MDBOT

Graduation Project: MDBOT

Supervised By

Prof. Dr. Sami Abdel Moneim
T.A Eng. Walaa Zaghlol

Topics

- Introduction
- Project SDLC
- System’s models
- References

INTRODUCTION

Data-driven, predictive and patient-centric solutions that adhere to today’s regulatory are the values that empower interactive healthcare nowadays, and those values worked together as a starting point for our graduation project.

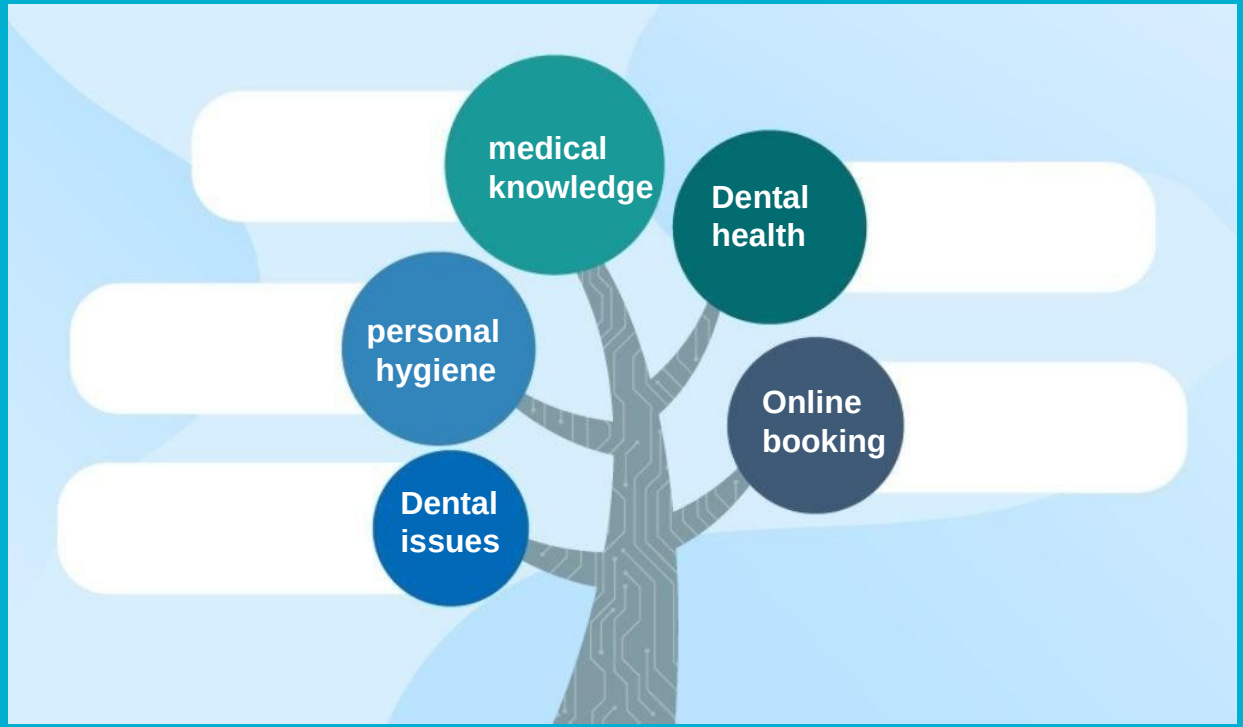
INTRODUCTION

1. **Purpose**

   To offer a medical solution that make use of the technological solutions and balance it with the actual problems faced by users, and keeping in mind the best practises and user’s most engaging communication styles.

INTRODUCTION

2. Problems



INTRODUCTION

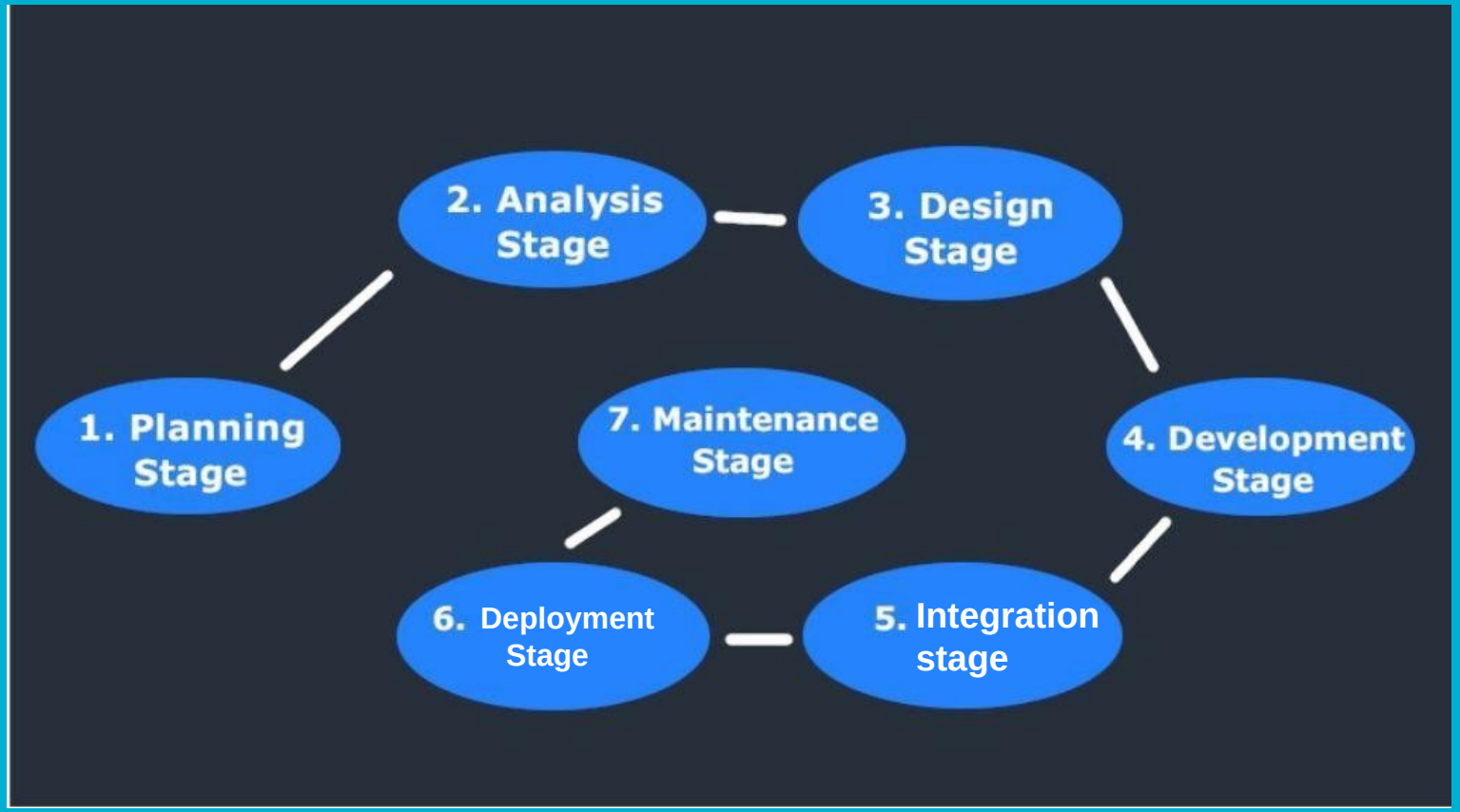
3. Proposed Strategy

- Easy-Access Website with optional booking appointments section
- Virtual Medical assistance
- dental hygiene self check for a better dental health management service
- chatbot system to assist users
- comprehensive diagnosis of diseases

Scope:

- Deployment channel
- Website medical fields
- Flow of the website
- The system effect on the actual Medical-related jobs
- User Characteristics

Project SDLC



Project SDLC

1. **Planning Stage**

   - Objectives
   - Timeline and the project schedule

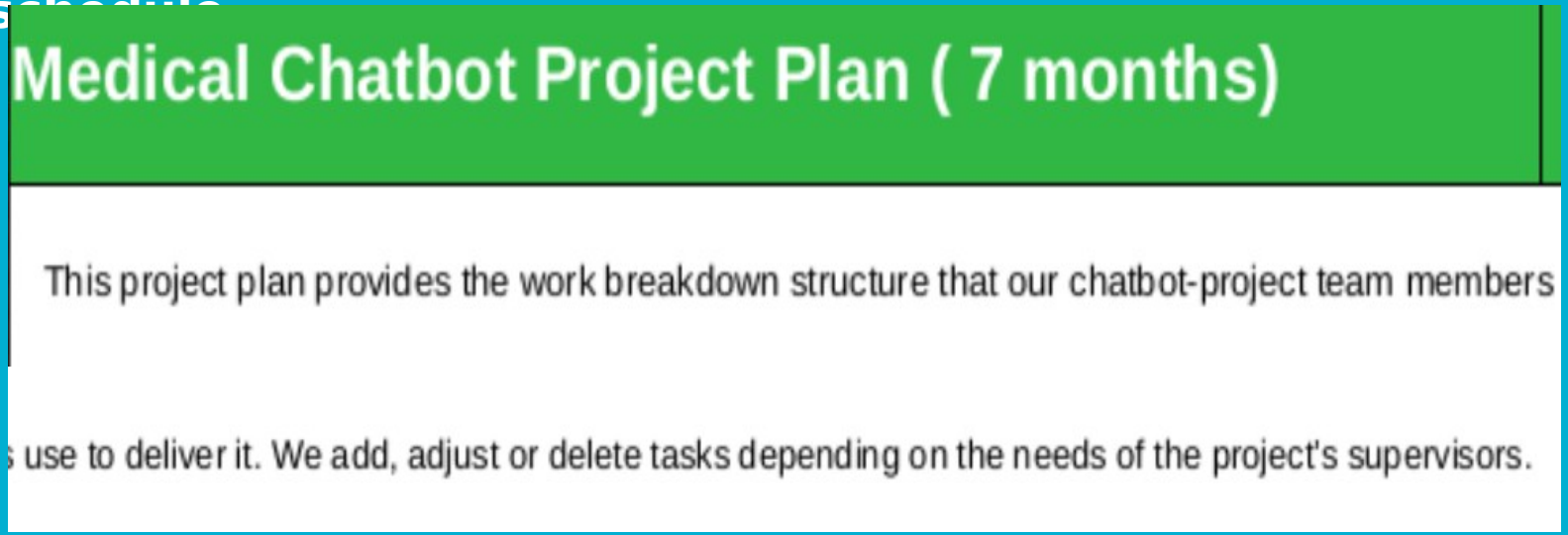
Project SDLC

Main Objective:

To have a Doctor in your pocket

Project SDLC

Timeline and the project schedule

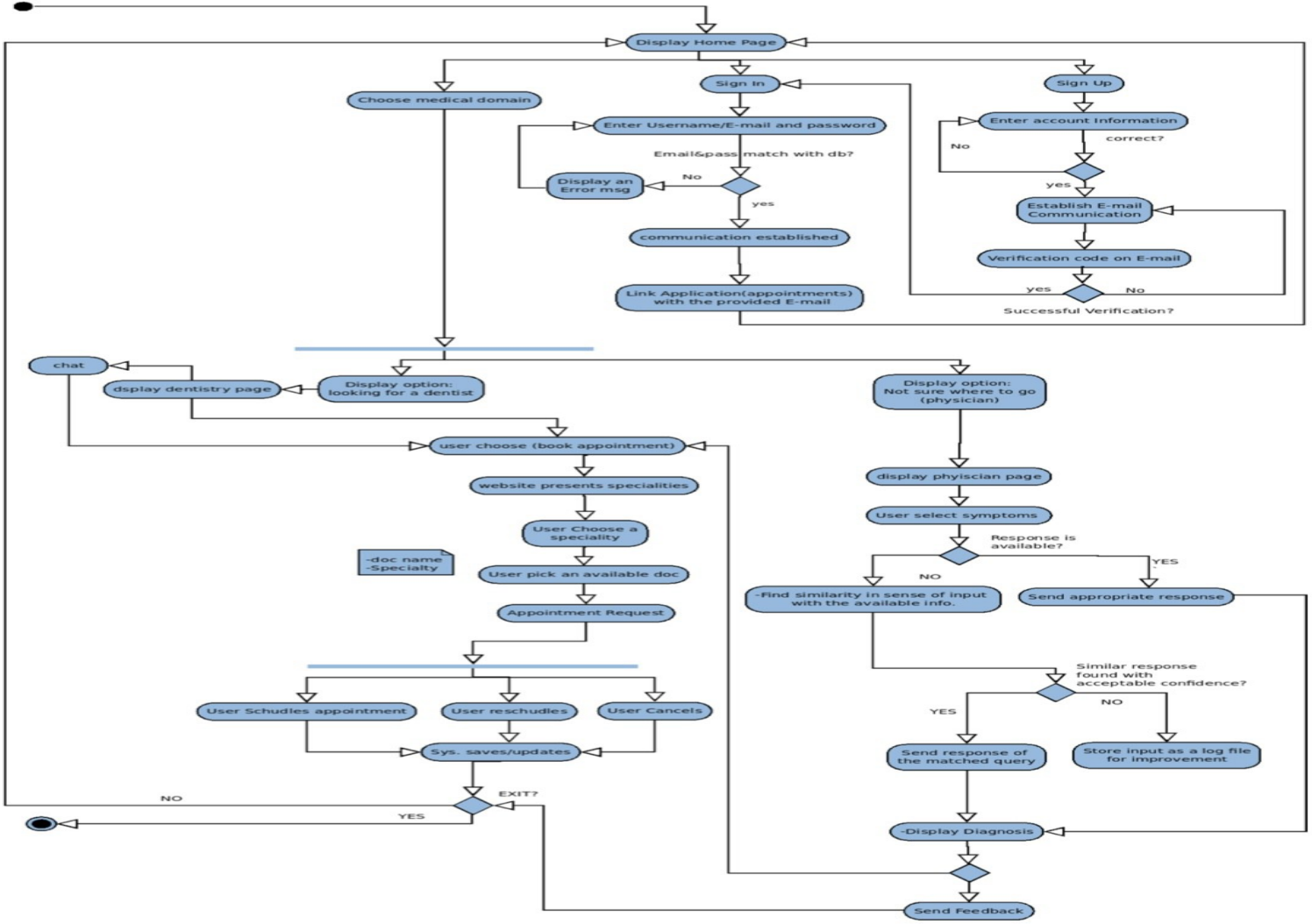


Project SDLC

1. **Analysis Stage**

   identify the requirements of a system and the information used

3.2.3 Activity Diagram



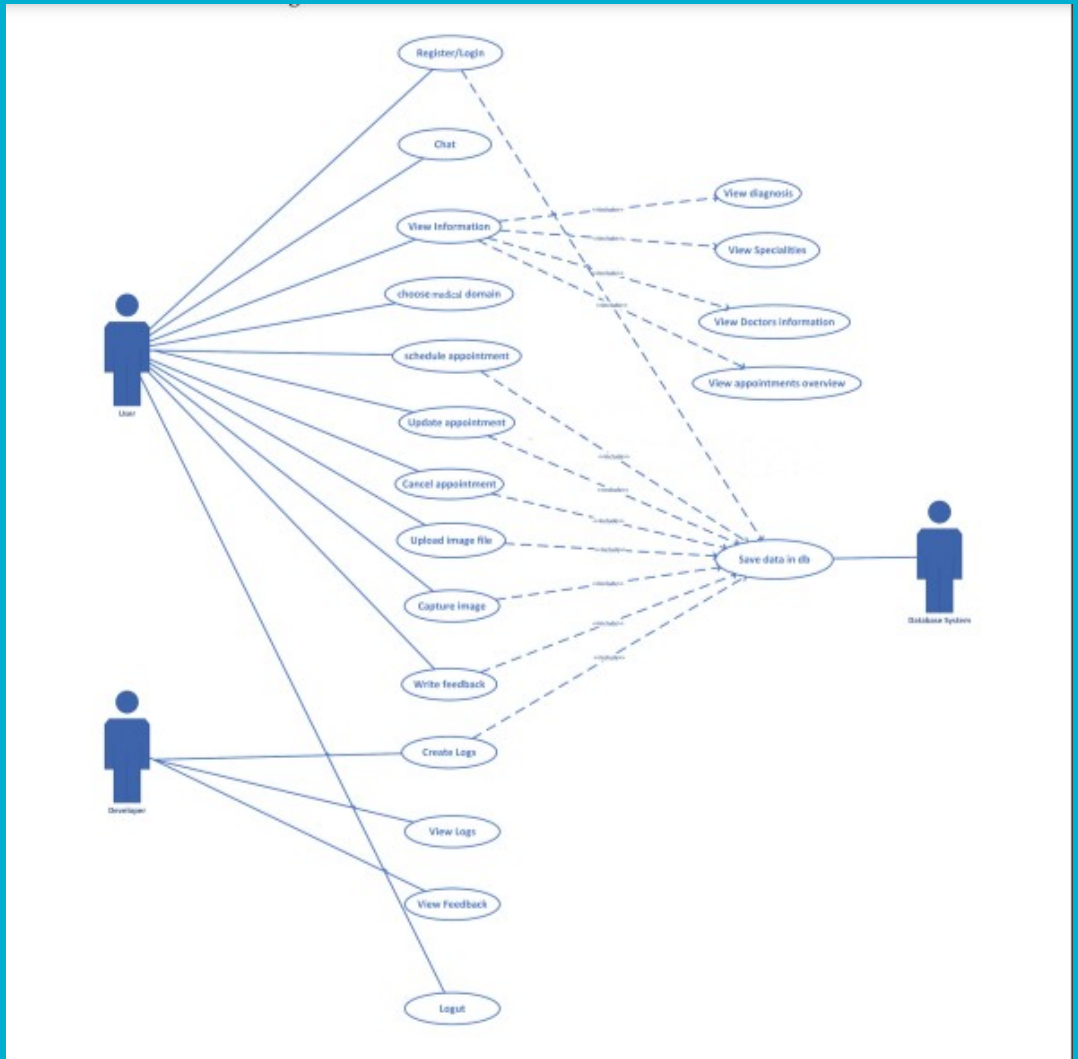
Project SDLC

Development Stage

developers writing code and building the application according to
the earlier design documents and outlined specifications.

- Python developers: Building the physician and Dentistry sections
- Front-end developers: Building the graphical user interface of a website
- Database Administrators: Storing and organizing data, and designing the db

Use Case diagram



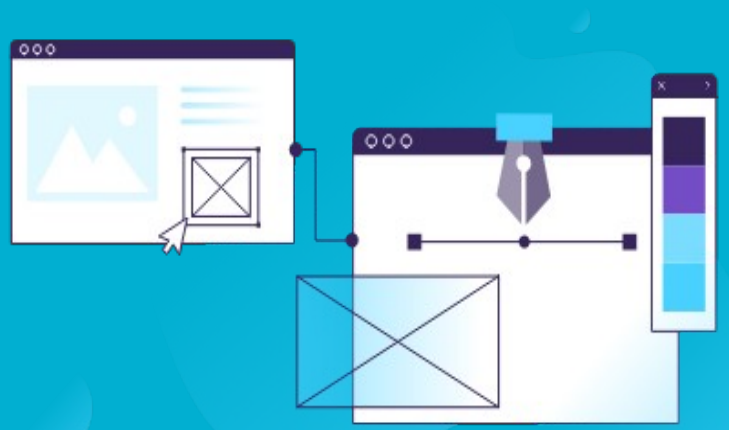
Use case	**Register/Login**
Description	The user Login to his account or creates a new account
Actors	user
Included use cases	Save data in database
Preconditions	home page loaded with two options (Register or Login). user have a valid account.
Postconditions	app displays the Home page
Main flow	1. user selects Register or Login 2. app prompts user to enter Information 3. user enter his Information 4. app verifies data and Login to the user's account
Alternative flows	• User enter invalid Register/Login information ◦ app displays an Error message ◦ the user reenters the information ◦ app verifies the new information

Project SDLC

Integration Stage

combining all the Development parts

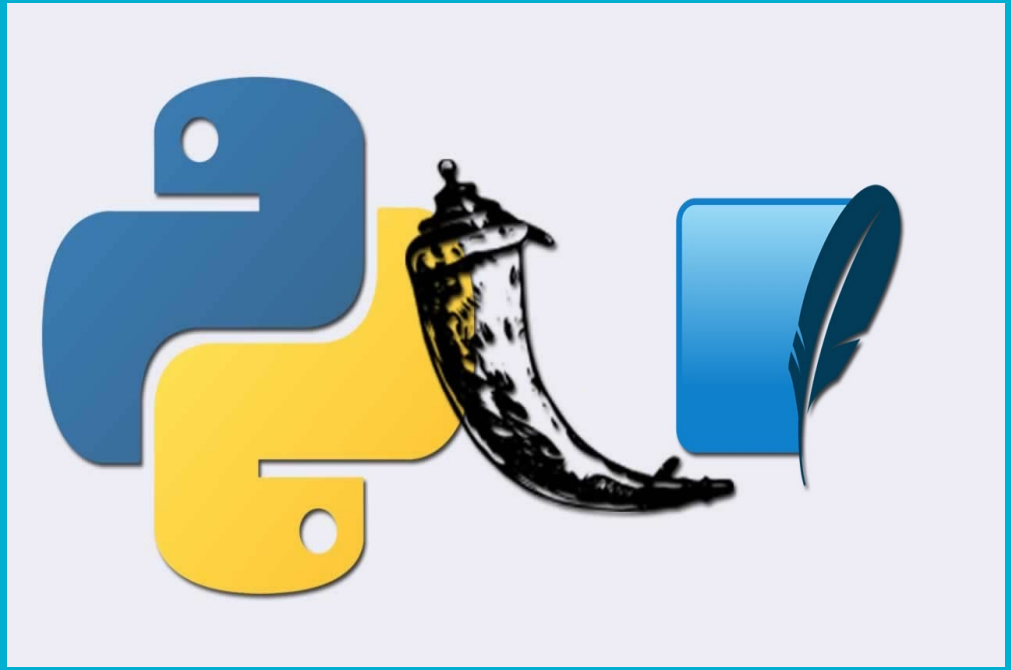
together into one system.



Project SDLC

5. Deployment Stage

development server deployment



Project SDLC

5. Maintenance Stage

- repair and fix the website when breaking.
- save logs and review them for error checks

Models

Target

- Reliable disease prediction based on symptoms.
- Based on the symptoms, the system will suggest doctors in the matched
  specialization.

How to reach?

AI algorithms that utilizes a set of data that enables it to recognize certain patterns (Symptoms). This allows it to reach a conclusion or make a prediction (Diseases) when provided with sufficient information.

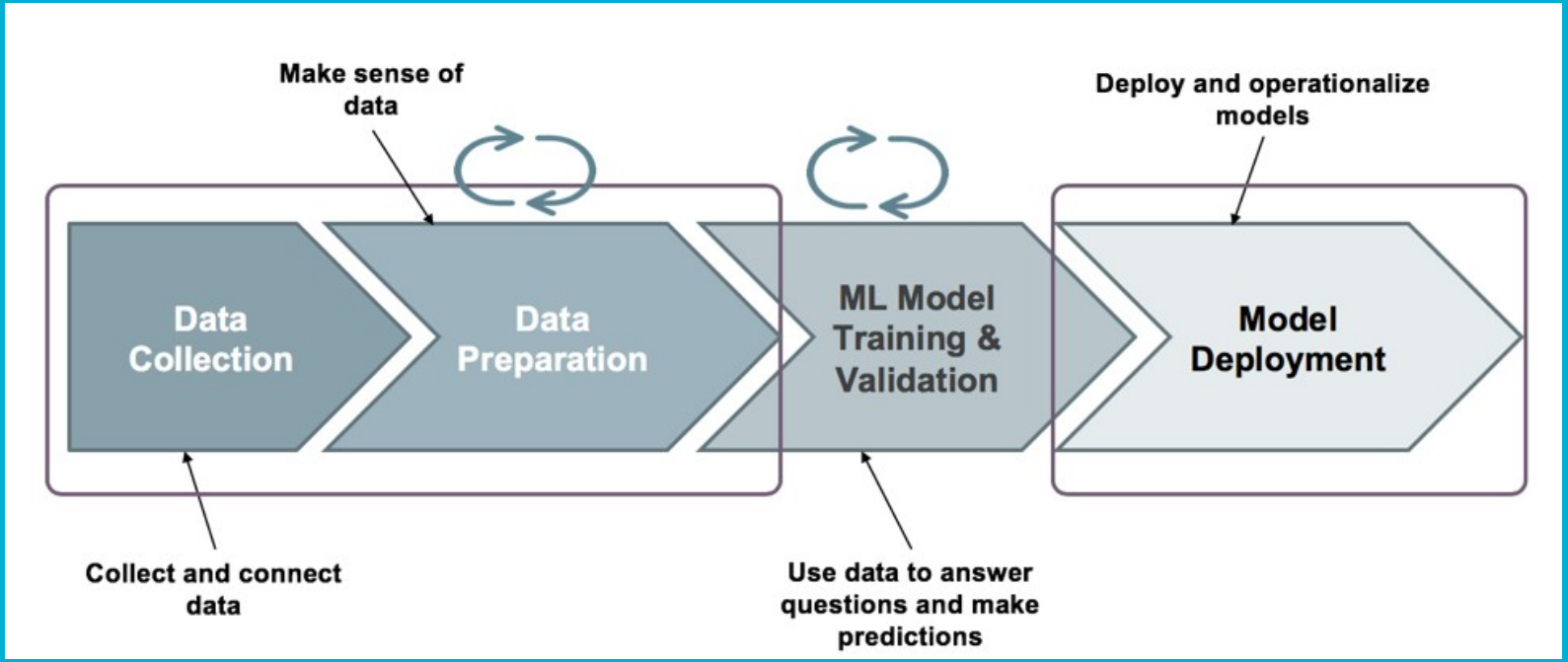
Models

1 .	2 .	3 .
chatbot	naive bayes	Decision Tree

Models

- Chatbot
  - Bag of words
  - Normalizer
  - Intents
  - Train

Models


Make sense of data
Deploy and operationalize models
Data Collection
Data Preparation
ML Model Training & Validation
Model Deployment
Collect and connect data
Use data to answer questions and make predictions

Models

Data Collection

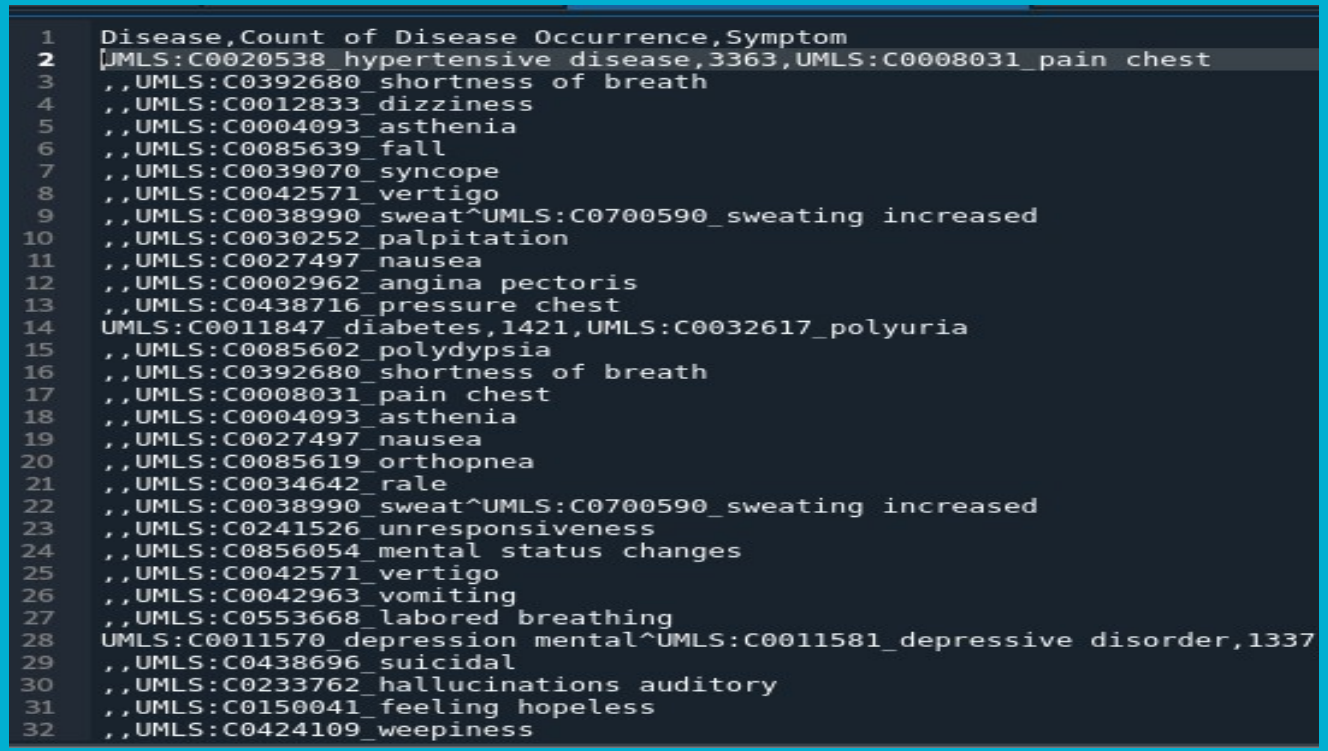
Columbia University dataset

Disease-Symptom Knowledge Database

Disease	Count of Disease Occurrence	Symptom
UMLS:C0020538_hypertensive disease	3363	UMLS:C0008031_pain chest
		UMLS:C0392680_shortness of breath
		UMLS:C0012833_dizziness
		UMLS:C0004093_asthenia
		UMLS:C0085639_fall
		UMLS:C0039070_syncope
		UMLS:C0042571_vertigo
		UMLS:C0038990_sweat^UMLS:C0700590_sweating increased
		UMLS:C0030252_palpitation
		UMLS:C0027497_nausea
		UMLS:C0002962_angina pectoris
		UMLS:C0438716_pressure chest
UMLS:C0011847_diabetes	1421	UMLS:C0032617_polyuria
		UMLS:C0085602_polydypsia
		UMLS:C0392680_shortness of breath
		UMLS:C0008031_pain chest
		UMLS:C0004093_asthenia
		UMLS:C0027497_nausea
		UMLS:C0085619_orthopnea
		UMLS:C0034642_rale
		UMLS:C0038990_sweat^UMLS:C0700590_sweating increased
		UMLS:C0241526_unresponsiveness
		UMLS:C0856054_mental status changes

Models

Data Collection (after web scraping)



```
Disease,Count of Disease Occurrence,Symptom
UMLS:C0020538_hypertensive disease,3363,UMLS:C0008031_pain chest
,,UMLS:C0392680_shortness of breath
,,UMLS:C0012833_dizziness
,,UMLS:C0004093_asthenia
,,UMLS:C0085639_fall
,,UMLS:C0039070_syncope
,,UMLS:C0042571_vertigo
,,UMLS:C0038990_sweat^UMLS:C0700590_sweating increased
,,UMLS:C0030252_palpitation
,,UMLS:C0027497_nausea
,,UMLS:C0002962_angina pectoris
,,UMLS:C0438716_pressure chest
UMLS:C0011847_diabetes,1421,UMLS:C0032617_polyuria
,,UMLS:C0085602_polydypsia
,,UMLS:C0392680_shortness of breath
,,UMLS:C0008031_pain chest
,,UMLS:C0004093_asthenia
,,UMLS:C0027497_nausea
,,UMLS:C0085619_orthopnea
,,UMLS:C0034642_rale
,,UMLS:C0038990_sweat^UMLS:C0700590_sweating increased
,,UMLS:C0241526_unresponsiveness
,,UMLS:C0856054_mental status changes
,,UMLS:C0042571_vertigo
,,UMLS:C0042963_vomiting
,,UMLS:C0553668_labored breathing
UMLS:C0011570_depression mental^UMLS:C0011581_depressive disorder,1337
,,UMLS:C0438696_suicidal
,,UMLS:C0233762_hallucinations auditory
,,UMLS:C0150041_feeling hopeless
,,UMLS:C0424109_weepiness
```

Models

Cleaned Data

Source	Target	Weight
hypertensive disease	pain chest	3363
hypertensive disease	shortness of breath	3363
hypertensive disease	dizziness	3363
hypertensive disease	asthenia	3363
hypertensive disease	fall	3363
hypertensive disease	syncope	3363
hypertensive disease	vertigo	3363
hypertensive disease	sweat	3363
hypertensive disease	sweating increased	3363
hypertensive disease	palpitation	3363
hypertensive disease	nausea	3363
hypertensive disease	angina pectoris	3363
hypertensive disease	pressure chest	3363
diabetes	polyuria	1421
diabetes	polydypsia	1421
diabetes	shortness of breath	1421
diabetes	pain chest	1421
diabetes	asthenia	1421
diabetes	nausea	1421
diabetes	orthopnea	1421
diabetes	rale	1421
diabetes	sweat	1421
diabetes	sweating increased	1421
diabetes	unresponsiveness	1421
diabetes	mental status changes	1421
diabetes	vertigo	1421
diabetes	vomiting	1421
diabetes	labored breathing	1421
depression mental	feeling suicidal	1337
depression mental	suicidal	1337
depression mental	hallucinations auditory	1337
depression mental	feeling hopeless	1337
depression mental	weepiness	1337
depression mental	sleeplessness	1337
depression mental	motor retardation	1337

Models

Data Preparation

Source	Heberden's node	Joint pain, hip	Knee pain	Knee pain while moving	Knee swelling	Murphy's sign	Numbness of both lower limbs	Numbnes
Alzheimer's disease	0	0	0	0	0	0	0	
HIV	0	0	0	0	0	0	0	
Knee osteoarthritis	0	0	1	1	1	0	0	
Pneumocystiscariniipneumonia	0	0	0	0	0	0	0	
Sciatica	0	1	0	0	0	0	1	
accidentcerebrovascular	0	0	0	0	0	0	0	
acquiredimmuno-deficiency syndrom	0	0	0	0	0	0	0	
adenocarcinoma	0	0	0	0	0	0	0	
adhesion	0	0	0	0	0	0	0	
affect labile	0	0	0	0	0	0	0	
anemia	0	0	0	0	0	0	0	
anxiety state	0	0	0	0	0	0	0	
aphasia	0	0	0	0	0	0	0	
arthritis	0	0	0	0	0	0	0	
asthma	0	0	0	0	0	0	0	
bacteremia	0	0	0	0	0	0	0	
benign prostatic hypertrophy	0	0	0	0	0	0	0	
biliary calculus	0	0	0	0	0	0	0	
bipolar disorder	0	0	0	0	0	0	0	
bronchitis	0	0	0	0	0	0	0	
candidiasis	0	0	0	0	0	0	0	
carcinoma	0	0	0	0	0	0	0	
carcinoma breast	0	0	0	0	0	0	0	
carcinoma colon	0	0	0	0	0	0	0	
carcinoma of lung	0	0	0	0	0	0	0	
carcinoma prostate	0	0	0	0	0	0	0	
cardiomyopathy	0	0	0	0	0	0	0	
cellulitis	0	0	0	0	0	0	0	
cholecystitis	0	0	0	0	0	1	0	

Models

Training and Validation

After training the multinomial Naive Bayes classifier which is suitable for classification with discrete input features (decoded Symptoms), and then validating the model. The model prediction accuracy reached

Models

Decision Tree

root	attribute which has most effect on the Root
nodes	consists of attribute which has most effect on the outcome
leaf	tests for value of certain attribute

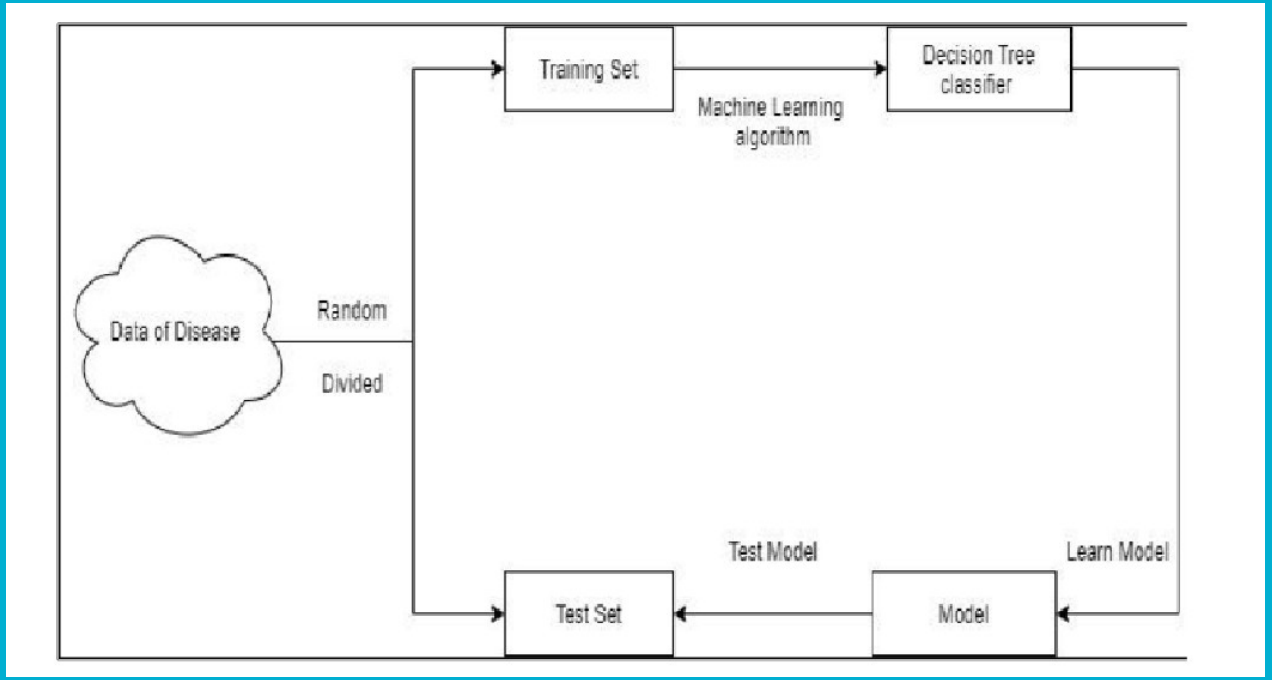
Models

Data: 148 diseases with their corresponding symptoms

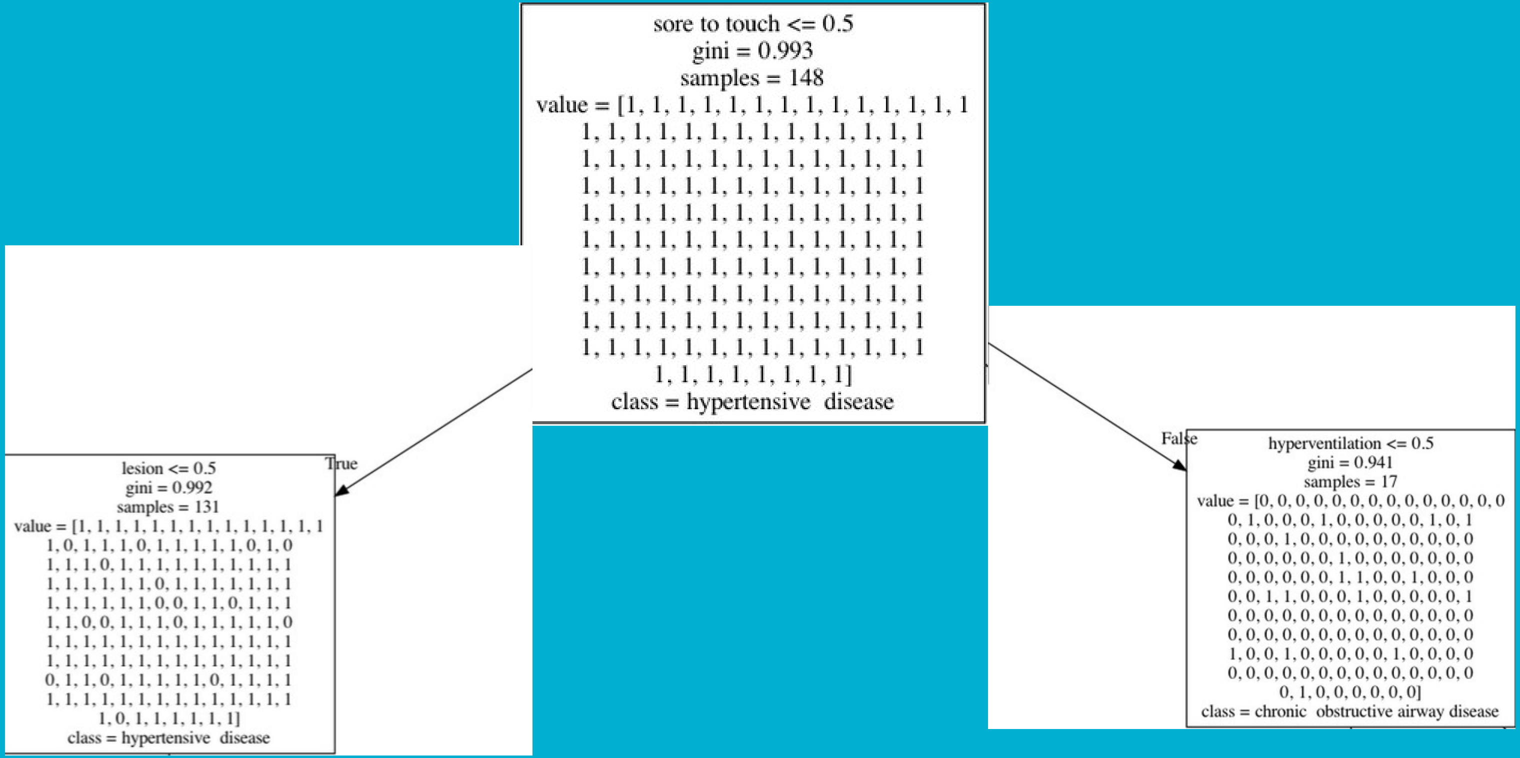
Disease	Count of Disease Occurrence	Symptom
UMLS:C0020538_hypertensive disease	3363	UMLS:C0008031_pain chest
		UMLS:C0012833_dizziness
		UMLS:C0004093_asthenia
		UMLS:C0085639_fall

Models

SYSTEM ARCHITECTURE



Models



sore to touch <= 0.5
gini = 0.993
samples = 148
value = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
1, 1, 1, 1, 1, 1, 1, 1]
class = hypertensive disease
True
lesion <= 0.5
gini = 0.992
samples = 131
value = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0
1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1
1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1
1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
0, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
1, 0, 1, 1, 1, 1, 1, 1]
class = hypertensive disease
False
hyperventilation <= 0.5
gini = 0.941
samples = 17
value = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1
0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0
0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0
0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
0, 1, 0, 0, 0, 0, 0, 0]
class = chronic obstructive airway disease

Deep learning models

diagnosing some of the most common dental diseases such as

- Cavities
- Decay
- Crooked teeth

Deep learning models

Data Collection:

collecting data for Crooked teeth model,
it was important to collect a large number of balanced data

For tooth decay model.
collecting a large number of pictures of the affected teeth

Deep learning models

Data Preprocessing:

create a formatted data set that could be fed to the model for training.

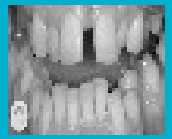
for Crooked teeth model.

Standardize the image size
convert it to grayscale
onvert to numpy array
label the data to build a binary classifier
Shuffle the data

example of image before and after preparing

image before preparing

image after preparing



X and y after converting to numpy array

```
X[1]
```

```
array([[[202],
        [199],
        [197],
        ...,
        [202],
        [199],
        [197]],

       [[199],
        [195],
        [193],
        ...,
        [195],
        [196],
        [195]],
```

```
[8]: random.shuffle(traning_data)
```

```
[9]: for sample in traning_data[:10]:
         print(sample[1])
```

```
1
1
1
0
1
0
1
1
1
1
```

Deep learning models

Data Preprocessing:

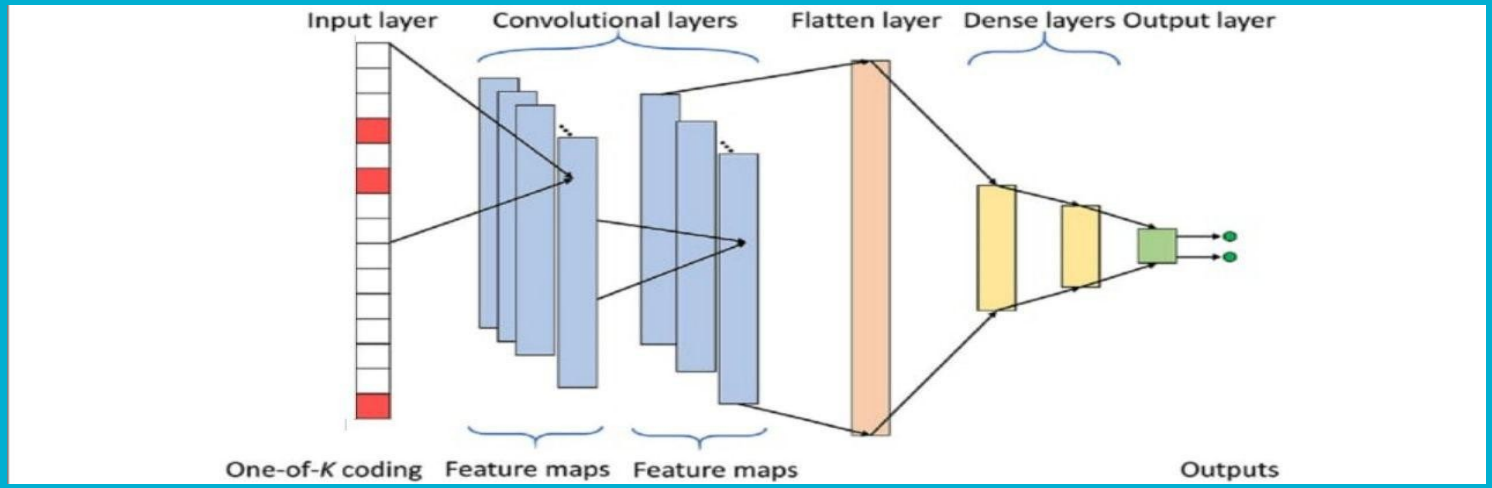
for tooth decay model

colored images
Labels the data

Deep learning models

Implementation:

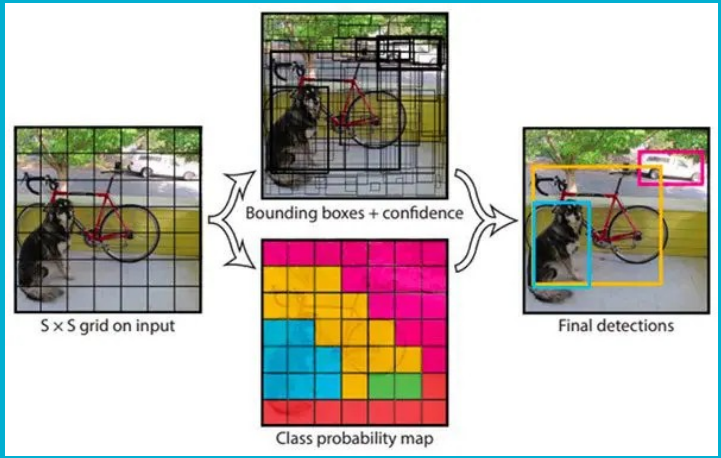
The model is a binary classifier based on convolutional neural networks



Deep learning models

Implementation:

for the second model.
The model is based on YOLO algorithm.
is popular because of its speed and accuracy.



Reference

Disease-Symptom Knowledge Database

https://people.dbmi.columbia.edu/~friedma/Projects/DiseaseSymptomKB/index.html