Higher National School of Computer Science

Skin cancer detection and classification using machine learning

- President : BADSI Hichem
- Examiner : BEKKOUCHE Mohamed
- Supervisor : MEDDAH Ishak

Submitted by :

- BALBAL Oussama
- KHODJA Moussa

General introduction

In this presentation, we are going to present our master's and engineering theses, which revolves around skin cancer detection using machine learning algorithms. In the master's thesis we have done a comparative study of machine learning methods used for skin cancer detection and classification and in the engineering thesis we will present a web based platform for melanoma skin caner detection using a CNN model.

Plan of the presentation

Master

- Introduction
- Goal
- Medical background
- AI background
- State of the art
- Results and discussion (our commentary and observations)
  - We provided a line of reasoning that could help future researchers and contributors
- Conclusion

Plan of the presentation

PFE

- Goal
- Background: software + hardware used + chosen development concepts
- Present the chosen CNN architecture and the final model
- After that, we gonna talk about the web platform

Plan of the presentation

- Execute our application, while showcasing the different features and use cases.
- The execution plan is going to be talked about after the presentation.

MASTER

Introduction

Skin cancer is one of the most common cancers in the world, and it can be fatal if not treated early, that is why its early diagnosis is considered to be the best treatment for it. And under the light of recent advancements in computational power and in the artificial intelligence field, it is considered to be one of the best ways for early skin cancer diagnosis.

Introduction

That is why in this article we are going to do a comparative study of recent methods and algorithms applied in skin cancer analysis, detection and classification.

Introduction

Our comparison is going to be based on different types of datasets used for training, different algorithms applied, and famous performance metrics calculated by researchers such as accuracy, specificity, AUC (area under curve) ...etc.

Goal

Better understand the problem at hand and its applied solutions, and understand some new explored ideas and challenges faced by researchers and contributors and finally this article will help new researchers to comprehend what is ahead of them and get a general view on the various applied methods before engaging and contributing to this field.

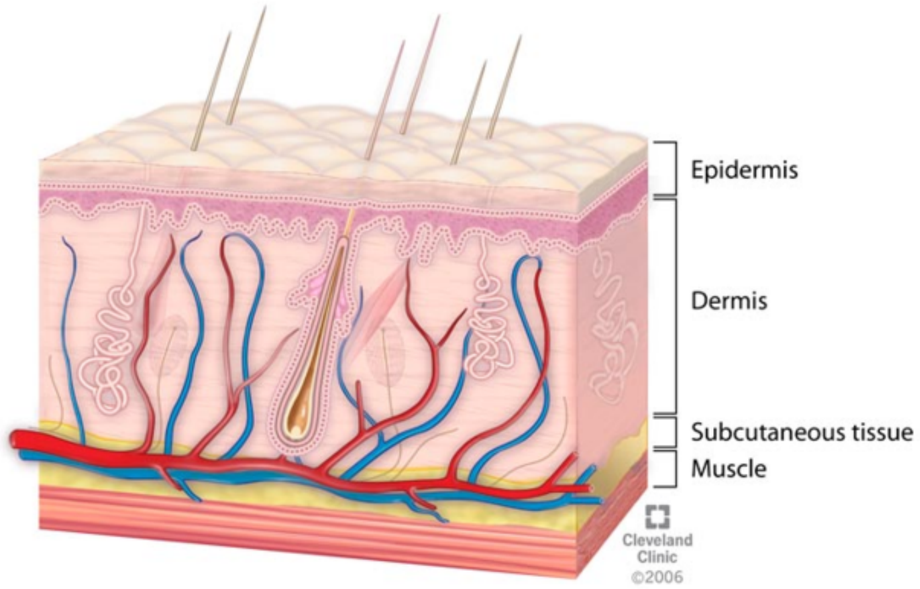
Medical background

Skin

The skin is a complex organ , it is interactive, self renewing and represents the first and primary defense line against hostile environment, and it has several characteristics and functions. It represents the largest sensory organ (15% of total body weight and a total area of 1.86 m2), it has a highly adaptive structure that makes it vital for the survival of the human body, the balance between its static and dynamic properties makes it highly adaptive to the variations of the outer world.

Skin anatomy

- **Epidermis** : regeneration, pigment
- **Dermis** : flexibility, supports epidermis
- **Hypodermis** : main support and shock absorption
- **Other entities contained in the skin** : hair and, sweat glands

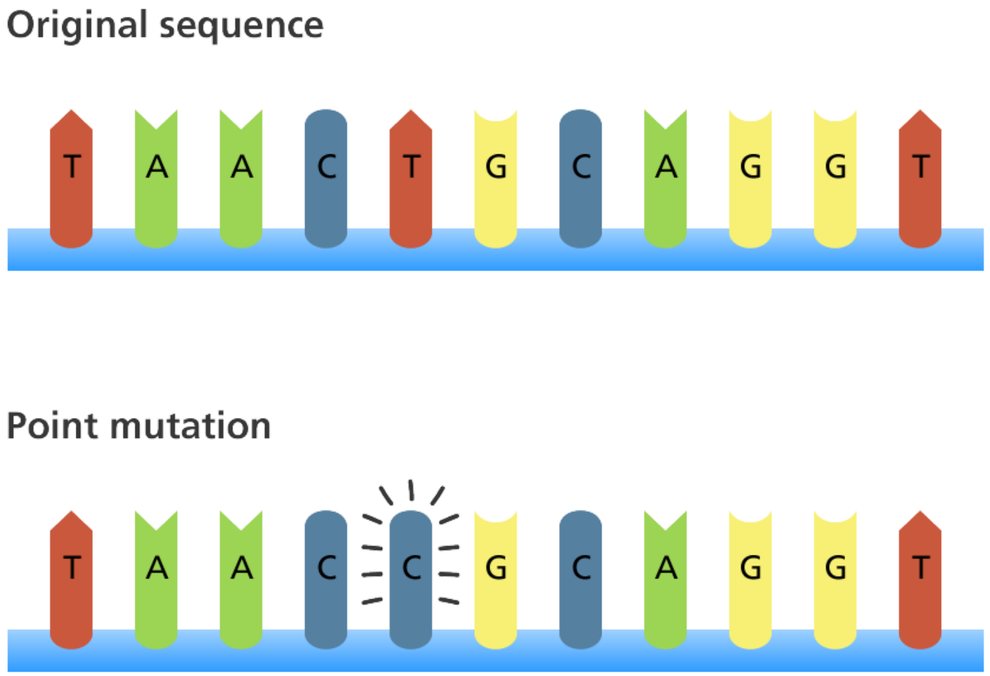


Cancer

Cancer is an illness caused by the uncontrolled division and spreading of normal cells unlike other diseases, cancer is caused by our own bodies and not by foreign entities, and it is one of the biggest causes of death among human beings nowadays, and that is because of the ineffectiveness of traditional treatment methods such as hormones, surgery, radiation, and chemotherapy. Their ineffectiveness is due to their side effects that lead the body to deteriorate more and more.

Origin

genetic mutations or rearrangements of DNA bases or various elements in our environment such as chemicals in tobacco, ultraviolet rays... etc.



Types

- benign : non fatal, doesn't spread to other organs
- malignant : fatal, spreads to other organs, hard to remove

Skin cancer

Skin cancer is the abnormal growth of cells found in the epidermis (the outer layer of the skin) , it is one of the most common cancers in the world and it falls under the category of a malignant tumor that is formed by fast multiplication of cells which is caused by mutations/damage in the DNA of those cells.

Main Causes

Caused by exposure to ultraviolet rays which can come from various sources but the most common are sunlight and tanning beds.

Most Common Types

The most common types of skin cancer are

- basal cell carcinoma(BCC)
- Squamous cell carcinoma(SCC)
- Melanoma.

The good news is that if it is discovered in an early stage or pre cancerous stage, it can be treated easily without leaving a scar.

Symptoms

It has various signs.

- Bump
- Brown scar
- Brownish spot

How doctors detect skin cancer?

- Naked eye analysis (depends on experience)
- Using a Dermatoscope
- Lab tests (biopsies)

Artificial intelligence

”Artificial intelligence (AI) is a wide-ranging branch of computer science concerned with building smart machines capable of performing tasks that typically require human intelligence.”

— https://builtin.com/artificial-intelligence

Machine learning

”Machine learning is a subset of artificial intelligence. Simply put, machine learning uses data, statistics, trial and error to “learn” a specific task without ever having to be specifically coded for the task.”

— https://builtin.com/machine-learning

types

- Supervised Learning: trained with pre-labeled data.
- Unsupervised Learning: trained with Unlabeled data, tries to find commonalities or lack of commonalities.
- Semi-Supervised Learning: middle ground, trained with both labeled/unlabeled data, it tries to find a pattern in the unlabeled data through the pattern of the labeled data.

Deep learning

”Deep learning is a subset of machine learning, where machines employ artificial neural networks to process information. Inspired by biological nodes in the human body, In essence, deep learning enables computers to do what humans do naturally, which is learning by past-experience and examples.”

— https://builtin.com/artificial-intelligence/deep-learning

Latest advancements in deep learning

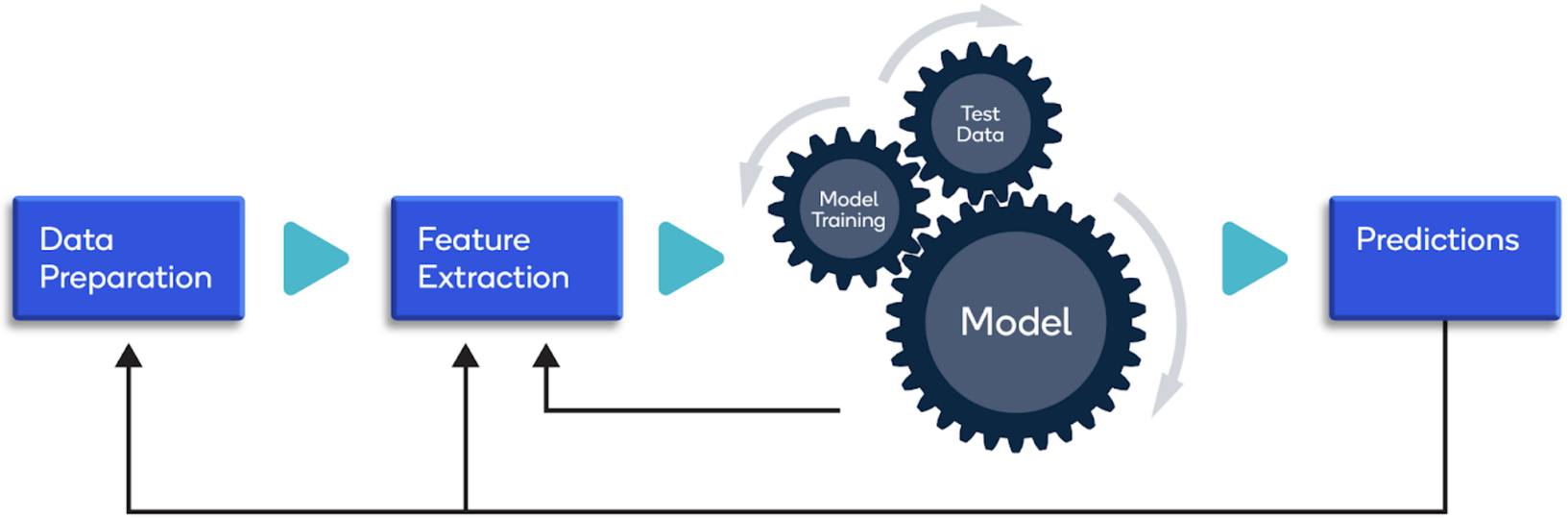
- **One shot learning and NAS (neural architecture search)**
  - One-shot : less data needed.
  - NAS : an algorithm searchs for the best neural network architecture.
- **GANS (Generative Adversarial Networks)**
  - 2 networks put against each other.
- **Auto ML**
  - All we have to do is provide data, and ML algorithms will find the design of the network.

Computer vision

Building algorithms that understand, analyses and extracts useful information from images just like us human beings.

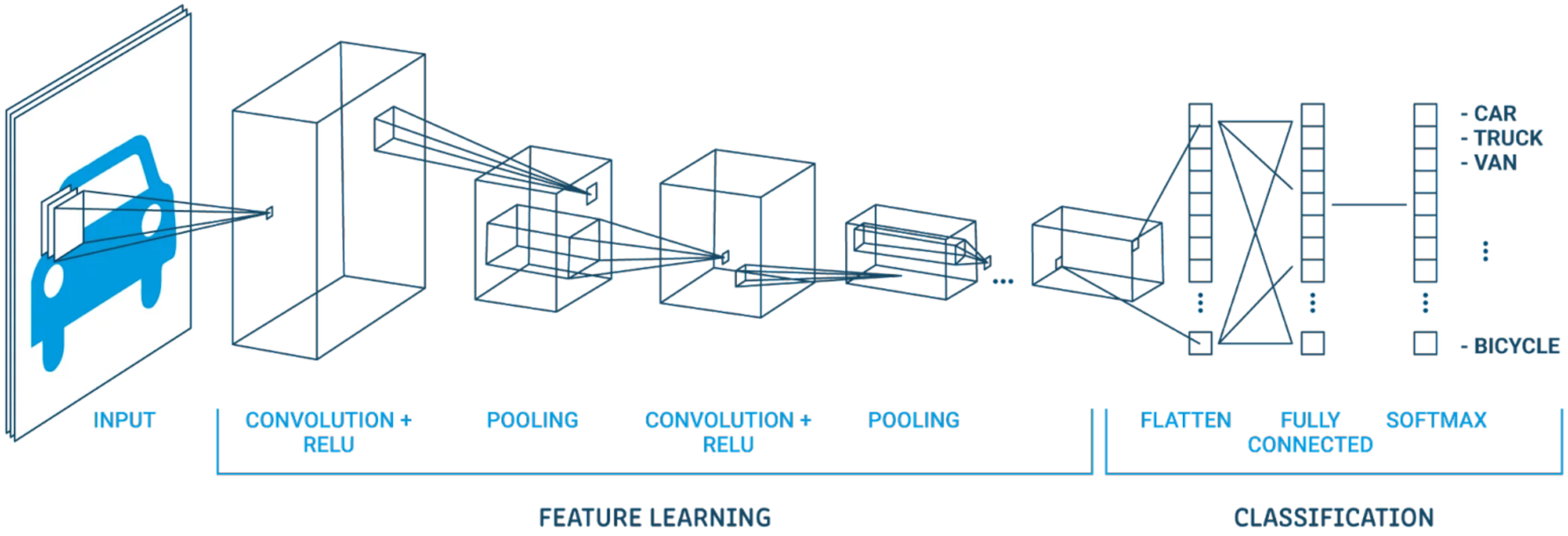
Computer vision using ML

- Images as input
- Preprocess and feature extraction done by developers
- Algorithms train using extracted features
- Prediction



Computer vision using DL (CNN)

- Images as input
- Preprocess and feature extraction done by algorithm
- Algorithms train using extracted features
- Prediction



State of the art

State of the art

In this chapter we did a recap of recent articles and research papers talking about skin cancer detection and classification, in the beginning we talked about traditional methods then we explored new and novel ideas.

Raman spectra

- **dataset** : They brought tissue samples and applied a laser on these tissues to get back Raman scatter radiation. The intensity of this radiation was calculated according to its different frequencies (800- 1800 cm^{-1}) (y=f(x), x: frequency, y: intensity) and that is their data set.

- **Features** : From these obtained spectrums they extracted different statistical measures: arithmetic mean, derivative ... etc

- **Classification** : LightGBM (complex decision tree)

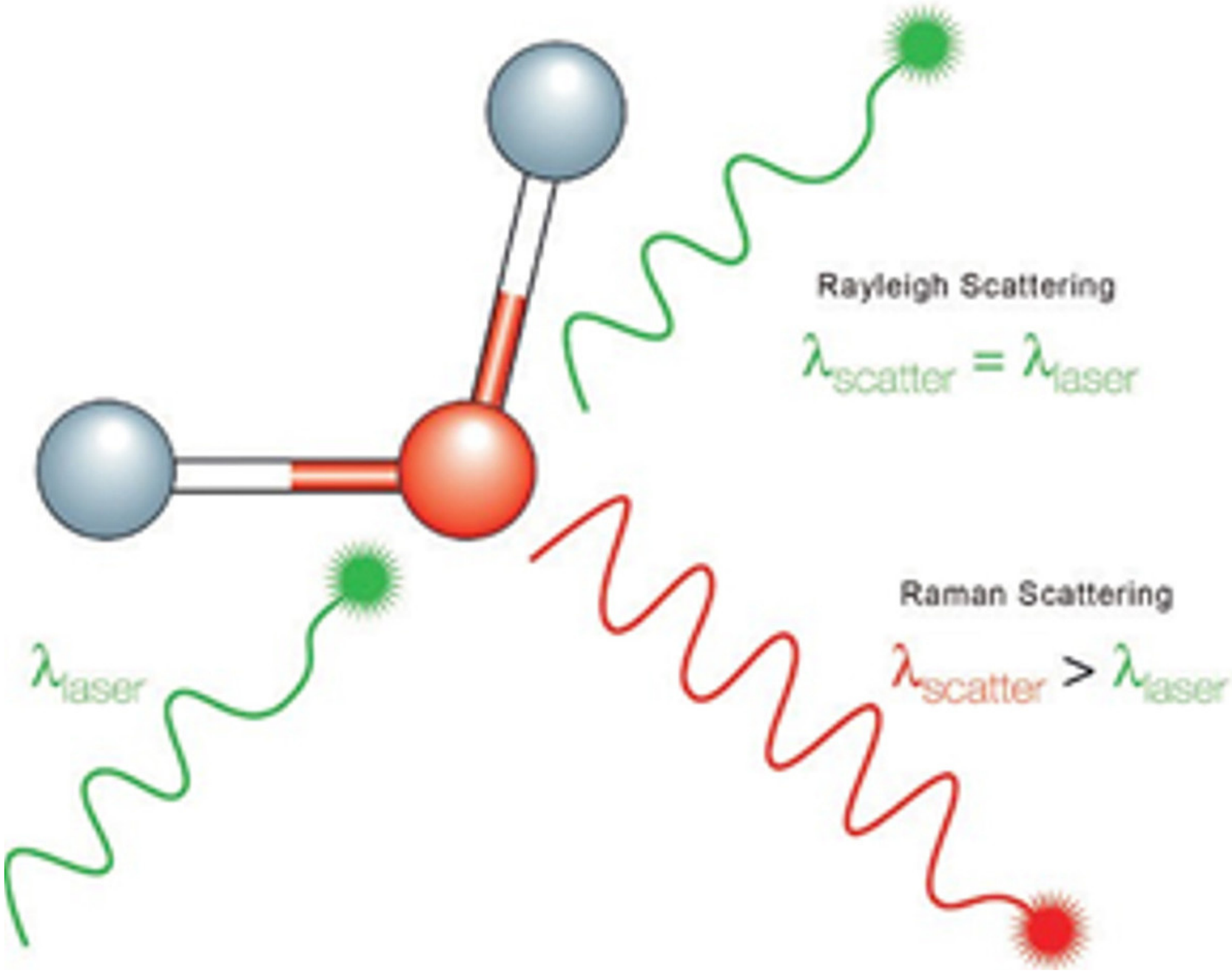
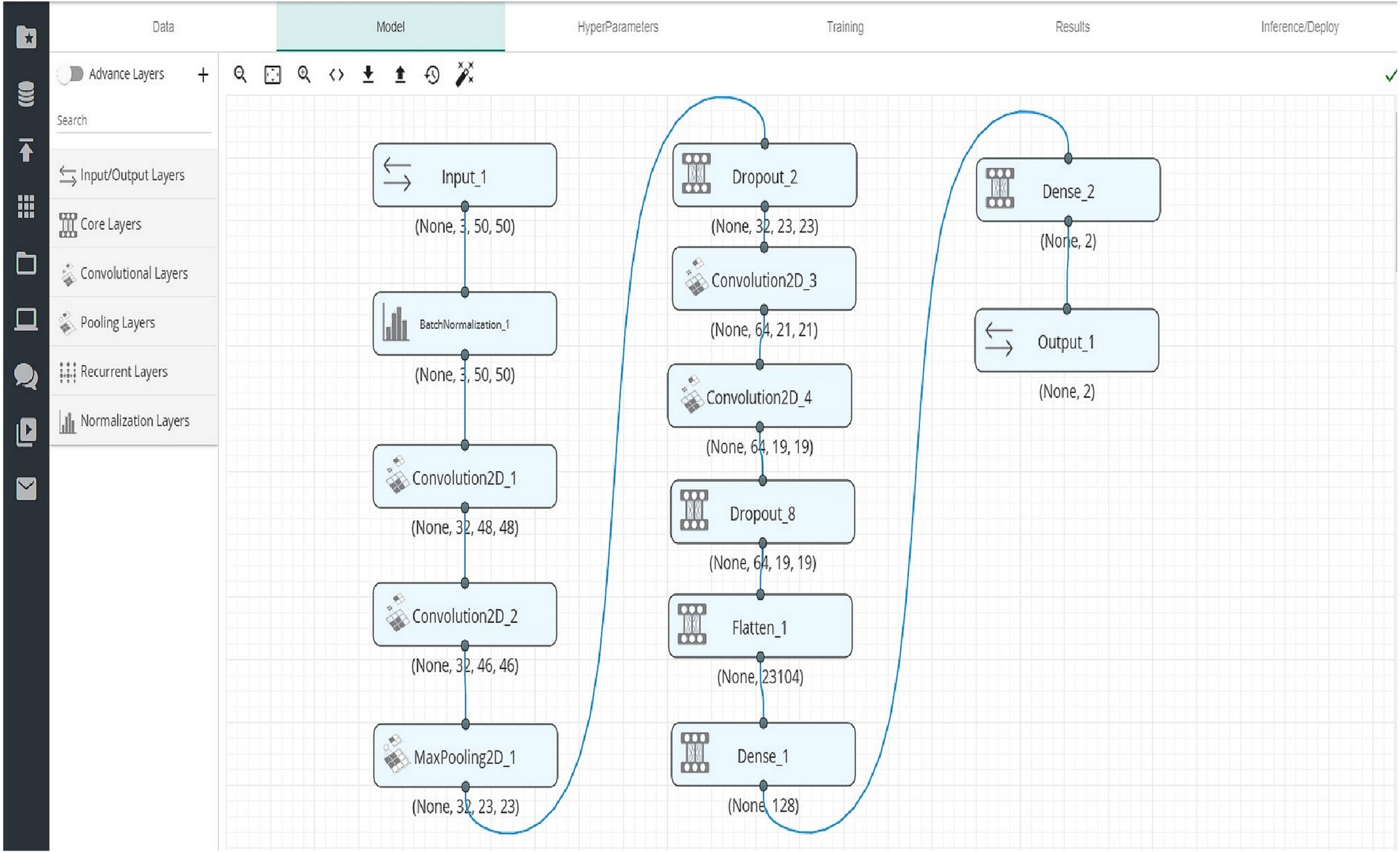

Rayleigh Scattering
$\lambda_{scatter} = \lambda_{laser}$
Raman Scattering
$\lambda_{scatter} > \lambda_{laser}$
λ_{laser}

Image & Clinical data

They wanted to combine image data with clinical information (such as: age, place of lesion, degree of pain...etc.), but before combining they applied a NN feature reducer on the image features to create the balance between image features and clinical features.
The effect of clinical data: almost all metrics, f1-score, accuracy... improved by 7%

DLS (deep learning studio)

Drag and drop interface for building models with desktop/cloud versions + automatic hosting as a Rest API + possibility to download the source code.

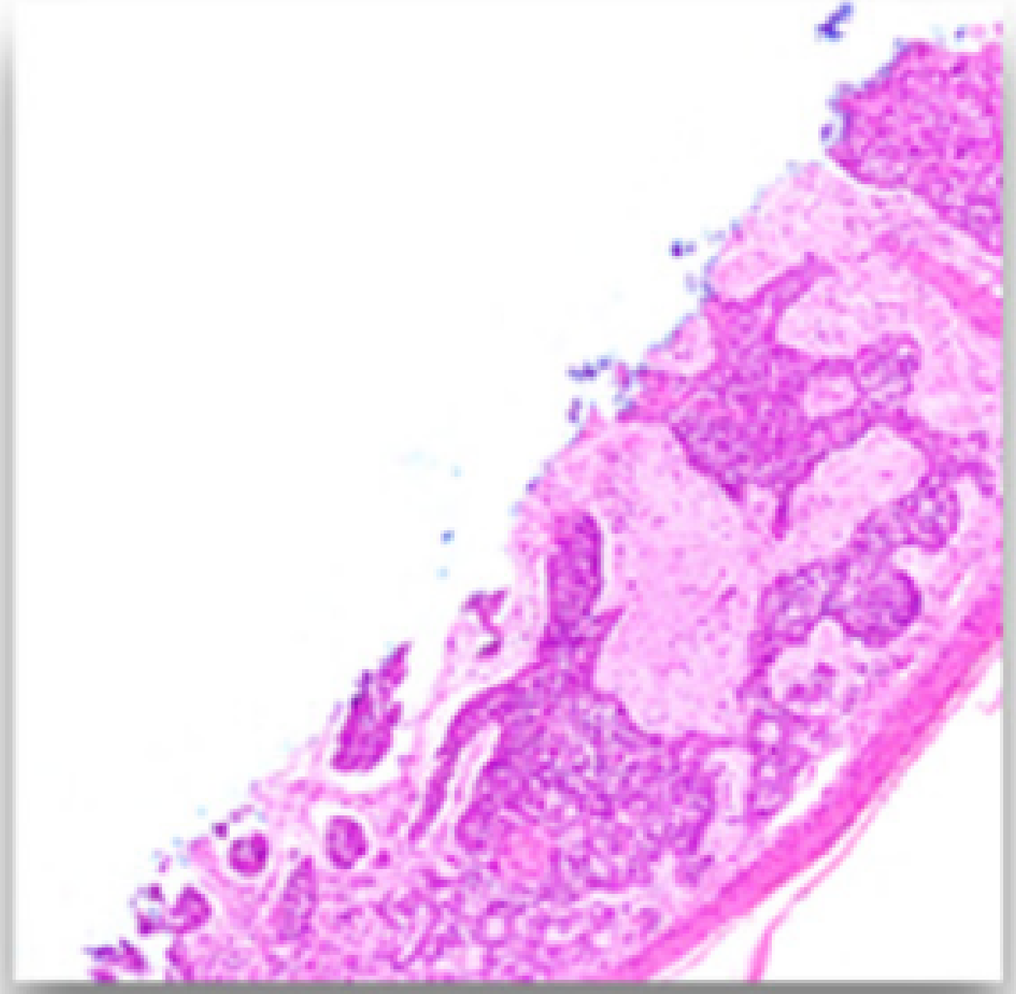

Data
Model
HyperParameters
Training
Results
Inference/Deploy
Advance Layers
Search
Input/Output Layers
Core Layers
Convolutional Layers
Pooling Layers
Recurrent Layers
Normalization Layers
Input_1
(None, 3, 50, 50)
BatchNormalization_1
(None, 3, 50, 50)
Convolution2D_1
(None, 32, 48, 48)
Convolution2D_2
(None, 32, 46, 46)
MaxPooling2D_1
(None, 32, 23, 23)
Dropout_2
(None, 32, 23, 23)
Convolution2D_3
(None, 64, 21, 21)
Convolution2D_4
(None, 64, 19, 19)
Dropout_8
(None, 64, 19, 19)
Flatten_1
(None, 23104)
Dense_1
(None, 128)
Dense_2
(None, 2)
Output_1
(None, 2)

Interpretability

Most ML algorithms are black boxes => can't be trusted in high stake scenarios.

The researchers wanted to implement an understandable CNN model by training it to detect cancer the same way a real doctor would, which is by analyzing the whole tissue and for that they implemented 2 models, one for semantic segmentation and the second for classification so the input image (which is a microscopic image) is going to pass through 2 model, the first will segment it into 12 semantic (dermis, hair, sweat glands...) classes and the second will receive the result of the first model as input and then will output the classification result (healthy, BCC, SCC, IEC)
side effect: because their method takes into consideration the whole tissue, they say that it can be enhanced to detect surgical margins and much more

Input ($h \times w \times 3$)



Output ($h \times w \times 12$)

Recap

All methods follow the same 6 step process.

- Input data
- Preprocessing
- Segmentation
- Feature extraction
- Classification
- Output data

And the differences are in datasets used, preprocessing techniques, chosen features and, the performance achieved by each method.

Datasets

- ISIC 2019
- PH2
- Hallym, problem performance depends on ethnicity
- Dermquest, problem unsatisfactory for non-melanoma
- DermIS, problem unsatisfactory for non-melanoma
- HAM10000 ISIC 2018, problem biased for diverse dataset (but works good for binary classification)
- Dermnet, problem able to identify only few skin lesions
- ISBI
- Atlas Contents

Datasets

Compile your own dataset if you have a new idea (such as: Raman spectra, clinical data + image)
A general problem met with all datasets is the preprocessing and cleaning (such as smoothing and hair removal)

Preprocessing

- Gray scale conversion
- Gaussian and median filter for noise removal and enhancement
- Dull Razor method (hair removal)

And other.

Feature extraction

- GLCM(Grey Level Cooccurrence Matrix)
- ABCD(Asymmetry, Border, Color and Diameter)
- LBP (local binary pattern)
- Autocorrelation
- Correlation

And other statistical features (such has arithmetic mean, standard deviation for Raman spectra).

Feature extraction

All of these features are proven to play an important role in the classification process but depending on your situation and the dataset you use you could find for example a single feature that presents the same performance as the group of features (like in the case of where they used the derivative only) and you can simply check that by training the model with different combination of features until you find an important difference or by using statistical methods that tells you the degree of discrimination of each feature.

Classification

- SVM
- Hybrid adaboost SVM (Adaboost: adaptive boosting, it can be used with various algorithms to increase the performance)
- MSVM (multi-class)
- LightGBM (complex decision tree, open source software)
- ANN (trained with pre extracted features)
- DNN (deep neural networks) with Dragonfly optimization algorithm
- CNN and its famous architectures and other statistical features (such has arithmetic mean, standard deviation for Raman spectra).
  - LeNet-5
  - AlexNet
  - VGG-16
  - Inception-v3
  - ResNet-50
- DLS (deep learning studio)

Evaluation

Both deep learning and machine learning is the same, with deep learning being slightly better. The best of both sides are: (MSVM reaching an accuracy of 96.25% and LightGBM achieving an AUC >97%) and (Inceptionv3 CNN architecture reaching and AUC of 99%).

Line of reasoning

Dataset

- consider the different limitations
- consider using a hybrid dataset
- Use a novel dataset (as we have seen with: Raman spectra, microsopic images)

Extracted Features

A lot of predefined features, use one of them or a combination.

Patient Experience and Target Users

Invasive or non-invasive?
Is it affordable or not?

Methods

In ML :

- Needs less data
- MSVM
- LightGBM

In DL :

- Requires a huge amount of data
- +optimization algorithms like: Dragonfly, Whale
- ANN trained with extracted features
- CNN
  - Inceptionv3
  - ResNet

For multiclassification use transfer learning.
For binary classification you can train from scratch.

No programming experience

Some tools help in that such as DLS(deep learning studio)

Interpretability

If your model is intended for real world application, you really need to consider the interpretability aspect.

Master conclusion

We conclude that there are a lot of things to take into considerations when building a machine learning model, there is no right or wrong way. but in the hopes of drawing a guidance map for future contributors we did a recap of most famous methods, algorithms and datasets used.

PFE

Goal

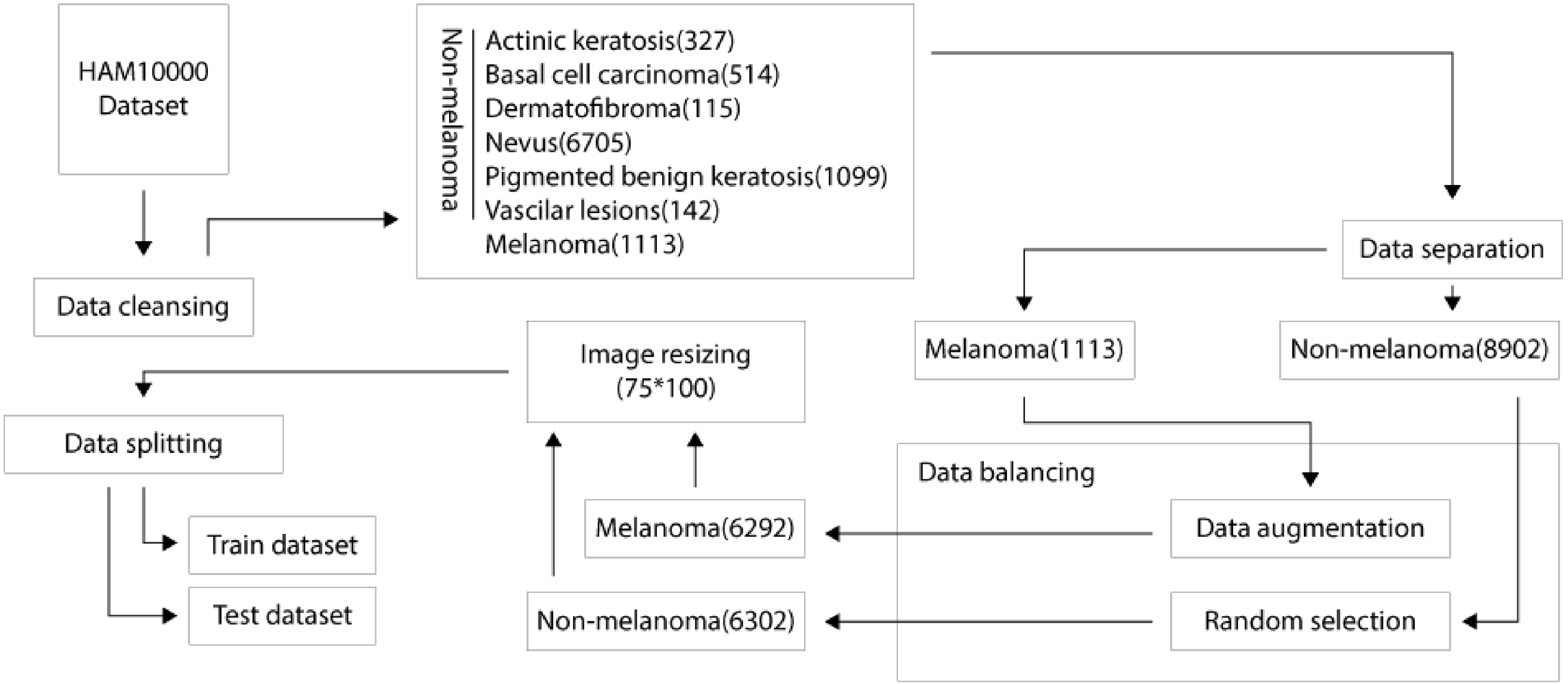
Our system is going to be a web-based platform, easy to use and, accessible to everyone through the Internet.

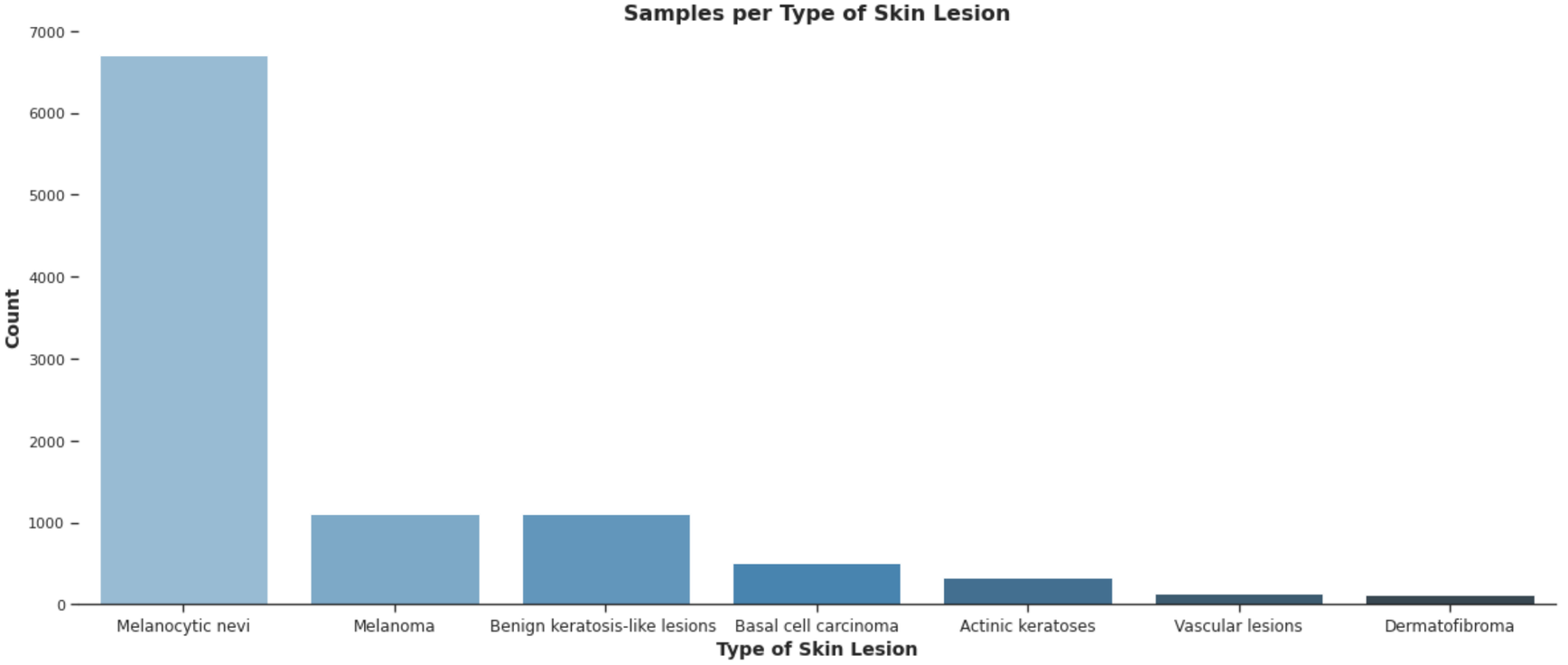
The Model

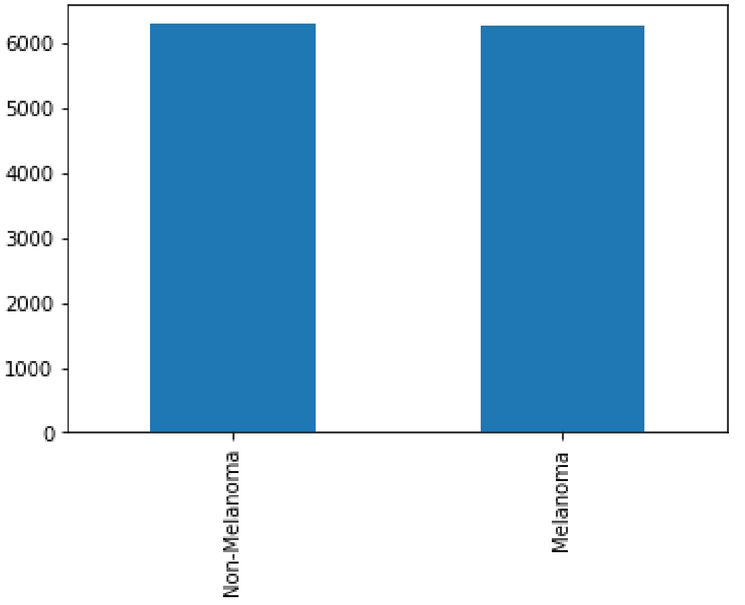
Dataset

MNIST- HAM10000 : is composed of 10.015 labeled images for 7 different types of skin lesions. The data was collected from Australian and Austrian patients.

Pre-processing




Samples per Type of Skin Lesion
Count
7000
6000
5000
4000
3000
2000
1000
0
Melanocytic nevi
Melanoma
Benign keratosis-like lesions
Basal cell carcinoma
Actinic keratoses
Vascular lesions
Dermatofibroma
Type of Skin Lesion


6000
5000
4000
3000
2000
1000
0
Non-Melanoma
Melanoma

Architecture

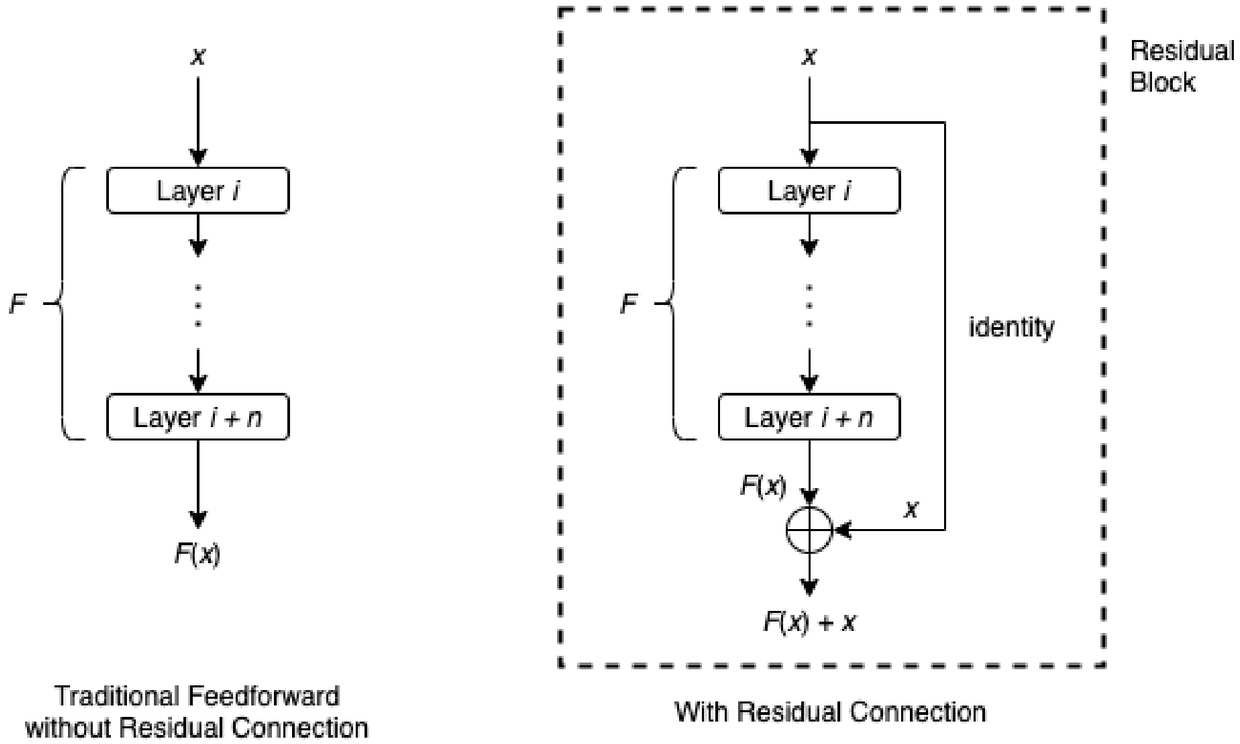
The Problem

In our case we have a limited dataset, so we thought about building a deep neural network to get acceptable accuracy and performance. The intuition behind adding more layers is that these layers progressively learn more complex features.

However, when we add more layers to the neural network, it gets more difficult to train them, and their accuracy begins to saturate and ultimately decline caused by exploding and vanishing gradients.

Residual Network

ResNet is a powerful CNN that can overcome the problem exploding and vanishing gradients.
The main idea of ResNet is to create an “identity shortcut link” that skips one or more layers.



ResNet50

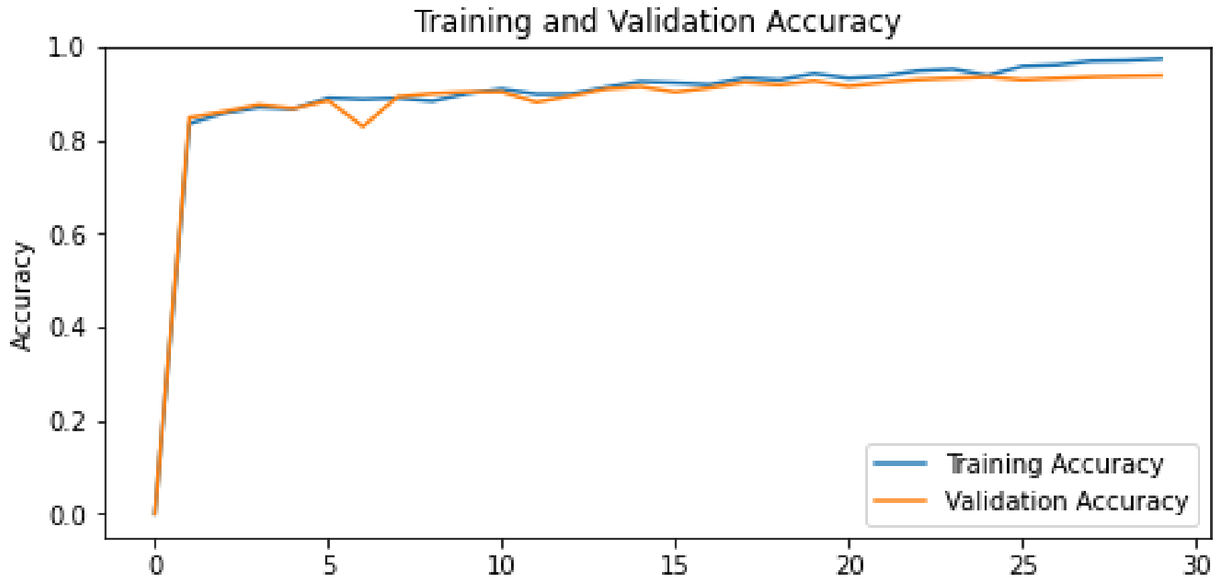
Residual Network has many versions according to the number of layers used such as ResNet50, ResNet152 and many others. we chose ResNet50 according to our available computational power used for training and its high achieved accuracy when applied to various datasets.

Model name	Number of parameters	ImageNet Accuracy
ResNet50	26 M	77.15%
ResNet152	60 M	78.57%

Environment

- **Train and test environment** : Dell XPS with GTX1050 and 4 GB of VRAM
- **Prediction environment** : Amazon EC2 c5a.xlarge instance with 4vCPU and 8 GB of RAM

Evaluation

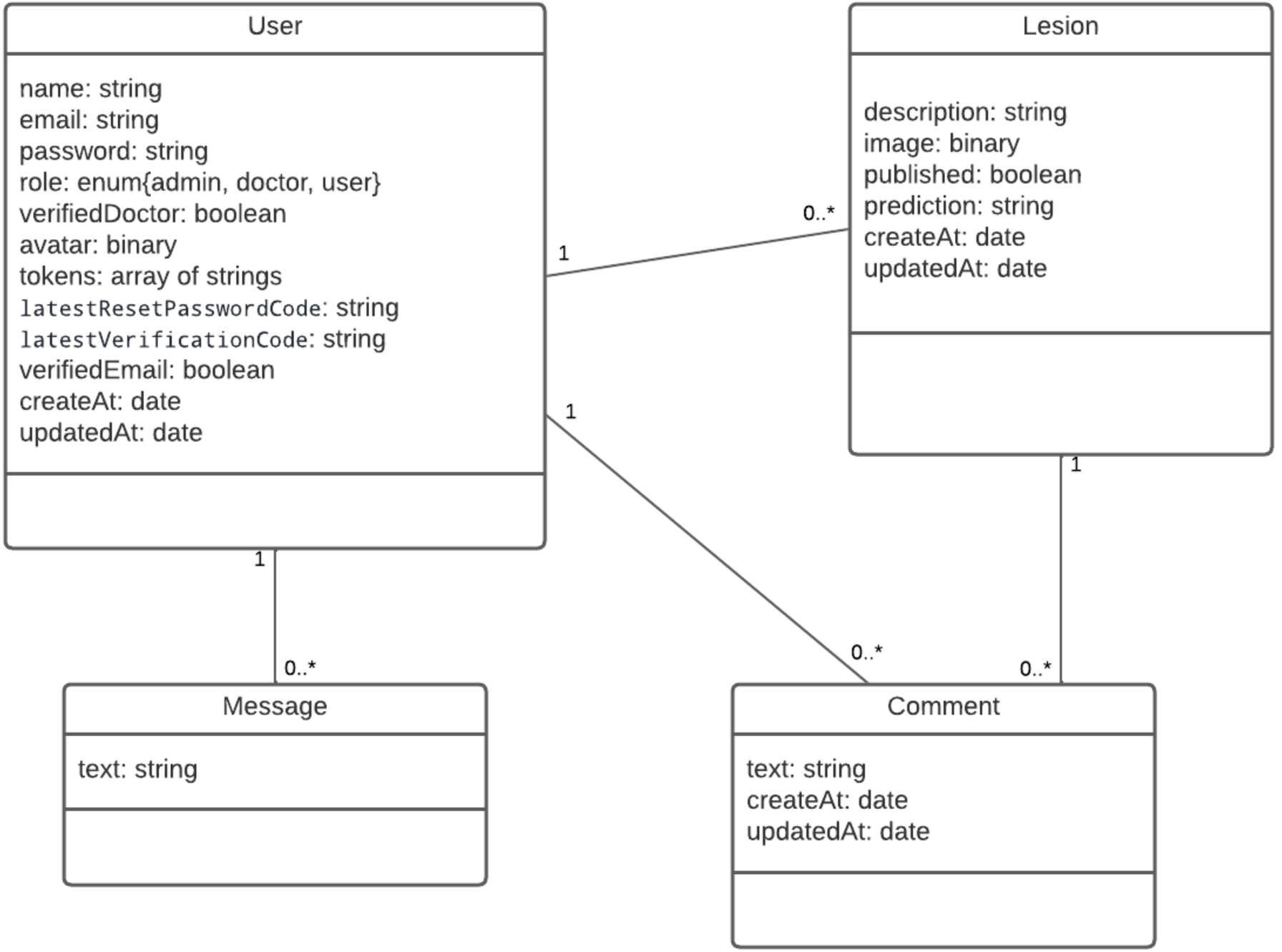

Training and Validation Accuracy
Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
Training Accuracy
Validation Accuracy

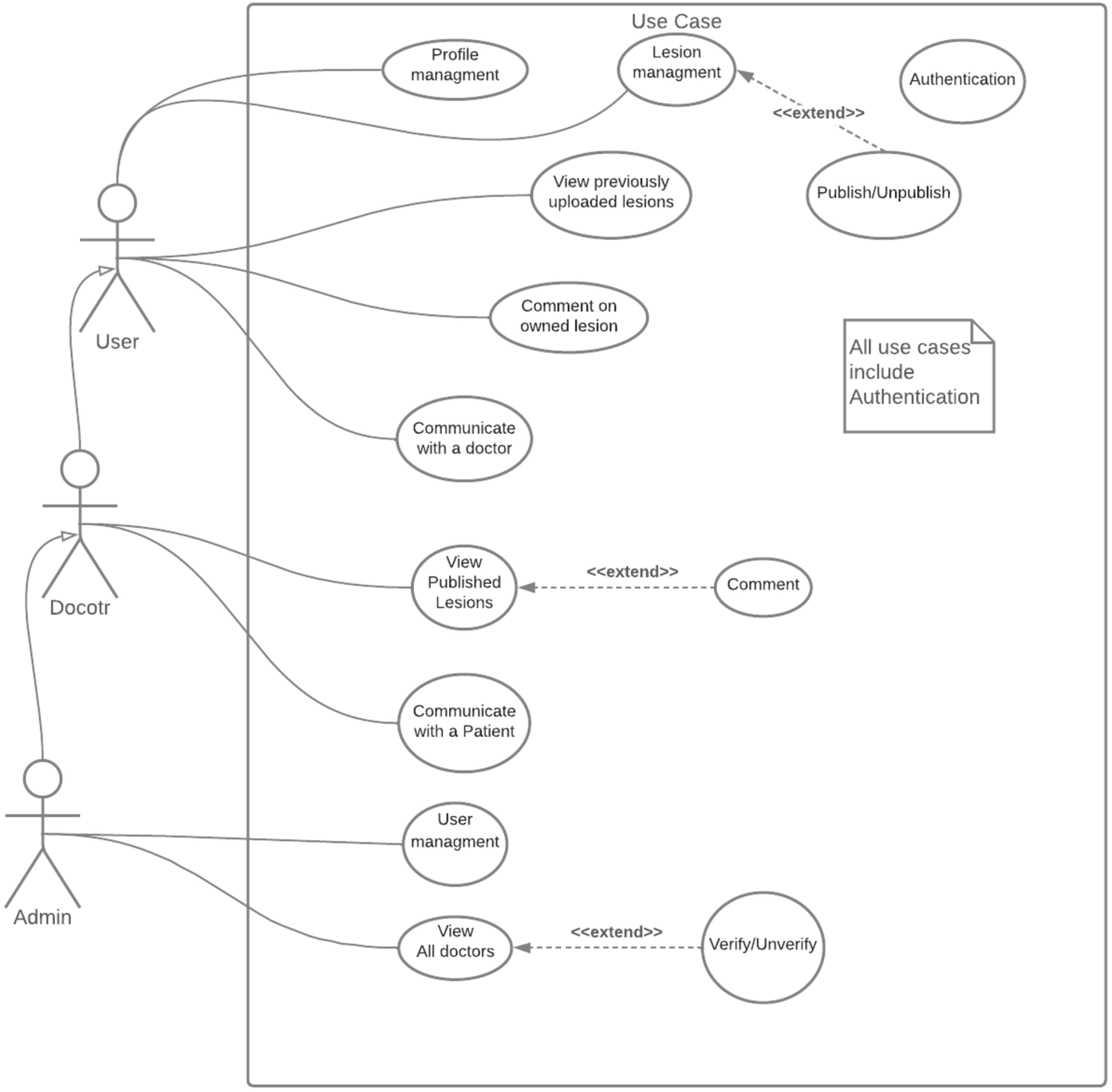

Training and Validation Loss
Cross Entropy
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
epoch
Training Loss
Validation Loss

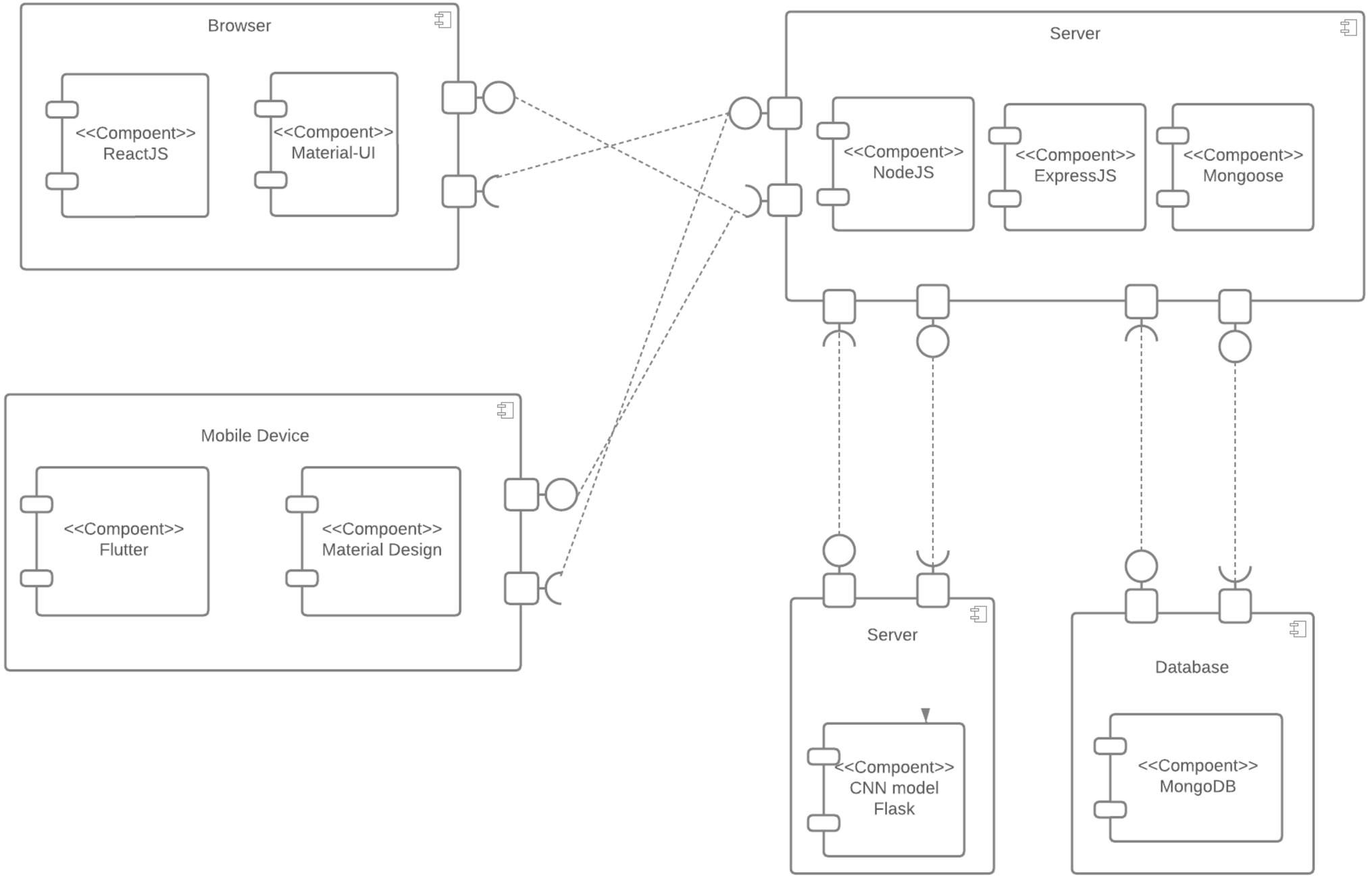
The web based platform

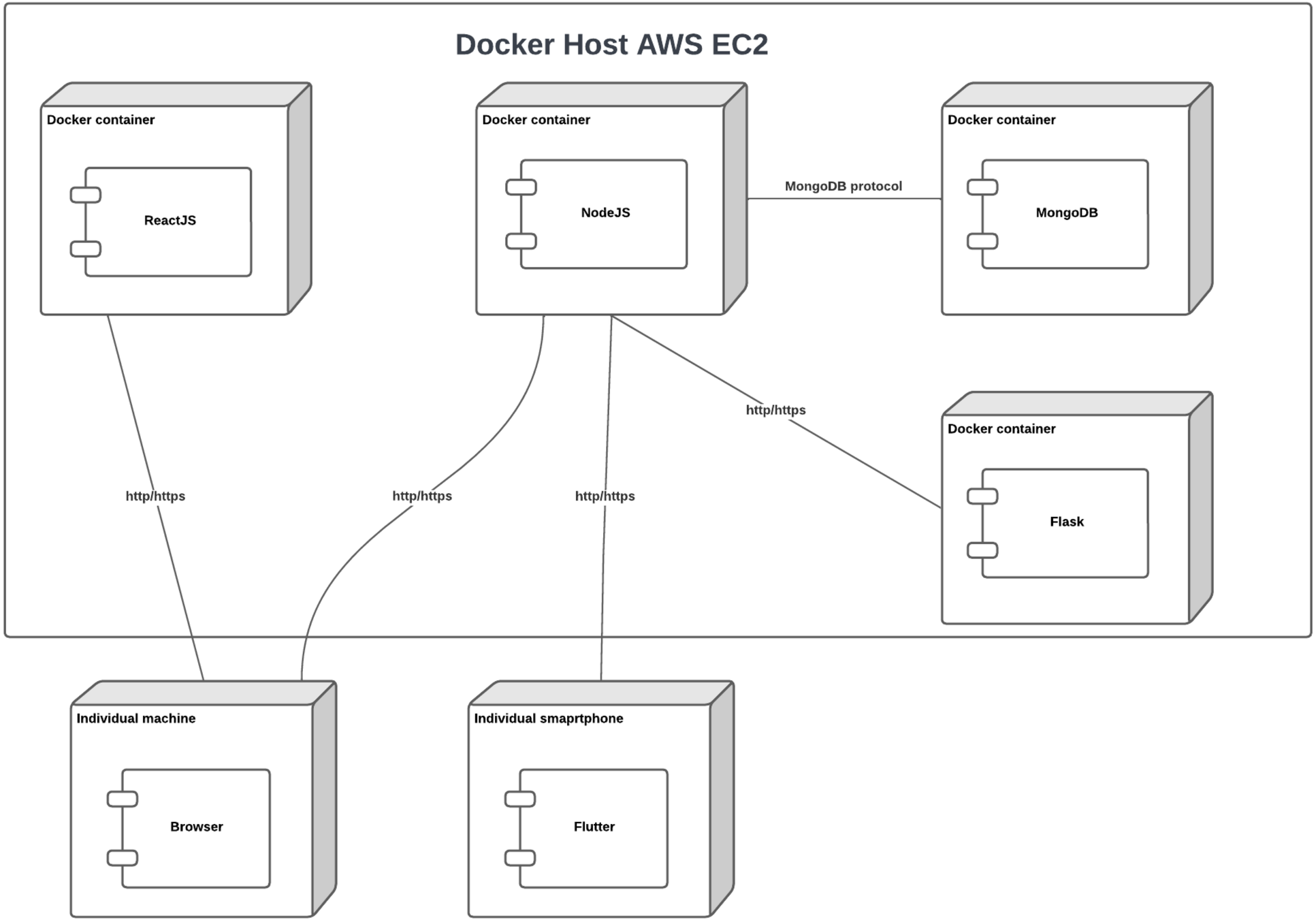
Conception

- **Main goal :** facilitate the diagnostic process
- **Patients :**
  - confirmation, second opinion, verification
  - decide whether or not to go to a dermatologist for further testing
- **Doctors :**
  - help in the initial testing phase
  - naked eye analysis replacement or confirmation
  - decide whether or not to do further Lab testing (biopsies)


User
name: string
email: string
password: string
role: enum{admin, doctor, user}
verifiedDoctor: boolean
avatar: binary
tokens: array of strings
latestResetPasswordCode: string
latestVerificationCode: string
verifiedEmail: boolean
createAt: date
updatedAt: date
Lesion
description: string
image: binary
published: boolean
prediction: string
createAt: date
updatedAt: date
Message
text: string
Comment
text: string
createAt: date
updatedAt: date
1
0..*
1
0..*
1
0..*
1
0..*


Use Case
Profile managment
Lesion managment
Authentication
<<extend>>
Publish/Unpublish
View previously uploaded lesions
Comment on owned lesion
All use cases include Authentication
Communicate with a doctor
User
Docotr
View Published Lesions
<<extend>>
Comment
Communicate with a Patient
User managment
Admin
View All doctors
<<extend>>
Verify/Unverify


Browser
<<Compoent>>
ReactJS
<<Compoent>>
Material-UI
Server
<<Compoent>>
NodeJS
<<Compoent>>
ExpressJS
<<Compoent>>
Mongoose
Mobile Device
<<Compoent>>
Flutter
<<Compoent>>
Material Design
Server
<<Compoent>>
CNN model
Flask
Database
<<Compoent>>
MongoDB


Docker Host AWS EC2
Docker container
ReactJS
Docker container
NodeJS
MongoDB protocol
Docker container
MongoDB
http/https
Docker container
Flask
http/https
http/https
http/https
Individual machine
Browser
Individual smaprtphone
Flutter

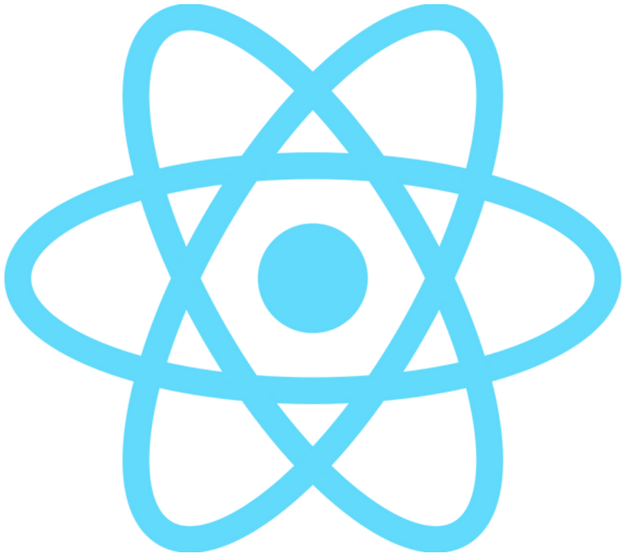
Mobile application

Features

- Account management
  - Create an account
  - Login
  - Reset password
  - Email verification
  - Logout
  - Upload profile image
- Lesion management
  - Upload lesion
  - Publish & Unpublish lesion
  - Delete lesion
  - Read comments

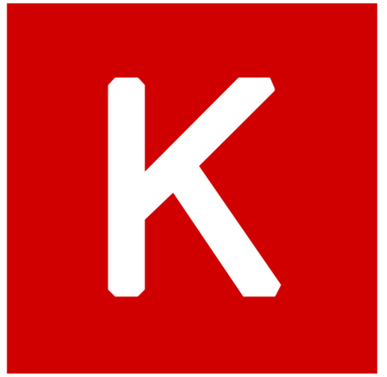
Software

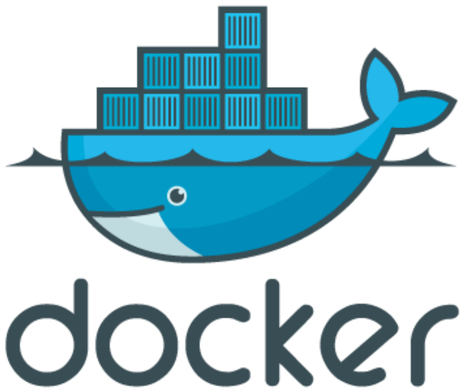
Frontend

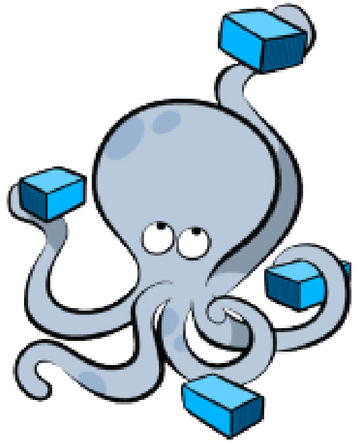


Backend

Flask







express

Future Work

To extend the usability of our project

- Detect and classify more lesions and not just melanoma.
- Take model interpretability into account, so it will be more acceptable in real world scenarios.
- Present more in-app features to facilitate and enhance user experience.

Conclusion

we have built a diagnosis system that could facilitate the process of detection and early discovery of melanoma skin cancer, it is web-based, easy to use and accessible to everyone, all you need is internet connection, a laptop or a smartphone, and you are set to go. It can either be used by normal users who think that they might have melanoma, they can check that using the app before actually going to a doctor and do invasive tests (such as biopsy), or it can be used by doctors to facilitate the process of diagnosis and improve patients experience.

General conclusion

General conclusion

The work of an engineer is to find solutions, either by using existent tools or inventing new ones, and our work on our master's thesis and final year project (PFE) has helped us a lot to grow as engineers. Our theoretical work helped us to develop as beginner researchers, and our practical work gave us the feel and experience of real world work environment.

So we want to thank our school and all of its staff for providing us with the opportunity to acquire both degrees.

Scenario of execution

Mobile :

- Create user account
- User uploads lesion image
- Publish his lesion image

Web site :

- Create a doctor account
- Verify doctor account by admin
- View published lesion images
- Comment on a lesion image
- Chat between doctor and user
- View new dataset by admin