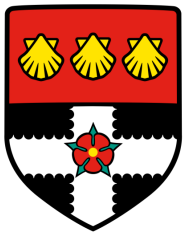


University of Reading

Department of Computer Science

An AI-assisted decision making system for thyroid nodule classification

Stefanos Stefanou

Supervisor: Huizhi Liang

A report submitted in partial fulfilment of the requirements of
the University of Reading for the degree of
Bachelor of Science in *Computer Science*

April, 2021

Declaration

I, Stefanos Stefanou, of the Department of Computer Science, University of Reading, confirm that all the sentences, figures, tables, equations, code snippets, artworks, and illustrations in this report are original and have not been taken from any other person's work, except where the works of others have been explicitly acknowledged, quoted, and referenced. I understand that if failing to do so will be considered a case of plagiarism. Plagiarism is a form of academic misconduct and will be penalised accordingly.

Stefanos Stefanou
April, 2021

Abstract

Deep learning has found numerous applications in the health care community. Recently, a massive explosion of research on the relevant field, driven by large amounts of available data, has generated important disease prevention and identification results. Fine Needle Aspiration (FNA) is the dominant procedure for thyroid nodule classification. FNA has associated risks and expenses, and in this project, we will try to reduce both using the recent advancements in Artificial Intelligence and Deep Learning. Our primary goal is to bring closer the radiologists 'on the field' with those complex algorithms and provide value to real patients by providing an interface, in the form of a web application, for probabilistically predicting the severity and the category of a given module.

Contents

Acknowledgments

Acknowledgments section is optional. You may like to acknowledge the support and help of your supervisor(s), friend(s), or any other person(s), department(s), institute(s).

Chapter 1

Introduction

Deep learning has found numerous applications in the health care community. Recently, a massive explosion of research on the relevant field, driven by large amounts of available data, has generated important disease prevention and identification results.[Eun Ju Ha (2021)]

The Dominant process for thyroid nodule identification and classification is called FNA (or Fine Needle Aspiration/Biopsy). FNA is an expensive process requiring expensive lab equipment and specialized personnel.[*Fine Needle Aspiration Biopsy of Thyroid Nodules* (2019)]

There is no way to determine the category of a thyroid nodule apart from performing FNA on a sample. Our vision is to create an application to act as a bridge between the academic community working on theoretical Deep Learning models to predict a nodule's category and the radiologists working with actual patients and accurate data. Our hope is that by establishing a common language(the application) we will improve the research process as experimental models will work on accurate nodule scans, providing instant feedback to the researchers for further analysis.

Our system needs to be as generic as possible to support any prediction model, reliable, easy to maintain, and expand. It needs to be optimized to handle the Deep Learning models and finally needs to be as secure as possible because it will eventually work with actual patients on sensitive data.

Abbreviations

FNA(Fine Needle Aspiration), AI(Artificial Intelligence), DP(Deep Learning)

Keywords

FNA, AI, DP

Chapter 2

Literature Review

2.1 Introduction

This section will note the essential sources needed to be studied and to be revised to complete this project. The sources are carefully selected to include theoretical, practical, and best practices knowledge in order to cover the wide variety of topics needed to fulfill the requirements of this project.

2.2 Brief Table of books

ISBN	Name	Type
N/A	ST1PS-18-9A: Probability and Statistics (2018/19)	Module Lectures
9780030105678	Linear Algebra and Its Applications	Book
9780131687288	Digital Image Processing	Book
9780262035613	Deep Learning	Book
9780128104088	Deep Learning for Medical Image Analysis	Book
9781491962244	Hands-on machine learning with scikit-learn and tensorflow	Book

2.3 Brief Table of papers

- Ye, H., Hang, J., Chen, X. et al. An intelligent platform for ultrasound diagnosis of thyroid nodules. Sci Rep 10, 13223 (2020). https://doi.org/10.1038/s41598-020-70159-y
- Nguyen DT, Pham TD, Batchuluun G, Yoon HS, Park KR. Artificial Intelligence-Based Thyroid Nodule Classification Using Information from Spatial and Frequency Domains. J Clin Med. 2019;8(11):1976. Published 2019 Nov 14. doi:10.3390/jcm8111976
- Manivannan T, Ayyappan N. Classification of thyroid nodules using ultrasound images. Bioinformation. 2020;16(2):145-148. Published 2020 Feb 29. doi:10.6026/97320630016145
- Nguyen DT, Kang JK, Pham TD, Batchuluun G, Park KR. Ultrasound Image-Based Diagnosis of Malignant Thyroid Nodule Using Artificial Intelligence. Sensors (Basel). 2020;20(7):1822. Published 2020 Mar 25. doi:10.3390/s20071822
- Chen J, You H, Li K. A review of thyroid gland segmentation and thyroid nodule segmentation methods for medical ultrasound images. Comput Methods Programs

Biomed. 2020 Mar;185:105329. doi: 10.1016/j.cmpb.2020.105329. Epub 2020 Jan 9. PMID: 31955006.

- Ha EJ, Baek JH. Applications of machine learning and deep learning to thyroid imaging: where do we stand? Ultrasonography. 2021 Jan;40(1):23-29. doi: 10.14366/usg.20068. Epub 2020 Jul 3. PMID: 32660203; PMCID: PMC7758100.

Chapter 3

Requirement Analysis

3.1 Introduction

Before we even start exploring this project and its features, it is important to strictly define the requirements need to be fulfilled and the scope of this project. failing to perform an requirement analysis beforehand puts additional and unsessesary risks to the project due to the varying scope and target set.

Chapter 4

Entity Relation Analysis

4.1 Introduction

After the requirements has been set. We need to traslate them into workable relational entities, in order to be able to modelled though a classical relational database system (RDBMS).

4.2 Entities

We will start our exploration by defining our entities for this project.

4.2.1 Scan

A scan is the result of an ultrasound scan performed in a specific patient(see [4.2.2)]. A scan entity has certain attributes

Image	The image produced by the ultrasound scan. 360x560 pixels
Prediction	The result of the prediction algorithm. Acceptable Values = Maligrant, Benign
Results	The logs of the algorithm performed the prediction, Optional
Algorithm	The algorithm used to perform the prediction. Acceptable Values = SVC,RES
Token	The scan identifier across the application services. token type is UUID [Leach (2005)]

4.2.2 Patient

A patient is a physical person that is suspected to have a thuroid nodule. A person may have 0 up to n scans, where n is the theoretical maximum number of records(no limit is enforced by the database or the application). A patient has characteristics explaned below

First Name	The first name of the patient.
Last Name	The last name of the patient
NiNo	The National Insurance Number(NiNo) of the patient
Enrolled Date	The Date that the patient was registered in the system
Ascosiate Doctor	The Doctor identification number, handling the case of the patient(see 4.2.3)
Comments	The Doctors(see 4.2.3) comments for the particular patient

4.2.3 Doctor

A doctor is a physical person with access on the system. Is the end-user of the system and has rights of uploading ultrasound image scans and retrieve predictions for those scans. It can also provide feedback to the system for a given prediction to be used for further research and develepoment. A doctor has specific characteristics presented below.

Username	Plain-text username
Password	MD5 Hashed[Rivest (1992)] and salted[Manber (1996)] password
First Name	National Insurance Number[*National Insurance Manual* (2021)]
Last Name	The date that the patient was registered in the system
Title	The title of the doctor.
Enrolled Date	The date and time of the user enrolled to the system
Last Seen	The date and time of last login of the user
Online Status	The status of the user, acceptable values are Connected,Not Connected
Tasks	The number of scans uploaded by the user

4.2.4 Notification

A notification is a short message from the system to the end-user(The doctor). Its sole purpose is to inform the user about various events that may interest the end-user. An example of this may be that the scan results for a given scan task are ready to view. A notification has specific characteristics witch are displayed and explained below.

Message	The message in question
Ascosiated Doctor	The receipient doctor identification number
Created Date	The Date and Time where the event in question where happened

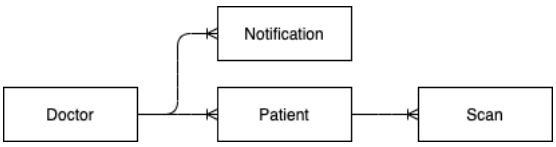
4.3 Entity Relations

The aforementioned entities have well defined relations. An exhaustive list is given below

- A doctor has many patients (1 - $+\infty$)
- A doctor has many notifications(1 - $+\infty$)
- A Patient has many Scans(1 - $+\infty$)

A above relations can be summarized in the following E-R[1] Diagram

Figure 4.1: Entity-Relation Diagram



[1] Entity-Relation

Chapter 5

Users Perspective

5.1 Introduction

In this section, we will start our exploration of the application and its infrastructure and features. As the nature of the requirements of this the system is complicated. Unavoidably the system will be complex as well. Taking this into a consideration, we will follow a natural top-to-bottom approach explaining its internals, starting as our end-users and seeing the system as a black box. In this section, we will analyze its functionality from the user's perspective. Our definitions will be rigorous as we will use UML Case diagrams for that purpose. on

5.2 Our Users

In Chapter we briefly mentioned our system's purpose as ...

Definition 1. *An AI-assisted decision making system for thyroid nodule classification.*

As we are going to see later, various scientific methods are applied, taking into account multiple parameters, in order to produce a probabilistic result. This fact implies that the end-user will be an expert on Radiology, to understand the terms, and carefully interpret the results. From now on we assume that our system users will be experianced Radiologists

5.3 Our Users

In Chapter we briefly mentioned our system's purpose as ...

Definition 2. *An AI-assisted decision making system for thyroid nodule classification.*

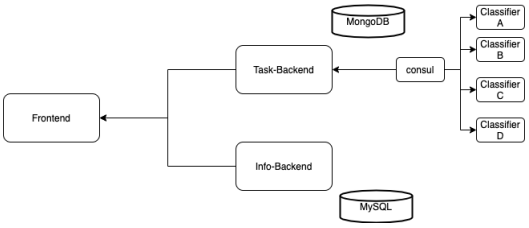
As we are going to see later, various scientific methods are applied, taking into account multiple parameters, in order to produce a probabilistic result. This fact implies that the end-user will be an expert on Radiology, to understand the terms, and carefully interpret the results. From now on we assume that our system users will be experianced Radiologists

Chapter 6

Abstracted View of the System

In this chapter, we will introduce the architecture of our system, explaining the essential elements that is composed of, and their interactions.

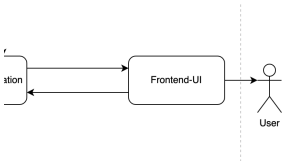
Figure 6.1: Essential System Components



6.1 Frontend Web App

The Frontend component has the responsibility of being the edge in our system.

Figure 6.2: Our system's edge



Every action from our users, should be channeled through the frontend.Our frontend is a web-based application, and as an consequence of that, a design desision is that the API between the web app and the backend application will not be a public one. This desision will increase the security of our system, as the proccess of writing spam bots will be significantly harder without a known API. More information about the API(Application programming interface) will be given below.

6.2 Backend

There is a number of design choices that we have made on our Backend System, in order to increase security, and decrese complexity. Our Backend System follows the design principles of the microservice pattern. Microservice pattern tries to decrease complexity and increase safety by splitting the internal logic of a system into several components called 'Microservices'.

Each microservice is essentially a server that handles a small portion of the systems logic. As opposed to the monolithic services, microservices have a number of advantages such as

- Highly maintainable and testable
- Loosely coupled
- Independently deployable
- No Single Point of Failure

6.2.1 Information Backend

The first of our services is the Information Backend. This service will have the responsibility to handle the information related to a scan, as well as its statistics and ascosiations between scans and patients. The majority of the models composing our systems will be available though this service, via a well designed API.

6.2.2 Task Backend

This microservice will have the responsibility to trigger prediction and classification tasks for our system. The whole procedure, due to its CPU Intensive nature, will have to be asyncronous and to be executed in the background. The Frontend will sent a request for a given task, and the server will have to return a token, ascosiated for that particular task. Later, The frontend may request to learn the progress of its task or its results(if completed) by using the relevant token. This design choice is unavoidable given that the HTTP protocol has embedded the notion of 'timeout', it is just impossible and impractical to wait untill a given task is complete. Another great advantage of this asyncronous design is the fact that multiple users may request Tasks without eliminating the server's resources, such as CPU time and amount of RAM available. Indepedent of the number of requests, the server will implement a queue FIFO (First-In-First-Out) strategy and it will inform its users when the task is ready to be seen.

6.2.3 Classification Backend

By using multiple classification techniques, our system will reduce the probability of an false prediction further. So one of our

Chapter 7

The Frontend

7.0.1 Technology Stack

In the construction of our system, we will need a number of open source technologies, libraries and standards to support our develepoment. An exhaustive list is given below

- HTML5
- CSS3
- Javascript
- React.Js
- Boostrap

HTML5

HTML5 is a markup language mainly used for structuring content on the World Wide Web. The its last major version(version 5.0) it is recommended by the World Wide Web Consortium (W3C). The responsible organisation WHATWG (Web Hypertext Application Technology Working Group) is a consortium of the major browser vendors(Apple, Google, Mozilla, and Microsoft)**?**.

CSS3

CSS stands for Cascading Style Sheets with an emphasis placed on "Style." While HTML is used to structure a web document, CSS comes through and specifies your document's style—page layouts, colors, and fonts are all determined with CSS**?**. We will use CSS, version 3, to make our frontend application aesthecaly pleasing and easy-to-use for our end-users.

Javascript

longside HTML and CSS, JavaScript is one of the major technologies of the World Wide Web. JavaScript makes possible interactive web pages and is an integral part of web applications.

React

React (also known as React.js or ReactJS) is an open-source, front end, JavaScript library[3] for building user interfaces or UI components. It is maintained by Facebook and a community

of individual developers and companies.[4][5][6] React can be used as a base in the development of single-page or mobile applications. However, React is only concerned with state management and rendering that state to the DOM, so creating React applications usually requires the use of additional libraries for routing.[7][8] React Router[9] is an example of such a library.

Boostrap

Bootstrap is a free and open-source CSS framework directed at responsive, mobile-first front-end web development. It contains CSS- and (optionally) JavaScript-based design templates for typography, forms, buttons, navigation, and other interface components.Bootstrap is among the most starred projects on GitHub, with more than 142,000 stars, behind freeCodeCamp (almost 312,000 stars) and marginally behind Vue.js framework.[2]

Chapter 8

The Backend

...

8.1 ...

....

8.2 ...

...

8.2.1 ...

8.3 Summary

...

Chapter 9

The Service

...

9.1 ...

....

9.2 ...

...

9.2.1 ...

9.3 Summary

...

Chapter 10

Discussion, Conclusion and Future work

...

10.1 ...

....

10.2 ...

...

10.2.1 ...

10.3 Summary

...

Chapter 11

Reflection

...

11.1 ...

....

11.2 ...

...

11.2.1 ...

11.3 Summary

...

References

Eun Ju Ha, J. H. B. (2021), 'Ha ej, baek jh. applications of machine learning and deep learning to thyroid imaging: where do we stand? ultrasonography. 2021 jan;40(1):23-29. doi: 10.14366/usg.20068. epub 2020 jul 3. pmid: 32660203; pmcid: Pmc7758100.'.
URL: *https://pubmed.ncbi.nlm.nih.gov/32660203/*

Fine Needle Aspiration Biopsy of Thyroid Nodules (2019).
URL: *https://www.thyroid.org/fna-thyroid-nodules/¿*

Leach, P. (2005), 'A Universally Unique IDentifier (UUID) URN Namespace', Internet Requests for Comments.
URL: *https://tools.ietf.org/html/rfc4122*

Manber, U. (1996), 'A simple scheme to make passwords based on one-way functions much harder to crack', *Computers and Security* **15**(2), 171–176.
URL: *https://www.sciencedirect.com/science/article/pii/016740489600003X*

National Insurance Manual (2021).
URL: *https://www.gov.uk/hmrc-internal-manuals/national-insurance-manual/nim39110*

Rivest, R. (1992), 'The MD5 Message-Digest Algorithm', Internet Requests for Comments.
URL: *https://tools.ietf.org/html/rfc1321*