Name: Swarnadeep Bhattacharya

Courses atteneded and corresponding marks:

Semester I		
Subjects	**Full Marks**	**Marks Obtained**
AI-ML	100	90
Cryptology-I	100	94
Automata and Formal Language	100	91
Discrete Mathematics	100	84
Semester II		
Advanced Cryptology-II	100	40
Design and Analysis of Algorithm	100	93
Quantum Information and Cryptology	100	87
Research Methodology	100	81
Total Marks	**800**	**660**

Research Papers read:

1) Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn,Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer,Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby. AN IMAGE IS WORTH 16X16 WORDS:TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE, *arXiv preprint* *arXiv:2010.11929v2*.

2) Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez,Łukasz Kaiser, and Illia Polosukhin. Attention is all you need**.** In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, pages 5998–6008, 2017.

3) Shin Ando and Chun Yuan Huang. Deep Over-sampling Framework for Classifying Imbalanced Data**.** In *21st PKDD / 28th ECML 2017*, pages 770-785, 2017.

4) Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint* *arXiv:1607.06450*, 2016.

5) Johnson, J.M., Khoshgoftaar, T.M. Survey on deep learning with class imbalance. *J Big Data* **6,** 27 (2019). https://doi.org/10.1186/s40537-019-0192-5

Study and Research Work:

I'm furnishing below, in brief, the topics I've studied and have been studying and working on related to my research work in the field of Vision Transformer:

1) Artificial Neural Network (ANN) primarily from the Machine Learning course taught by Andrew Ng in Coursera platform and I've implemented it from scratch in python using Numpy library. In course of doing this I also learnt the concept of **Backpropagation** and incorporated it in my ANN code.

2) Convolutional Neural Network (CNN) primarily from the course 'CS231n: Convolutional Neural Networks for Visual Recognition' of Stanford University. I've implemented the same from scratch using PyTorch library.

3) Recurrent Neural Network (RNN) referring to the Andrej Karpathy's blog and various blogs from 'towards data science' website and implemented the same from scratch in python using Numpy library.

All the above codes have been tested using the scikit learn toy dataset 'handwritten digits' with performance accuracy over 90%.

4)
a) Long Short Term Memory(LSTM) networks, a special kind of RNN referring primarily to the excellent blog of Christopher Olah.
b) Learnt the concept of **Encoder-Decoder structure** built on top of LSTM and the way they are employed to design **sequence to sequence (seq2seq) model** for word level Neural Machine Translation from 'towards data science' website.

5) Subsequently I've studied and implemented the following topics from the Dive into Deep Learning online book. I've also referred to the excellent write up by tensorflow's documentation page.
a) Attention Pooling
b) Attention Scoring Function
c) Multi-Head Attention
d) Self-Attention and Positional Encoding

6)The concepts learnt from above helped me to go through the paper **'Attention Is All You Need' by Vaswani** et al. to understand the concept of **Transformer architecture**.
Equipped with all the above concepts learnt, I've gone through the highly influential paper **'AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE'** with an objective of the applications of the Transformer architecture in the domain of **Vision Transformer (ViT)** which attains excellent results compared to state-of-the-art convolutional networks requiring reasonably fewer computational resources to train. I also have consulted and studied various sites ilike huggingface.co, blogs from 'towards data science' for an implementational overview for ViT which is quite involved from coding perspective and eventually implemented it using pytorch library.

7) Now, I'm studying the **Class Imbalance Problem** and the way it has been dealt with Convolutional Networks with an objective to primarily use ViT instead.

Research Plans:

Transformers equipped with self-attention based architectures have performed exceedingly well in various tasks in NLP domain. The computational efficiency and scalability of Transformers have made it possible to train models of unprecedented size. However, its applications to computer vision remain limited with convolutional architectures being the dominant one. Nevertheless, the Transformers' scaling successes in NLP motivated the employment of a pure transformer in **Vision Transformer(ViT)**. In doing so, an image is split into patches and the sequence of linear embeddings of these patches are provided as an input to a Transformer-encoder. ViT attains excellent results compared to the state-of-the-art convolutional networks while requiring substantially fewer computational resources to train.

Class Imbalance is the problem when the number of examples available for one or more classes in a classification problem is far less than other classes. We initially intend to study various literatures regarding the impact of class imbalance on classification performance of convolutional neural networks (CNNs).

Then we've planned to use ViT and also ViT augmented with SMOTE (Synthetic Minority Over-sampling Technique) and study how our approach influences the effect of imbalance on classification and also to perform a comparative study with the approach adopted by CNN.