PES UNIVERSITY	**PES UNIVERSITY** **(Established under Karnataka Act No. 16 of 2013)** **100 Ft. Road, BSK III Stage, Bengaluru – 560 085** **DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING** **SESSION: Aug-Dec 2021**

Course Title: Ethical Algorithm Design	
Course code: UE18CS400SJ	
Semester : VII	**Team Id: 18**
SRN: PES1201800790	**Name: Varun Kadam**
SRN: PES1201801580	**Name: K. Sreesh Reddy**

ASSIGNMENT REPORT

Problem Statement

Create a classifier of images that would try to predict whether CT Images scanned for diagnosing COVID-19 are positive or negative in nature.

Description

In a real-world scenario, especially medical ones, it is considered essential that the sensitive data of individuals be kept as private and personal as possible. In our case, we consider a scenario wherein we are presented with images of CT scans and are asked to label them as COVID-positive and COVID-negative.

Dataset Description:

Our data has two folders, images and labels:

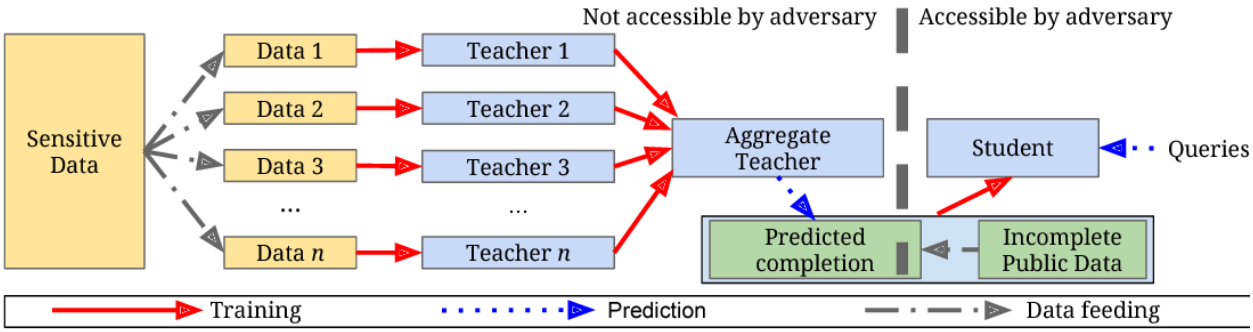
- The images folder contains all the images in png format.
- The labels folder contains text in which all the images are divided into a training dataset, testing dataset and validating data set.

This data is created using a custom data loader and imported from a COVID dataset on Github. COVID datasets aren't usually publicly available and rightly so, since the primary matter of concern is privacy.

Approach:

To answer questions such as how can you make use of these datasets without having direct access to them and how do you ensure that the data of these hospitals' patients is secure, we will use **Differential Privacy**. It works by trying to add noise either locally or globally, i.e. it can be applied at the input level or the output level. To calculate the right amount of noise that needs to be added, we use Privacy Budget. The **Privacy Budget** sets a restriction on how much data different systems can access or detect in a user's browser. In other words, the Privacy Budget permits the browser to divulge information up to a particular point based on pre-set thresholds before blocking access to or detecting more bits of data. It is represented using the symbol epsilon (ε).

The exact method which will be used to carry out differential privacy is known as **PATE** (Private Aggregation of Teacher Ensembles) whose definition is predominantly in its name. By carefully coordinating the activities of numerous independent ML (which are also known as teachers) models, the PATE system provides private learning. The resulting data will be combined and used to train a new public model (also known as student models) using unlabeled public data. The total resultant model will have measurable privacy assurances. As a result, using DP on instructors' responses can be seen as a proxy for protecting sensitive data privacy.



We also use the **RNM** (Report Noisy Max) technique to inject random noise into each model's output. This strategy ensures a significant and reliable level of privacy.

Since teacher models are to be trained using disjoint datasets, we segregate our dataset into 5 parts. To better simulate a real-world scenario, we treat these 5 parts as 5 hospitals that have been kind enough to provide us with data but would not want us peeking around said data. The basic flow of the process would be:

- ➔ The hospitals are asked to provide the datasets on which the models are trained. These obtained models are called the Teacher models.
- ➔ Since we are using multiple teacher models, for each image whose label is being predicted, we get 5 labels.
- ➔ To get the best output, we aggregate the outputs/labels and consider the majority and then add noise to them to render them differentially private.
- ➔ These aggregated results are now used to help train the student model which will finally be deployed.

The structure of the datasets which will be followed and applied is:

Whole Training Dataset	Dataset used for training teacher models.
Whole Testing Dataset	The Student dataset.
Whole Validating Dataset	Testing the final model’s performance.

The model being used is a simple 2D Convolutional Neural Network whose parameters will remain exactly the same for checking the accuracy without differential privacy and with differential privacy.

Output Screenshots

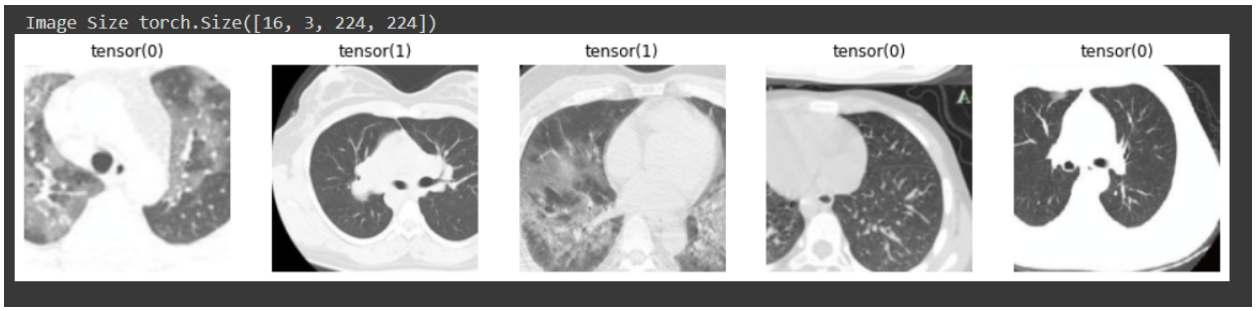


Fig: Example of some CT Scans being labelled as 0 or 1.

```python
class SimpleCNN(torch.nn.Module):
    def __init__(self):
        super(SimpleCNN, self).__init__() # b, 3, 32, 32
        layer1 = torch.nn.Sequential()
        layer1.add_module('conv1', torch.nn.Conv2d(3, 32, 3, 1, padding=1))

        #b, 32, 32, 32
        layer1.add_module('relu1', torch.nn.ReLU(True))
        layer1.add_module('pool1', torch.nn.MaxPool2d(2, 2))
        self.layer1 = layer1
        layer4 = torch.nn.Sequential()
        layer4.add_module('fc1', torch.nn.Linear(401408, 2))
        self.layer4 = layer4

    def forward(self, x):
        conv1 = self.layer1(x)
        fc_input = conv1.view(conv1.size(0), -1)
        fc_out = self.layer4(fc_input)

        return fc_out
```

Fig: Screenshot of the CNN Model.

```
(163, 5)
[1 1 1 1 1]
```

Fig: Screenshot of the shape of the predicted labels and 1 example with 5 teacher models giving their labels which are later aggregated as 1.

```
[0 1 0 1 1 1 0 0 1 1 1 0 1 0 1 0 0 0 1 0 1 1 1 0 1 1 0 0 1 1 1 1 1 1 1 0 1 0
 0 1 0 1 1 1 0 0 0 0 0 1 0 0 1 1 0 1 0 1 0 1 0 0 1 0 1 0 0 1 1 1 1 1 0 0 0 1 0
 0 1 0 0 0 1 1 0 1 0 1 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 0 1 1 0 0 1 1 1 1 1 0
 1 0 0 1 0 1 0 0 1 0 1 1 0 1 0 0 1 1 1 0 0 0 1 0 1 0 1 0 0 0 0 1 0 0 0 0 1 1 0
 1 1 0 1 0 1 1 0 0 0 1 1 0 0 0]
(163,)
```

FigL Screenshot of our 163 predicted labels.

```
Student Model
        Test Loss: 1.292508
        Test Accuracy: 59% (70/118)

======================
Normal Model
        Test Loss: 1.116361
        Test Accuracy: 65% (77/118)
```

Fig: Accuracy of both models.

Interpretation of efficiency

Although the accuracy of the model in the differentially private scenario is 6% lesser than the normal deep learning model, that kind of difference can be avoided to compensate for no loss of privacy.

Learning Outcome

Ethical Issues of the dataset:

There are currently few open-source datasets available for COVID-19 diagnostic purposes. It's understood that gaining access to such datasets is difficult due to patient confidentiality and legal concerns. This shows the importance of privacy-preserving approaches. We have therefore used a publicly available and scarce COVID-19 dataset.

Differential privacy is achieved by adding randomized noise to a cumulative query, which results in individual items being saved without altering the result.

Differentially private algorithms ensure that attackers can learn almost no more about a person than they would if that person's record were not there in the dataset. Differentially private algorithms are active in the field of research and especially medicine. Its versatile definition allows it to be used in a wide range of applications.

Name and Signature of the Faculty