

Configuration Manual

MSc Research Project
MSc in Cybersecurity

Srikanth Gattu
Student ID:x23421380

School of Computing
National College of Ireland

Supervisor: Niall Heffernan

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: Srikanth Gattu

Student ID: X23421380

Programme: MSc in Cybersecurity **Year:** 2025-2026

Module: MSc (Research) Practicum/Internship Part 2

Lecturer: Niall Heffernan

Submission Due Date: 11-12-2025

Project Title: SS-CURL: SELF-SUPERVISED CONTRASTIVE URL LEARNING FOR ADVERSARIALLY ROBUST PHISHING DETECTION

Word Count: 174 **Page Count:** 5

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: Srikanth Gattu

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

Srikanth Gattu
X23421380

1. Authorizing Colab to access files in Google Drive.

Step 1 : Launching the colab environment (navigating to google colab website)

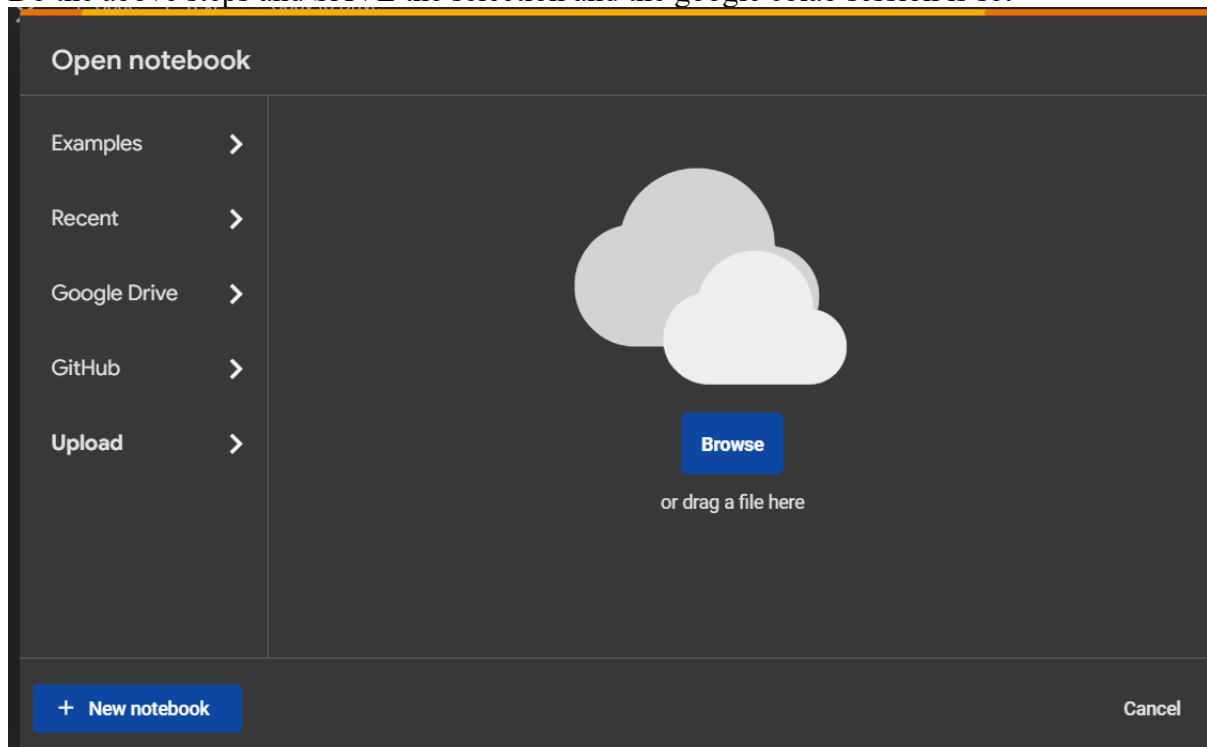
Step 2: Sign in into the google account

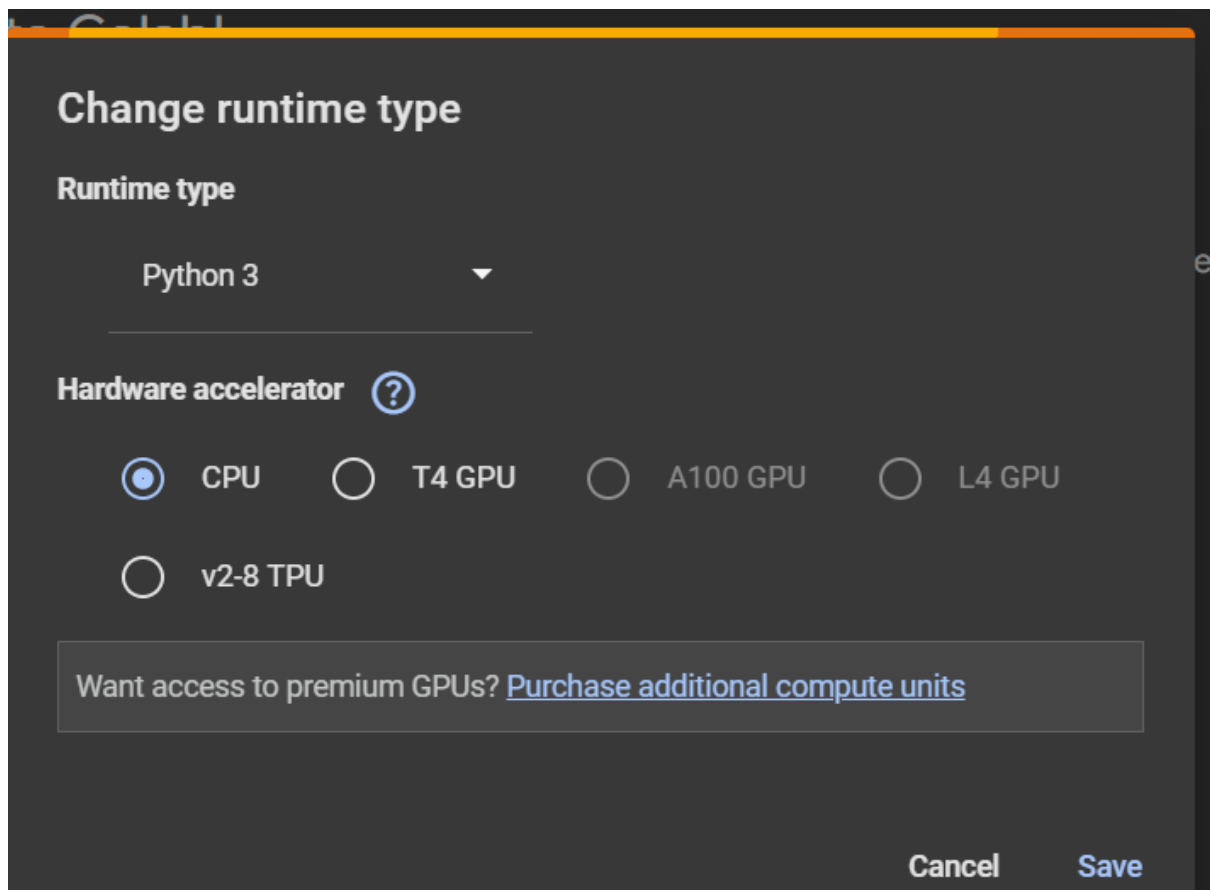
Step 3:uploading the ipynb file /notebook

Step 4: Enabling the GPU Acceleration (in this change the runtime)

Step 5 : change the hardware accelerator to T4 GPU

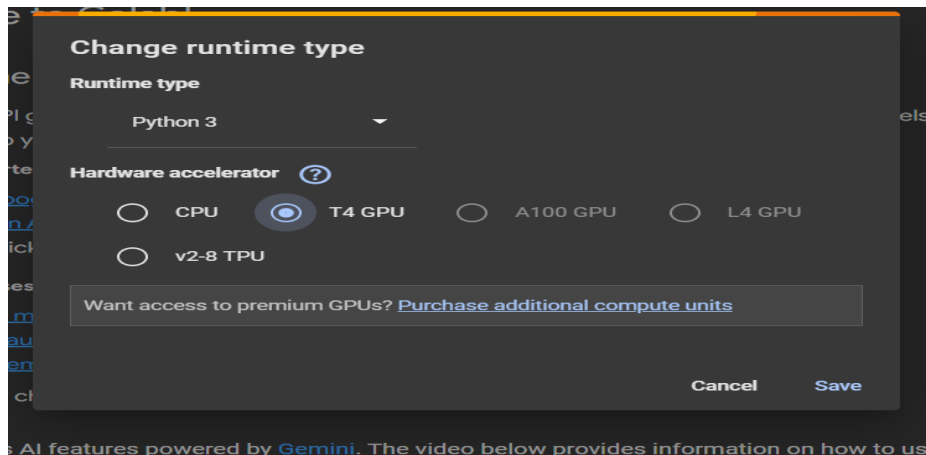
Do the above steps and **SAVE** the selection and the google colab session is set





2. Spinning up the runtime

- Step 1: open the google colab session
- Step 2: Sign in to your Gmail
- Step 3: start a new note book
- Step 4: Activate the GPU acceleration
- Step 5: Create a folder on the website
 - 1. Upload the document
- Step 6: You can see the results now



3. Fetching and loading the dataset

Step 1 : load the data into the notebook(PhiUSIIL_Phishing_URL_Dataset)

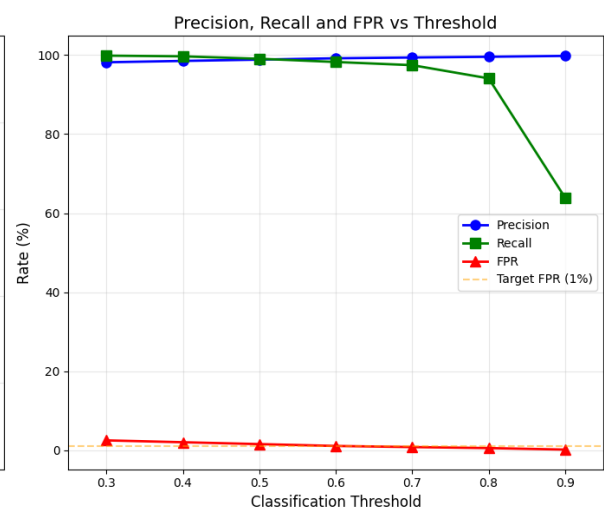
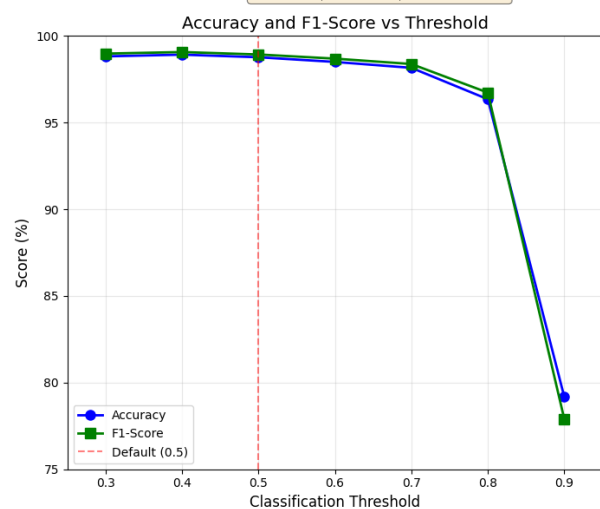
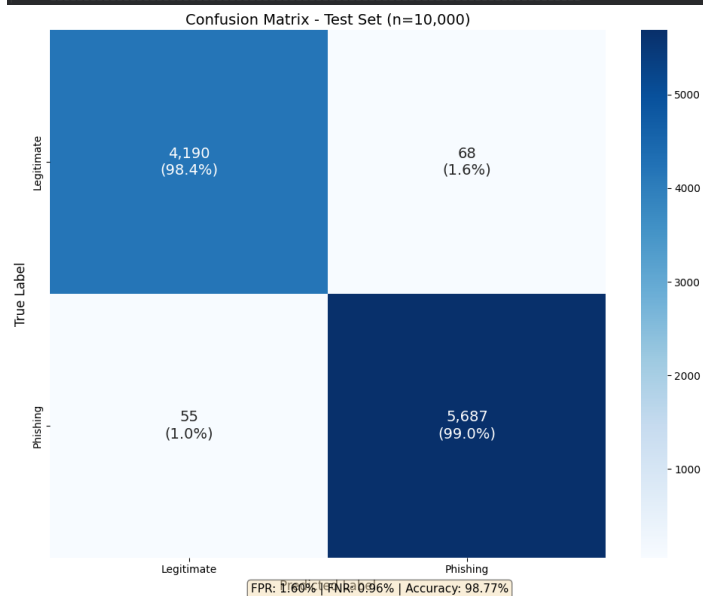
```

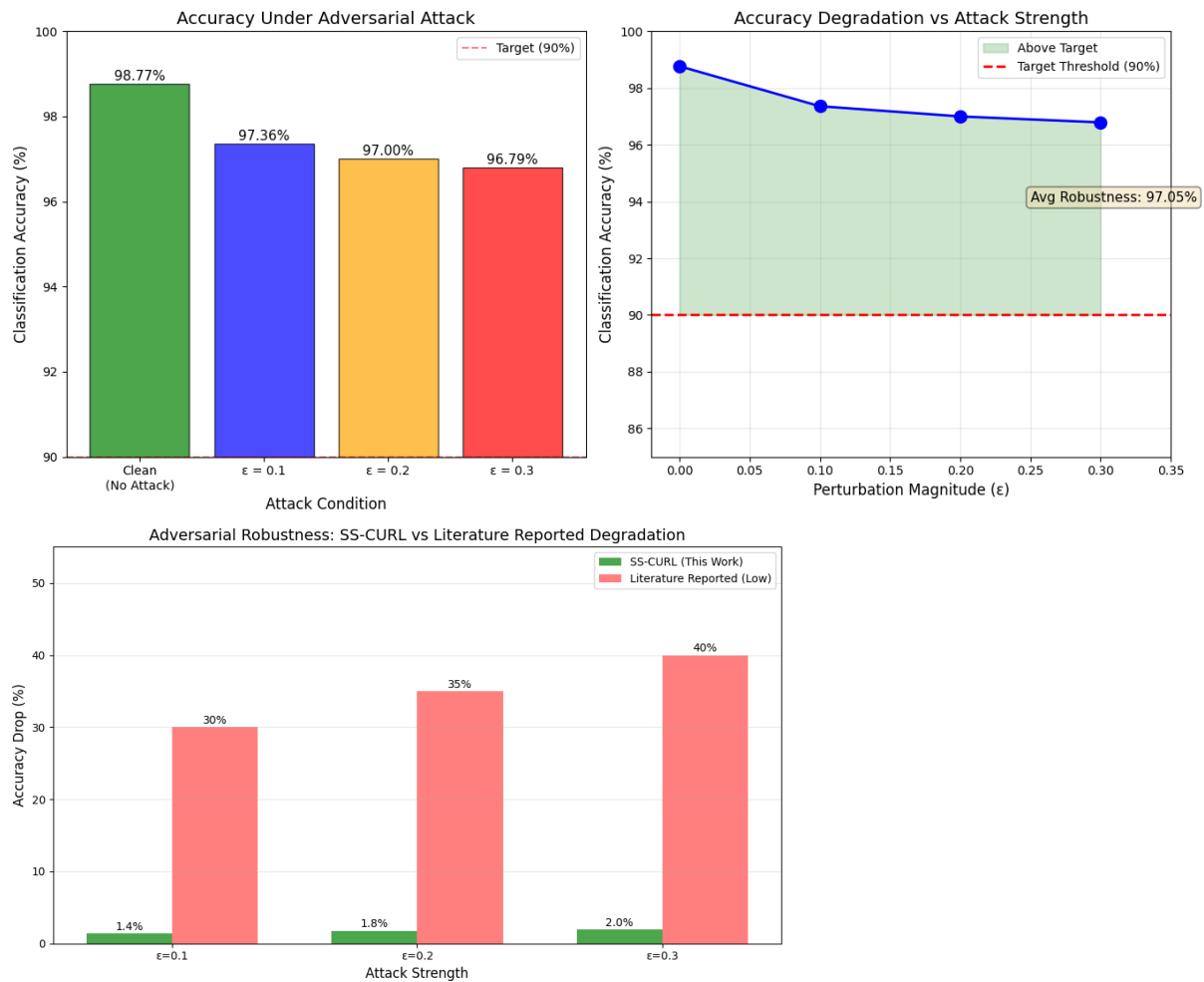
1  # Optimized configuration for better GPU utilization and results
2
3  class Config:
4      """Improved Configuration for SS-CURL"""
5
6      # Random seed
7      SEED = 42
8
9      # Data paths
10     DATA_PATH = "PhiUSIIL_Phishing_URL_Dataset.csv"
11
12     # Model architecture - IMPROVED
13     BERT_MODEL = "bert-base-uncased"
14     MAX_URL_LENGTH = 256
15     HIDDEN_SIZE = 768
16     PROJECTION_DIM = 256 # INCREASED for richer representations
17     NUM_HEADS = 12
18     NUM_LAYERS = 6
19     DROPOUT = 0.25 # Balanced dropout
20     ATTENTION_DROPOUT = 0.1 # NEW: Separate attention dropout
21
22     # Training - Pre-training - OPTIMIZED FOR FAST TRAINING
23     PRETRAIN_EPOCHS = 5 # Reduced - loss converges quickly
24     PRETRAIN_BATCH_SIZE = 512 # Large batch for GPU
25     PRETRAIN_LR = 1e-3 # Higher LR with large batch
26     PRETRAIN_WARMUP_RATIO = 0.1
27
28     # Training - Fine-tuning - OPTIMIZED FOR FAST TRAINING

```

4. Final Results

```
10 print(" results = main_training_pipeline(run_pretrain=False, run_finetune=True)")
11 print("\nTo predict URL:")
12 print(" predict_url('https://example.com')")
13 print("\nTo export metrics:")
14 print(" export_metrics_to_csv(results)")
15 print("\nTo view model summary:")
16 print(" print_model_summary()")
17 print("="*80)
18
19 # Run complete training pipeline
20 results = main_training_pipeline(run_pretrain=True, run_finetune=True)
21
22 # Export metrics to CSV after training
23 if results and 'finetune' in results:
24     export_metrics_to_csv(results)
25
26 ...
27
28 =====
29 SS-CURL Implementation Ready
30 =====
31
32 To run training:
33 results = main_training_pipeline(run_pretrain=True, run_finetune=True)
34
35 For quick test (skip pre-training):
36 results = main_training_pipeline(run_pretrain=False, run_finetune=True)
37
38 To predict URL:
39 predict_url('https://example.com')
40
41 To export metrics:
42 export_metrics_to_csv(results)
43
44 To view model summary:
45 print_model_summary()
46
47 =====
48 SS-CURL Training Pipeline - IMPROVED VERSION
49 =====
```





5. References

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

Haq, Q.E.U., Faheem, M.H. and Ahmad, I. (2024) 'Detecting Phishing URLs Based on a Deep Learning Approach to Prevent Cyber-Attacks', *Applied Sciences*, 14(22), p. 10086. doi: 10.3390/app142210086.