Name: Sahil Bharodiya

College: Indian Institute of Information Technology Sonepat

Brach: Computer Science and Engineering

Email: sahilbharodiya.ppsv2@gmail.com

Telephone: +91 9173661451

Instant Massaging: +91 9173661451 (WhatsApp or Signal)

Country of Residence: India

Time Zone: IST (UTC + 5:30)

Language of Communication: English

LinkedIn: Sahil Bharodiya | LinkedIn

GitHub: https://github.com/SahilBharodiya

I am a second-year (will complete in May 2022) regular degree student pursuing a B.Tech in Computer Science and Engineering (specialization in Machine Learning and Deep Learning) at the Indian Institute of Information of Technology, Sonepat. My semester will be complete in May 2022 leaving me enough time to get ready for my GSoC project. If I am selected, I shall be able to work around 30 hrs a week on the project, though am open to putting in more effort if the work requires it.

Why this Organization?

Most people are spending their time in the ML for business purposes. But the base behind the ML is dependent on science. By putting ML in science, we can achieve a business purpose. For that first, we need to use ML in science. This idea affected me. And I started my ML journey in science. While most GSoC organizations follow the method of bug fixing and new feature addition, ML4SCI follows the core values of science. I am as an engineering student with a science background, I would like to contribute to ML4SCI with my maximum efforts to achieve the goals of the ML4SCI organization.

Why the DeepLense project?

As I know DeepLense is a deep learning pipeline for particle dark matter searches with strong gravitational lensing. I have already read the paper under [1909.07346] Deep Learning the Morphology of Dark Matter Substructure (arxiv.org). I found this project interesting according to my technical knowledge of Machine Learning and Deep Learning. Moreover, I also tried the evaluation test for this project and I found very satisfying results related to this paper mentioned above on DeepLense. So, finally, I have committed to contributing and am ready for further development in this DeepLense project with all my effort.

Why me for this DeepLense project?

When the evaluation tests were released for this project, at the same time I picked up them and tested them on my hand. As I previously said that I have already read DeepLense's paper on arxiv, I found my results satisfactory on the amount of data provided for evaluation. I have achieved an AUC of about 0.95 for Common Test I. Multi-Class Classification and an AUC of about 0.93 for Specific Test IV. Exploring Equivariant Neural Networks. I assure you that if I am provided more data than the results will be more satisfactory.

My project timeline

Please notice that the project timeline mentioned below can be made flexible for mentors. This means that I am ready to give my full time to this project on holidays and weekends.

→ Week 1 [May, 31 - Jun, 12] Community Bonding period

(Because I have my end semester examination from 18th May to 30 May, my week 1 starts on 31st May.)

First of all, I will communicate with the mentor of my corresponding projects and know about further processing, their requirements, expectations, rules, etc. I will tell them my project requirements like data resources and expected time for coding. I also tell them about results that were achieved on evaluation data. If they give any suggestion regarding improvements or anything then I will consider it and goes to my work.

→ Week 2 [Jun, 13 - June, 19] Deep Analysis of the Prospective

I will first dive deeply into the given project of DeepLense. I will carefully analyze it. See what it says. What is it all about? What is the science behind it? What solutions exist for it currently? How ML can solve this?

→ Week 3 [Jun, 20 - June, 26] Look at the Data Provided

During this week, the data that they will provide I will look at that. Will analyze it. Extract meaningful information like frequency count, distribution, simulated or not, etc. Then I will be able to decide how I will start training. During this phase, I will also pre-process the data that actually can be implemented into the training phase. Finally, I will convert the data into a useful format for training.

→ Week 4 [Jun, 27 – Jul 4] Report 1 submission

I will submit a report, Report 1 based on previous three-week developments. I.e. brief description of the problem statement, and an overview of the data that is finalized for further processing.

→ Week 5 [Jul, 5 – Jul, 11], Week 6 [Jul, 12 – Jul, 19], Week 7 [Jul, 20 – Jul, 26] Training Phase

Let's finally come to our most important part of the project training. The useful data that I have created in previous weeks, will be used for training. Throughout the project, I will use Python as the programming language and TensorFlow as my main framework for this task. My initial trial is to create a model described in the DeepLense paper on arxiv (ResNet, AlexNet, VGG). After that, I will go for the model that I created during the evaluation test (Xception). For training purposes, I will use Google Collaboratory or Kaggle notebooks. If required then I am ready to use cloud resources like Google Colab Pro or AWS Sagemaker which are more powerful. I will use GPU and TPU during training.

→ Week 8 [Jul, 27 – Aug, 2], Week 9 [Aug, 3 – Aug, 9] Optimizing Phase

In these two weeks, I will try other methods like data augmentation if results during the training phase are not satisfactory. Also, I will try to train the model on data that has lower over the higher resolution of images.

→ Week 10 [Aug, 10 – Aug, 16] Report 2 submission

Now, time for our next report, Report 2. In this report, I will include all the strategies that I have applied in the training and optimizing phase.

I will include why I have chosen this or that model for training. Report 2 also includes the learning curves like loss improvements, Receiver Operating Characteristics (ROC) curve, and Area under the Curve. Results are also compared with the results mentioned in the DeepLense paper.

→ **Week 11 [Aug, 17 – Aug, 23] Further Processing Related to Accuracy Measurement improvements**

If mentors fill that this model should have more satisfactory results like more AUC, then I will try to achieve this.

→ **Week 12 [Aug, 24 – Aug, 30] Finalisation of the Models**

Now I have completed the complete model training and optimization. The time has come to pick up the best model that is most accurate. After the selection of the best model, I will convert them as required. The model directly can be saved and also its only weights are saved.

→ **Week 13 [Aug, 31 – Sep, 6] Final Report submission**

This is the final report in which I include All the aspects of the project, approach, metadata of the data and model, accuracy plots, ROS, and AUC. Final report will also include GitHub repo links, notebook links, etc.

→ **Week 14 [Sep, 7 – Sep, 12] Thanks**

After the completion of the task, I am eagerly waiting for the response related to my behaviour, approach, performance, etc from mentors.

My Projects on Science using ML

1) Brain Tumour Detection (Computer Vision, Deep Learning, AI in Health) – [Self]

   A complex neural network will detect the tumour present in MRI images of patients. With an accuracy of 94%.

2) Detection of Criminal Activities/Criminal through CCTV by analysing live footage for melee, mob formation, the body language of the suspect, etc. (Deep Learning, AI in Security)– [Self, SIH2022 Problem Statement]

   The state CCTV Control Room receives feeds from a number of CCTV Cameras across the state. It is not possible to monitor all the camera feeds in a live scenario. The solution Should be capable of reading the feeds being received from the cameras and analysing the feeds for any criminal activity. I developed an ML model that identifies many activities like robbery, Abusing, Vandalism, Burglary, etc. from real-time CCTV footage.

3) Early Diabetes Prediction (Deep Learning, AI in Health) – [Self]

   A Simple DNN based model predicts diabetes in early stages based on age, gender, polyuria, polydipsia, sudden weight loss, weakness, polyphagia, genital thrush, visual blurring, itching, irritability, delayed healing, partial paresis, muscle stiffness, alopecia, and obesity. With an accuracy of 96%.

4) Estimating Aqueous Solubility Directly from Molecular Structure (AI in Drug Discovery) – [Self]

Predicting log of solubility LogS based on molecule's Wildman-Crippen LogP value, molecular weight, number of rotatable bonds, and aromatic proportion.

5) AI-based real-time identification of Crop Nutrient Disasters in agricultural crops (Computer Vision, Deep Learning, AI in Agriculture) – [with FOSS4Gov and VJHackethon]

To develop an intelligent system for the detection of nutrient deficiency in major crops taken in a local area.

Online Courses and Certification

Machine Learning [Stanford University - Coursera]

Master The C language [Udemy]

Deep Learning [IBM]

Applied Data Science [IBM]

Problem Solving Basic [HackerRank]

Problem Solving Intermediate [HackerRank]