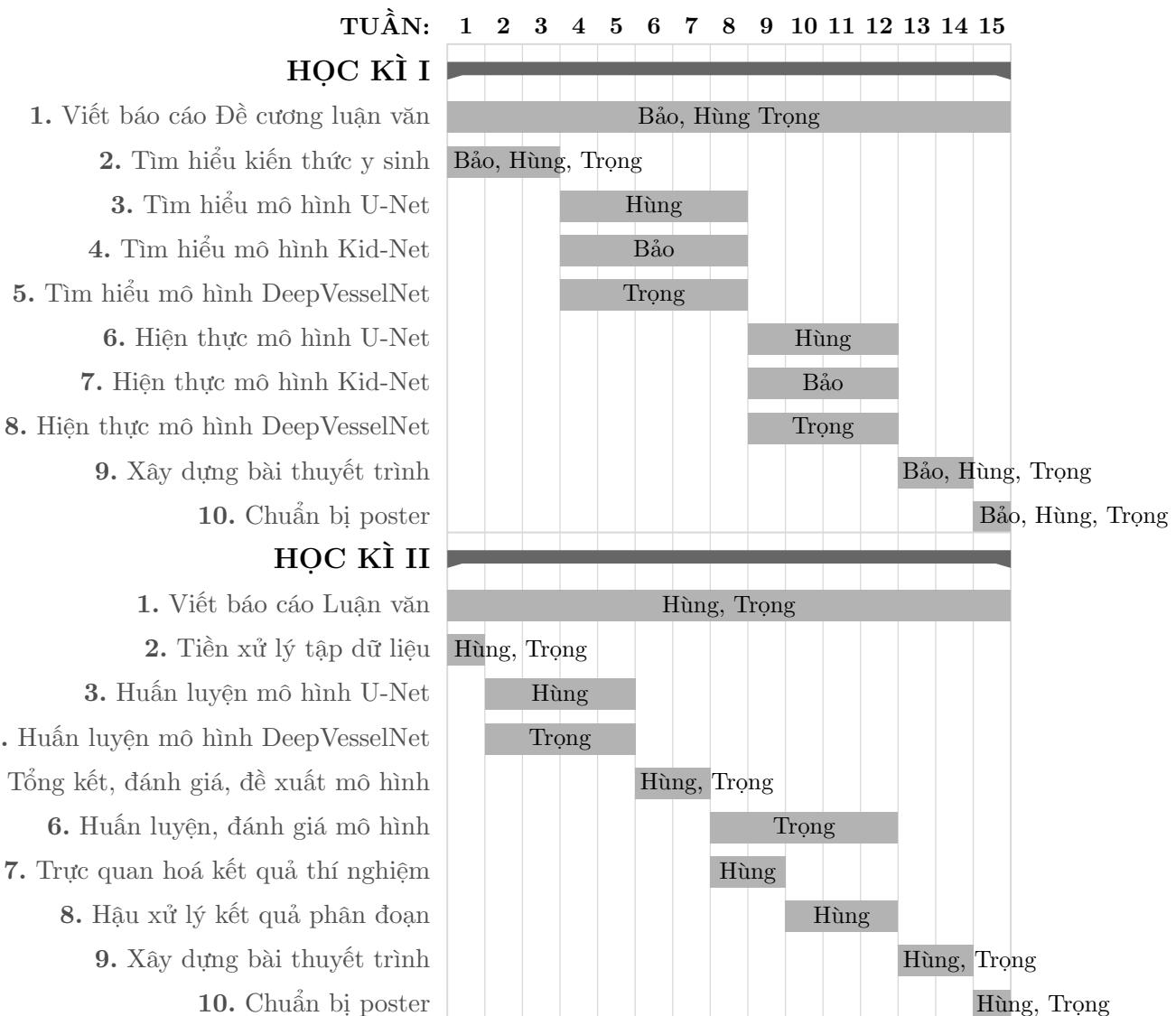


Figure B.1: Kế hoạch thực hiện luận văn.

BIBLIOGRAPHY

- [1] A. Joshi and V. Geetha, “Sql injection detection using machine learning.,” Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014 International Conference, 2014, pp. 1111–1115.
- [2] *Sqlmap*, <https://sqlmap.org/>, Accessed: 2022-09-30.
- [3] R. Rawat and S. Shrivastav, “Sql injection attack detection using svm.,” *International Journal of Computer Applications*, pp. 1–4, 2012.
- [4] F. Mereani and J. Howe, “Detecting cross-site scripting attacks using machine learning,” in Jan. 2018, pp. 200–210, ISBN: 978-3-319-74689-0. DOI: [10.1007/978-3-319-74690-6_20](https://doi.org/10.1007/978-3-319-74690-6_20).
- [5] A. Laughter, S. Omari, P. Szczurek, and J. Perry, “Detection of malicious http requests using header and url features,” in Jan. 2021, pp. 449–468, ISBN: 978-3-030-63088-1. DOI: [10.1007/978-3-030-63089-8_29](https://doi.org/10.1007/978-3-030-63089-8_29).
- [6] A. Alshammari and A. Aldribi, “Apply machine learning techniques to detect malicious network traffic in cloud computing,” *Journal of Big Data*, vol. 8, no. 1, p. 90, Jun. 2021, ISSN: 2196-1115. DOI: [10.1186/s40537-021-00475-1](https://doi.org/10.1186/s40537-021-00475-1). [Online]. Available: <https://doi.org/10.1186/s40537-021-00475-1>.
- [7] M. Alsaedi, F. A. Ghaleb, F. Saeed, J. Ahmad, and M. Alasli, “Cyber threat intelligence-based malicious url detection model using ensemble learning,” *Sensors*, vol. 22, no. 9, 2022, ISSN: 1424-8220. DOI: [10.3390/s22093373](https://doi.org/10.3390/s22093373). [Online]. Available: <https://www.mdpi.com/1424-8220/22/9/3373>.
- [8] L. N. M. Khoi, “Machine learning approaches to cyber threats detection.,” Jan. 2021.
- [9] IRCAD France. “3d-ircadb 01.” (2009), [Online]. Available: <https://www.ircad.fr/research/3d-ircadb-01/> (visited on 02/03/2019).
- [10] Medical Connections. “Hounsfield units.” (2010), [Online]. Available: <https://www.medicalconnections.co.uk/kb/Hounsfield-Units/> (visited on 03/02/2019).
- [11] D. V. Duy, L. T. Sach, and E. K. Karuppiah, “Medical images sequence normalization and argumentation: Improve liver tumor segmentation from small dataset,” 2017.
- [12] B. H. T. Nhat and P. H. Son, “Apply deep learning in tumor detection in ct images,” Engineer’s Thesis, Ho Chi Minh City University of Technology, 2019, pp. 50–51.

DANH MỤC TỪ KHOÁ

Danh sách dưới đây liệt kê các từ khoá được đề cập trong nội dung luận văn này theo thứ tự bảng chữ cái.

B	Left limit	41, 42
Background	24, 25, 42, 49	
Batchnorm	50, 51	
C		
Convolution	40	
F		
Foreground	24, 25, 42, 49	
H		
Hounsfield	21	
I		
Inference	50	
Intensity	21	
L		
Learning rate	50–53	
O		
Overfitting	69	
P		
PyTorch	44	
R		
Right limit	41, 42	
S		
Skeleton	68	
Slicer	46	
V		
VTK	46	