

Configuration Manual

MSc Research Project
Master of Science in Artificial Intelligence

Karthik Uppalapati
Student ID: 23289961

School of Computing
National College of Ireland

Supervisor: Dr. Anu Sahini

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: Karthik Uppalapati
Student ID: 23289961
Programme: Master of Science in Artificial Intelligence **Year:** 2025-2026
Module: Practicum
Lecturer: Dr. Anu Sahini
Submission Due Date: 24/01/2026
Project Title: MedPredict-DeBERTa: Enhancing Symptom-Based Disease Prediction through UMLS-Integrated Disentangled Attention
Word Count: 484 **Page Count:** 7

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: U.Karthik

Date: 11/12/2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Karthik Uppalapati
Student ID: 23289961

1. Connecting a Colab session to Google Drive for file access

Step 1: Launch the colab environment

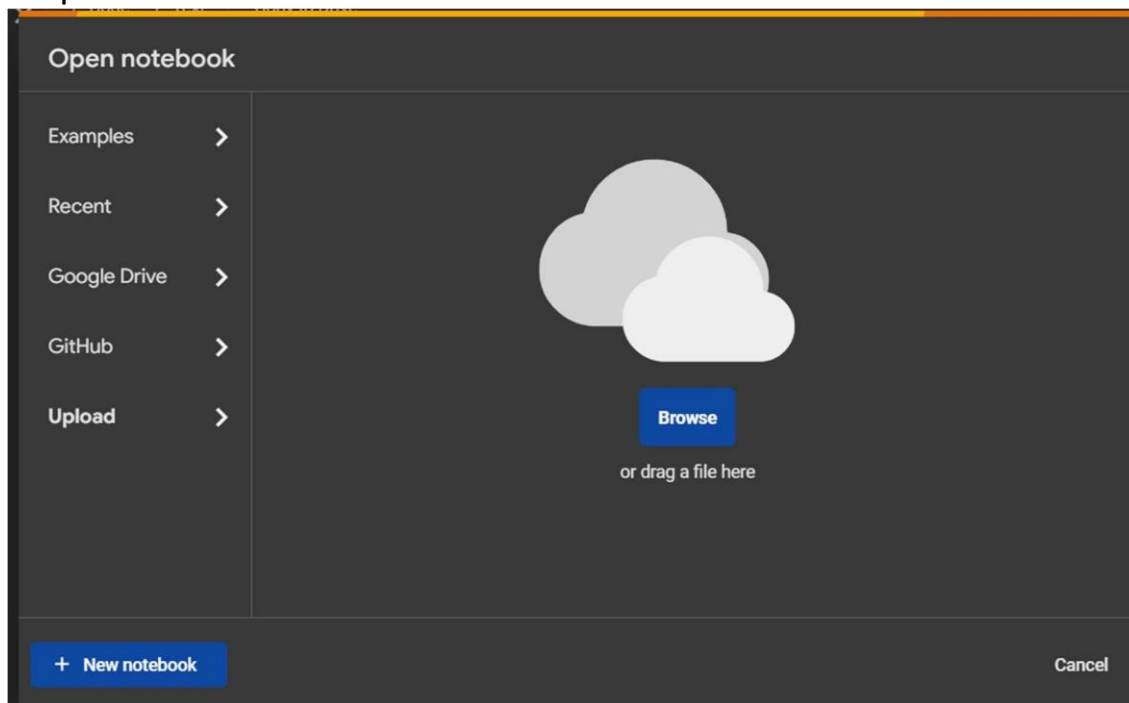
Step 2: Navigate the google colab website using google

Step 3: Now upload the File

1. Upload the ipynb or py files

Step 4: From Runtime change the GPU acceleration

Step 5: Save the runtime for next use



2. Beginning the session

- Step 1: Open Google in your browser and launch Google Colab.
- Step 2: Sign in using your Gmail account.
- Step 3: Open a new notebook or load your saved configuration file.
- Step 4: Enable the GPU runtime.
- Step 5: Upload and open your notebook file.
1. Open the data set file
 2. Now repeat the steps and open the ULM's integration file
 3. Now repeat the steps and open the Baseline
 4. Now open the ULM's analysis file
 5. Now open the UMI's Debeart file
- Step 6: Once everything is done you will get the results

Change runtime type

Runtime type

Python 3 ▼

Hardware accelerator (?)

☒ CPU ☐ A100 GPU ☐ L4 GPU ☐ T4 GPU

☐ v6e-1 TPU ☐ v5e-1 TPU

High RAM ☐

Runtime version (?)

Latest (recommended) ▼

Cancel Save

3. Ingesting the data

Step 1: Load the data

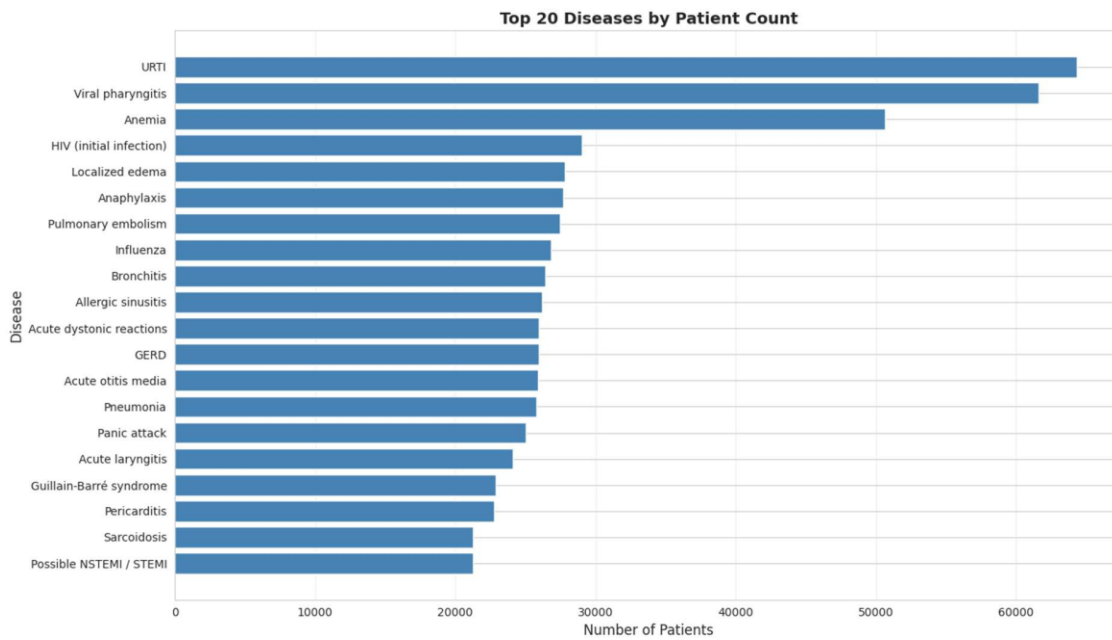
1. change drive path for the dataset and saving

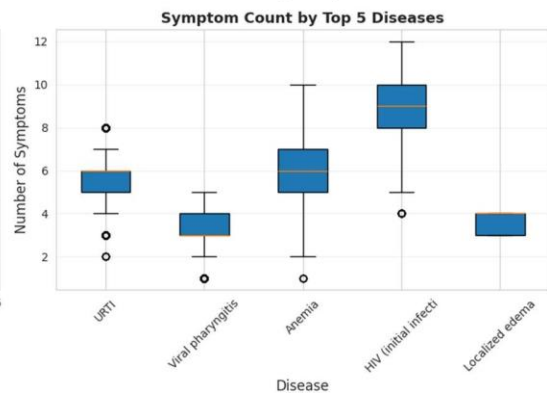
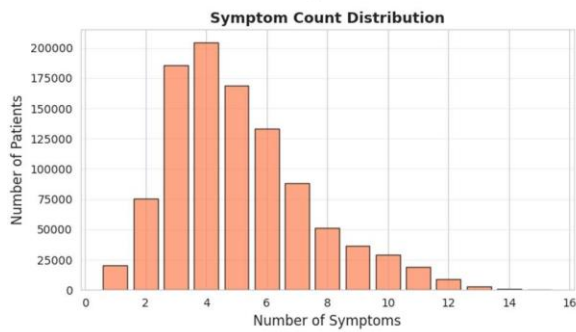
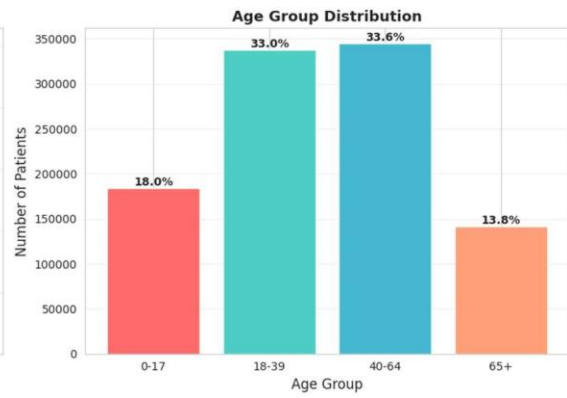
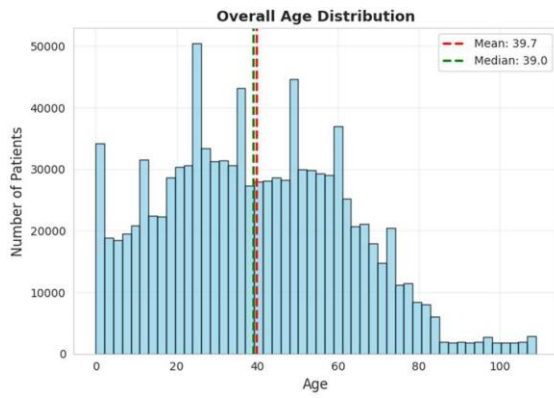
```
1 from google.colab import drive
2 drive.mount('/content/drive')

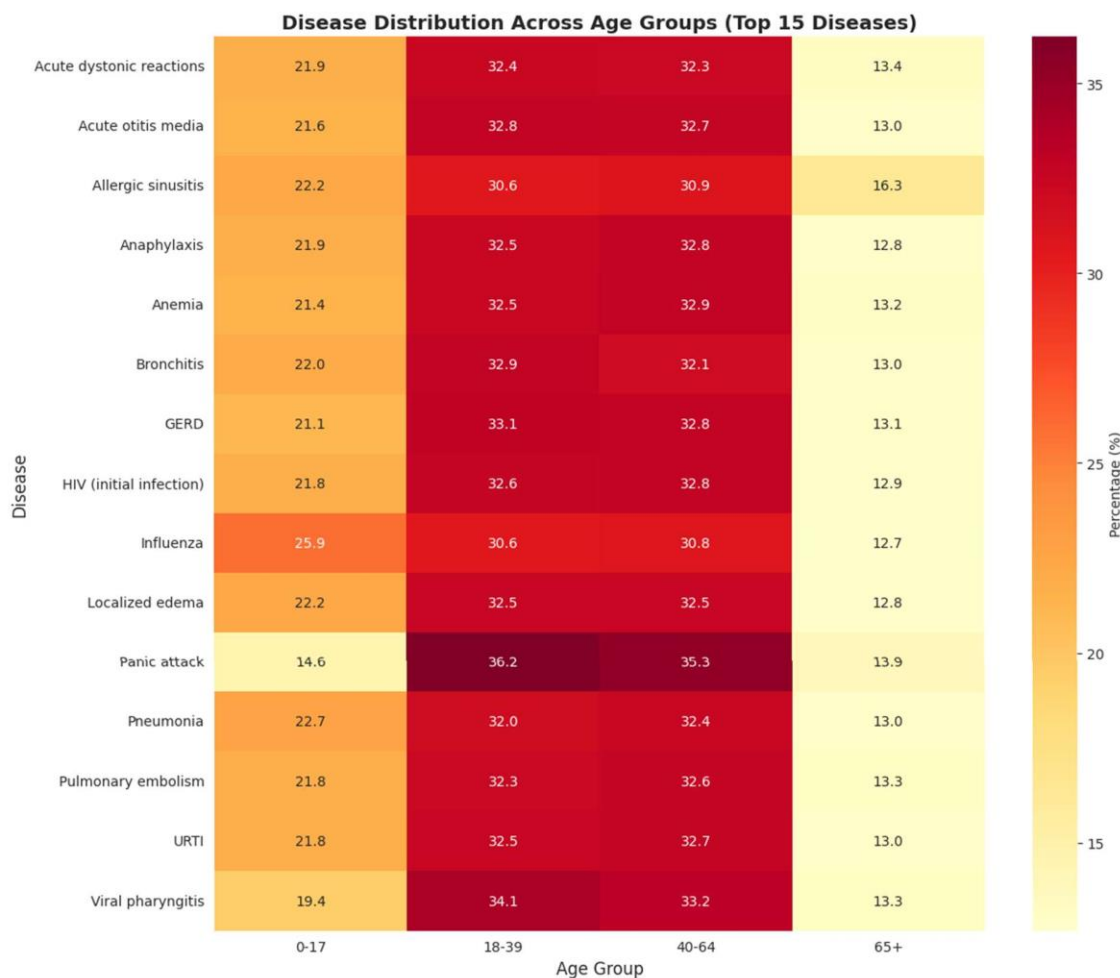
... Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

1 import os, zipfile
2
3 zip_path = "/content/drive/MyDrive/dataset.zip"
4 out_dir = "/content/dataset"
5
6 # make sure the target folder exists
7 os.makedirs(out_dir, exist_ok=True)
8
9 # extract all files
10 with zipfile.ZipFile(zip_path, "r") as zf:
11     zf.extractall(out_dir)
12
13 # quick check: list a few extracted items
14 for root, _, files in os.walk(out_dir):
15     for f in files[:20]:
16         print(os.path.join(root, f))
17     break
```

4. Results







DETAILED EVALUATION - TEST SET

Overall Metrics:

Accuracy: 0.9747 (97.47%)
 Precision: 0.9787
 Recall: 0.9718
 F1-Score: 0.9723
 Top-3 Acc: 0.9999 (99.99%)
 Top-5 Acc: 1.0000 (100.00%)

Baseline (DeBERTa-base, text only): 0.9747 (97.47%)
 Strategy 4 (DeBERTa-large + UMLS): 0.9747 (97.47%)
 Improvement: +0.00%

5 References

Google Colab. (n.d.). <https://colab.research.google.com/>

Fansi, A., Jehanno, E., Laberge, G., & Bhattacharyya, O. (2022). DDXPlus: A new dataset for automatic medical diagnosis. *Proceedings of the Conference on Health, Inference, and Learning*, 231–245. <https://proceedings.mlr.press/v174/fansi22a.html>