



Internship report

Presented by :

Ayda Zamanian

Supervised by:

Dr.Zohreh Azimifar

Presented to :

Dr.Farshad Khunjush

Summer 1402

Abstract:

This internship outlines my experience in Dr.Zohre Azimifar's company and highlights the knowledge of data science , software engineering , illustrating technical information and presenting them in the simplest way to non technical people.

The report consists all of my activities such as working and annotating ECG data and therefore reading many papers , attending workshops such as biobank workshop presented by Dr.Fatehi form Tehran university , segmenting in excite about 2'500 bioimage using SAM model And finally scraping the web inorder to answer the following question for Dr.Shakibafar : "How to edit and modify a template text using voice only to eliminate typers for doctors".

Table of contents

Abstract.....	2
Chapter One :Introduction and overview	4
Chapter Two : Activities and reports	5
Conclusion	18
References	19

Introduction and overview

In order to complete the tasks that I was given I needed to learn how to use python signal processing libraries such as Neurokit and Wfdb to illustrate ECG waves.

Neurokit2 :

NeuroKit2 is a user-friendly package providing easy access to advanced biosignal processing routines. Researchers and clinicians without extensive knowledge of programming or biomedical signal processing can analyse physiological data with only two lines of code. I needed to use its features to predict peaks and extract information to illustrate in future.

Wfdb :

Waveform Database (WFDB) is a set of file standards designed for reading and storing physiologic signal data, and associated annotations. There are various types of ECG recorded files such as “.atr” that I needed to use.

Pandas:

Pandas was needed to make graphs and the reason It was chosen was the powerfulness and easy to work with.

Activities and reports

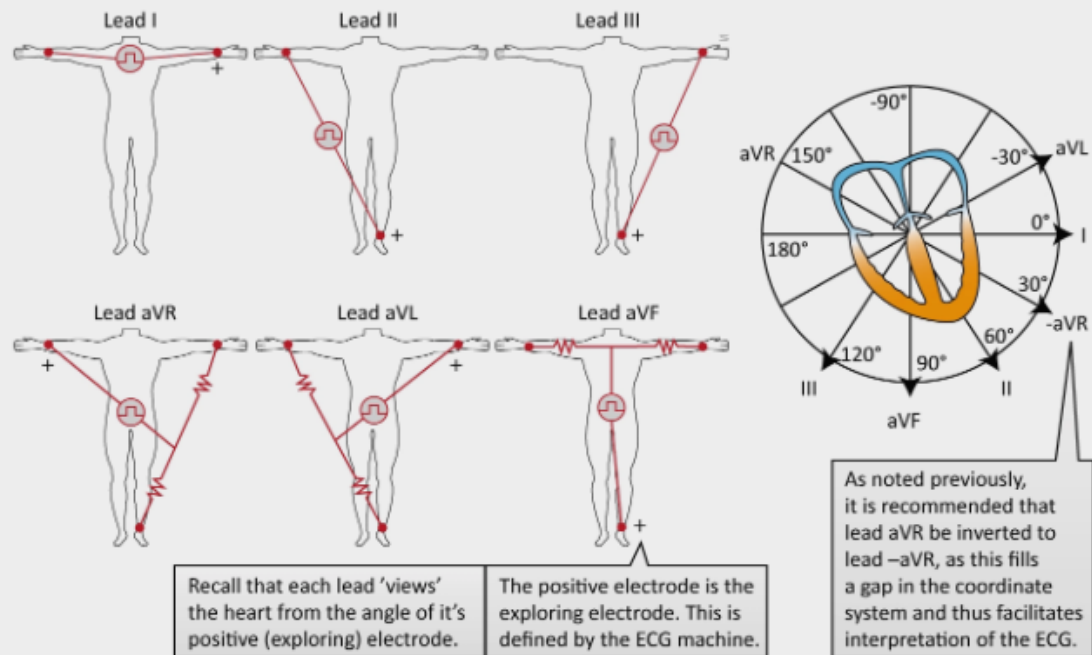
- Firstly I was assigned to a ECG project which it's first phase was visualising ECG data in a way that is understandable for Cardiologists and show the peaks extracted from the Neural network model so they can see how accurate our model would be when we constructed it.

I had to study about heart and ECG data and how it actually records the information and what it's meaning .I made a report of it and talked about all details in the paper and its database In a way that anyone can understand even non technicals. Finally the paper had some ambiguities that I needed to search about several topics like : power spectrum density and heart arrhythmias in order to understand what they have done in their research and explain it to my supervisor and Prof.Karimi - a part of the report is shown by figure 1:

This database contains heart rhythm disturbances necessary for research and is widely used for testing and comparative analysis of various algorithms for detecting arrhythmia threats. The signal is a modified limb leadII (MLII).

Lead II compares the left leg with the right arm, with the leg electrode being the exploring electrode. Therefore, lead II observes the heart from an angle of 60°

A) The limb leads and their view of the heart's electrical activity



B) Einthoven's triangle

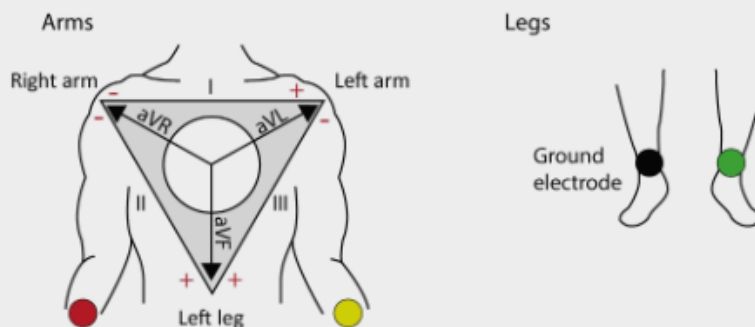


Figure 1

Wfdb was used to read the ECG data and then the output was modified to pandas dataframe then the data was ready to be given to Neurokit2 and detect the peaks and then plot it. Figure 2 shows the output:

```
[ ]
df = record.to_dataframe()

ecg_signal = df# nk.data(dataset="ecg_1000hz")
# Extract R-peaks locations
processed_data, info = nk.bio_process(ecg=ecg_signal, sampling_rate=721/5)

arr = np.array(processed_data['col 1'])
_, rpeaks = nk.ecg_peaks(arr, sampling_rate=721/5)

# Visualize R-peaks in ECG signal
plot = nk.events_plot(rpeaks['ECG_R_Peaks'], arr)
```

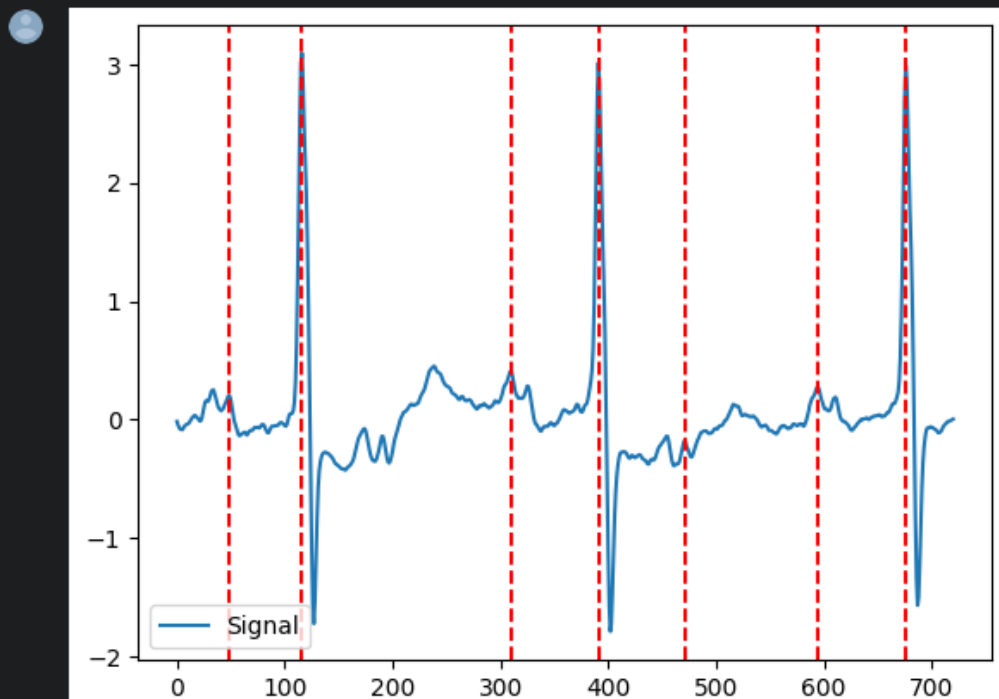


Figure 2

And made many changes to get the following output shown in figure 3 :

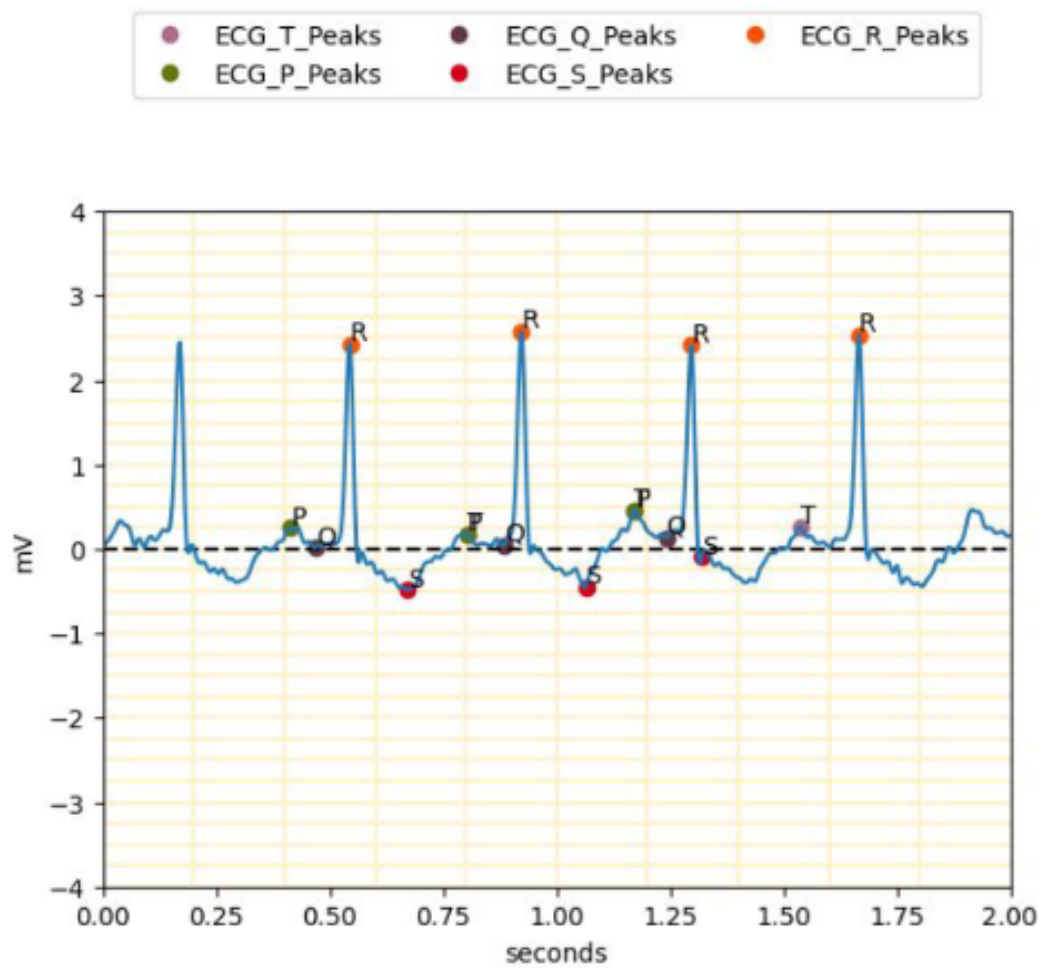
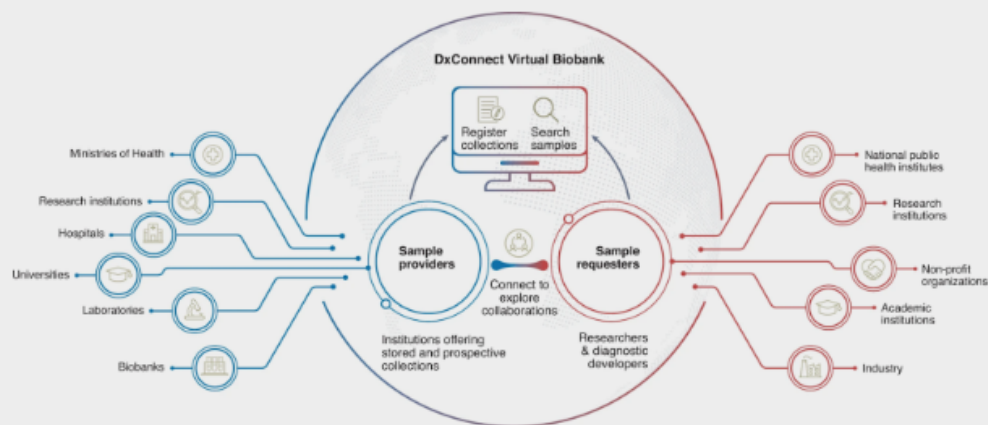


Figure 3

- I have participated in the biobank project of Shiraz medical university in which I gathered the related papers and information as Biobank multiple report.(figure 4)



A review of biobanks:

review article titled "Biobanking Past, Present and Future: Responsibilities and Benefits" that explores the field of biobanking as it has evolved from a simple collection of frozen specimens to the virtual biobank. The review discusses how biorepository and biospecimen science has evolved in response to external regulatory pressures, advances in the biological sciences, and the advent of the computer chip. Virtual biobanks are developed to assist investigators in locating biospecimens for testing and data mining from multiple biobanks in dispersed locations. Such virtual biobanks are accessed using specialized software or web portals designed to connect biobanks and investigators throughout the world¹.

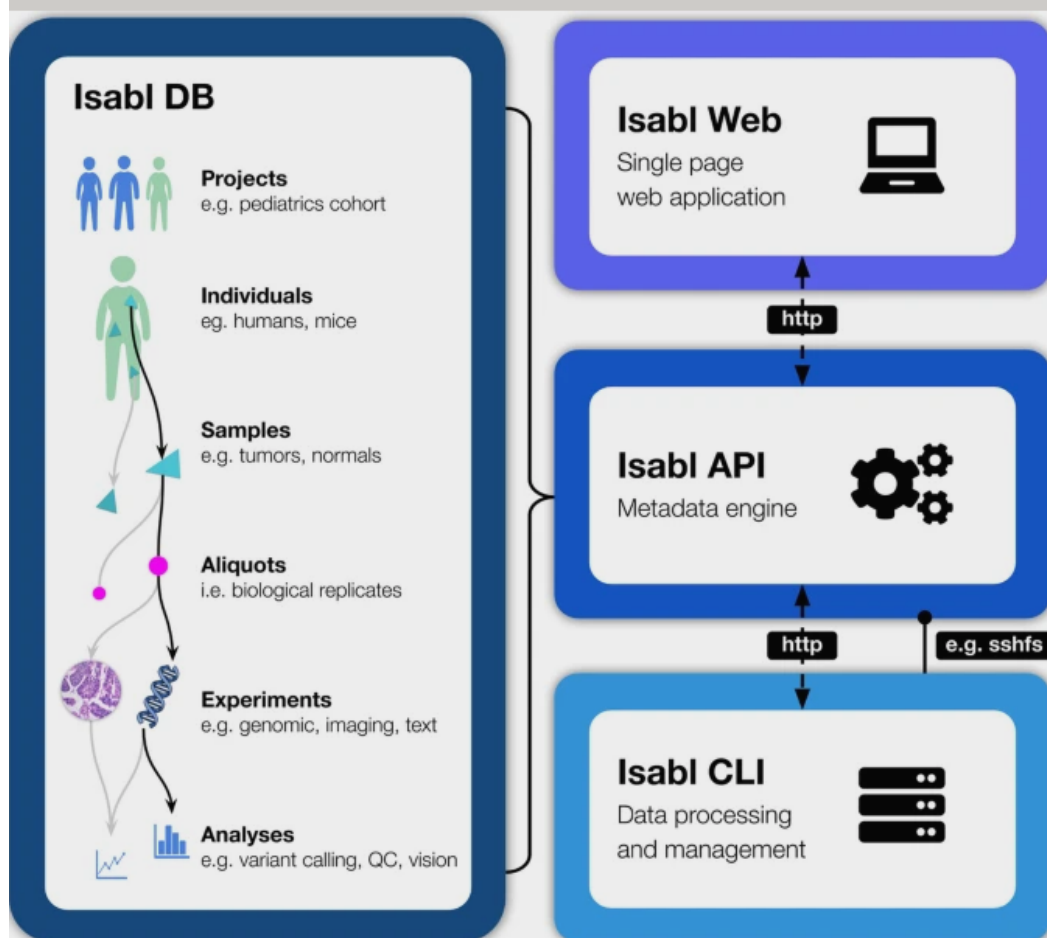
Another review article titled "The value of virtual biobanks for transparency purposes with respect to reagents and samples used during test development and validation" discusses how virtual biobanks represent a powerful mechanism for enhancing access to well-characterized material for standards and validation panels. This can enhance the quality of test development and validation pathways, allowing other parties to have greater confidence in the work done and to be able to repeat it themselves as required².

Figure 4

Next report was mainly about biobank overview and its software implementation challenges and see a literature of what others have done in biobank prospects . Figures 5 and 6

generation platforms, and analyses cohorts are also provided. Lines with one circle represent foreign keys, whilst lines with two circles represent many to many relations

b A brief description of these schemas with examples



Schematic representation of Isabl's microservice architecture. Isabl DB provides a p

Figure 5

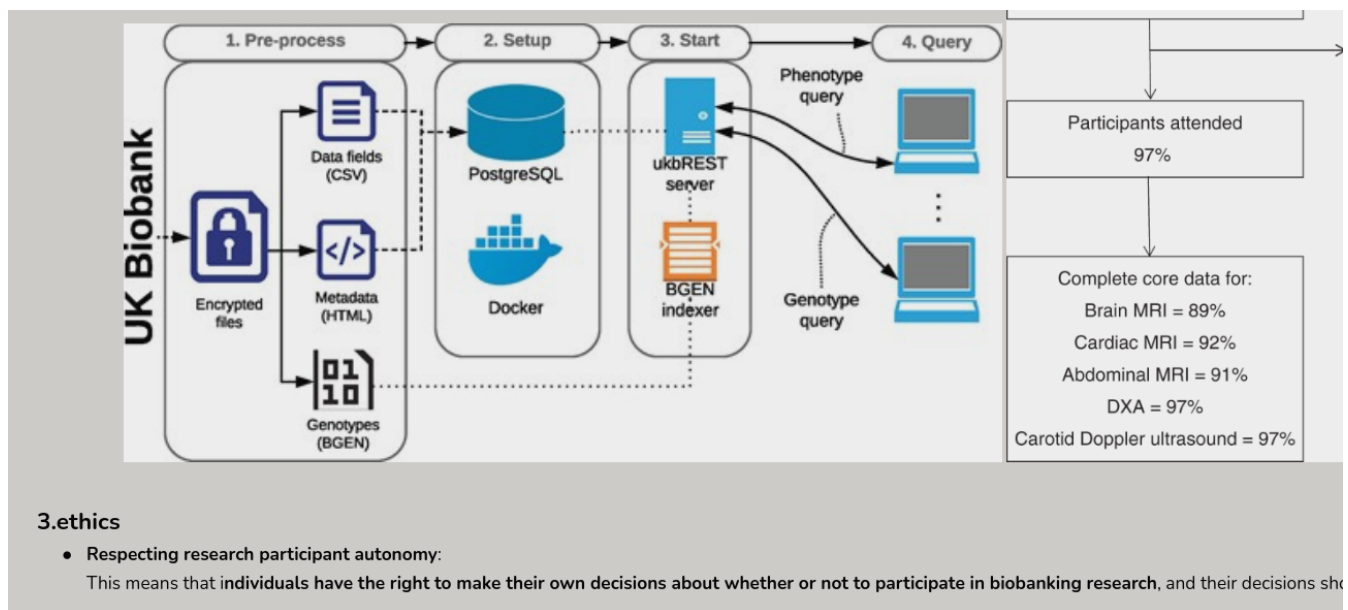


Figure 6

I attended the biobank workshop presented by Dr.Fatehi.
(shown in Figure 7)

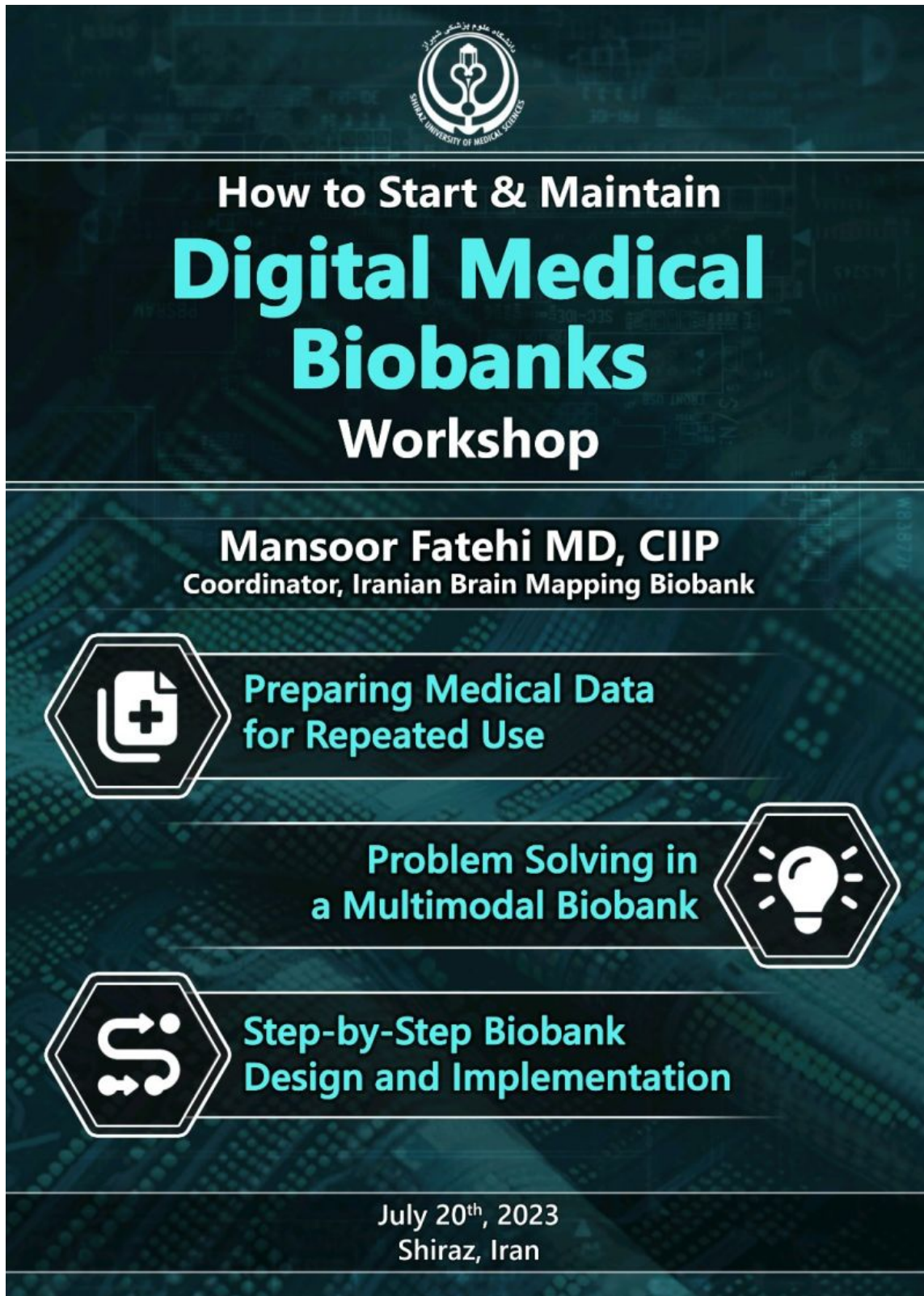


figure 7

Then I gathered all the information that I learnt in the workshop pulse reading extra papers to make a brief report on how to make a biobank step by step figures 8 and 9 show the introduction of the report:

Step of a Biobank design and implementation

Some concerns:

How to prevent Defacing? Should we ?

How to let others use our bank? Federated learning?

What structure is the right one? Dictionary is an answer.

What about the image quality?

How to actually query biodata in a way that is patients privacy is not violated and is most convenient for researchers?

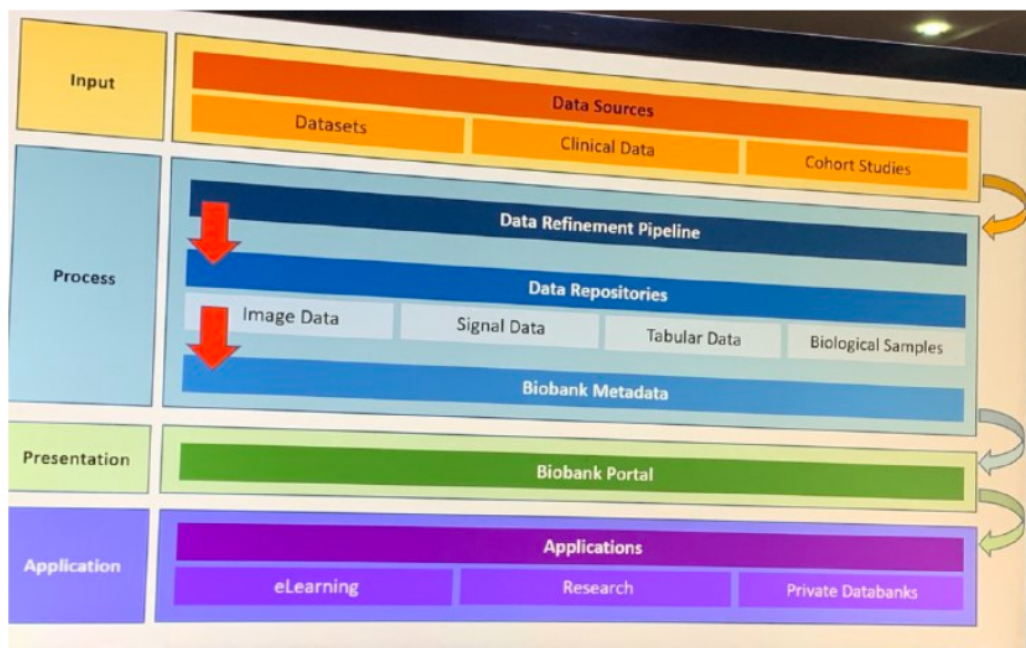


Figure 8

Main parts:

- Gathering information about devices that take images(MRI , ...)

- Making a unique structure to input our models and make every image from different places standard.

- Make sure we can accept donations form other banks

- Data cleaning

Figure 9

- Another responsibility that I had was segmenting and cleaning bioimage data using SAM (Segment anything model) which consisted of about 2500 images. Figure 10 shows the part of the python code used to segment images that sam was used.

```
import sys
sys.path.append("..")
from segment_anything import sam_model_registry, SamAutomaticMaskGenerator, SamPredictor

sam_checkpoint = "sam_vit_h_4b8939.pth"
model_type = "vit_h"

device = "cuda"

sam = sam_model_registry[model_type](checkpoint=sam_checkpoint)
sam.to(device=device)

mask_generator = SamAutomaticMaskGenerator(sam)

"""To generate masks, just run `generate` on an image."""

masks = mask_generator.generate(image)
```

Figure 10

For instance figure 11 was the input of the segment anything model and the outputs were figures 12 and 13 with some coding and changing names of the pictures.

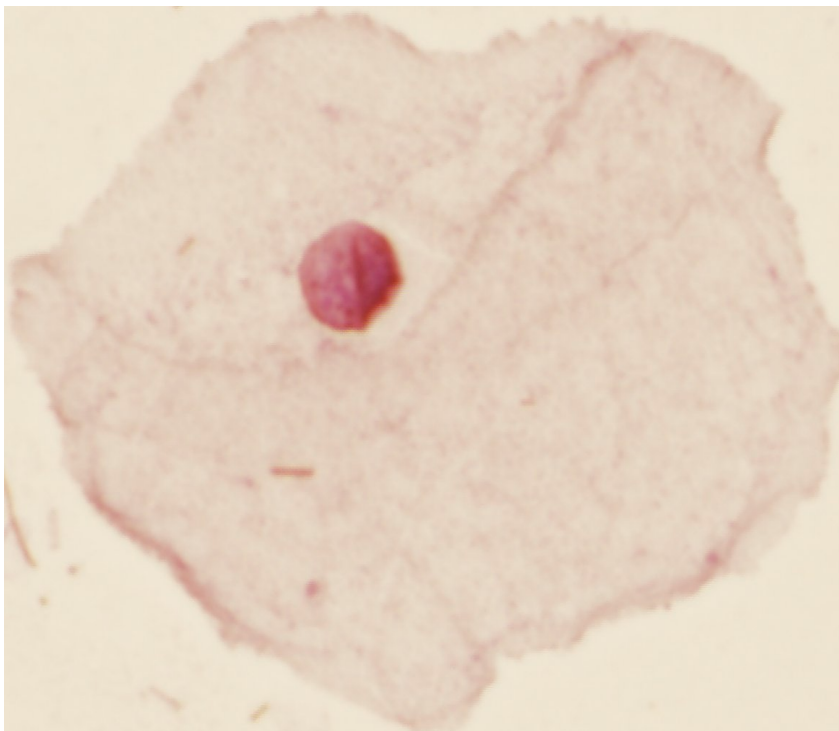


Figure 11



Figure 12

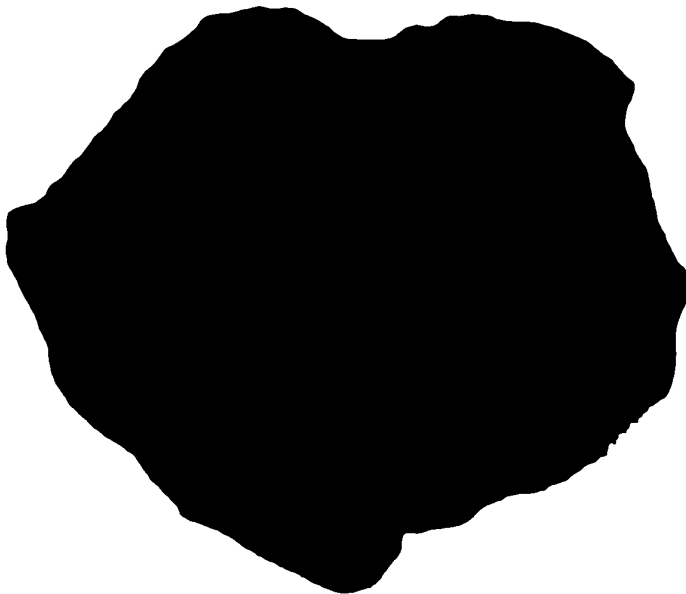


Figure 13

The main problem was to segment a cell into its nuclear and cytoplasm
No segmenting tool could distinguish these from each other except SAM

.

- Finally the most recent project of mine was to find existing solutions to the following question of Dr.Shakibafar :
“How to modify a template text over and over by just using voice in a way that there would be a need to typist.”

I constructed a report which the abstract is shown in figure 14

name	language	Voice command	price	Adopint process	Video links
Micorosoft word	English	<input checked="" type="checkbox"/>	0 in Iran	Fastest since it is already installed	link
Dragon	English	<input checked="" type="checkbox"/>	0 in Iran - 300 or 600 \$ If the cracked program was not desirable	Easy but needs to be installed	link
Chat Gpt Extension	Persian and english	Needed if voice in extension is used otherwise it's a chat to chat gpt	Virtual number price+ 40\$ for extention	Needs to sign up in openai.com and then add the extension to chrome once	link1(voice in extension) link(chatting with chat gpt with some extention)

Figure 14

Conclusion

By doing an internship under supervision of Dr.Azimifar I've learned how can doing projects with nontechnical people and customers can be tricky and sometimes time consuming and in order to save yourself lots of time and avoid conflict you need two things ; first clearly write the problem and inputs and outputs when both parties came to an agreement and secondly the importance of visualising data, preparing a simple demo can help the customer to decide to invest on a specific project or not and if not the sooner they know the better and more money can be saved. Above all technical aspects of this internship, having passion was the most valuable lesson that I've learned.

When there is no specific plan there is no specific destination so in order to do a project it's extremely important to first make a clear path and understand challenges and break the problem into subproblems and then conquer them and build a good solution.

references

ECG data was gathered from “ECG Fragment Database for the Exploration of Dangerous Arrhythmia” paper:

https://www.linkedin.com/feed/update/urn:li:activity:7083436632439173120/?updateEntityUrn=urn%3Ali%3Afs_updateV2%3A%28urn%3Ali%3Aactivity%3A7083436632439173120%2CFEED_DETAIL%2CEMPTY%2CDEFAULT%2Cfalse%29

Wfdb :

<https://wfdb.readthedocs.io/en/latest/>

Pandas:

<https://pandas.pydata.org/>

Segment anything:

<https://segment-anything.com/>

<https://github.com/facebookresearch/segment-anything/tree/main>

Neurokit2:

<https://pypi.org/project/neurokit2/>