April-June
2020

# CSI Adhyayan

*Discovering the Potential*

A national publication dedicated to IT education, research and student community

Computer Society of India
www.csi-india.org

# Know Your CSI

## Executive Committee (2020-21/22)

President
**Mr. R K Vyas**
(E) rkvyas06@gmail.com

Vice President cum President Elect
**Dr. Anil Kumar Saini**
(E) aksaini1960@gmail.com

Hon. Secretary
**Dr. Vipin Tyagi**
(E) dr.vipin.tyagi@gmail.com

Hon. Treasurer
**Dr. Durgesh Mishra**
(E) drdurgeshmishra@gmail.com

Immd. Past President
**Prof. A K Nayak**
(E) aknayak@iibm.in

### Regional Vice-Presidents

Region-I (2019-21)
**Mr. Arvind Sharma**
(E) arvindsha@hotmail.com

Region-II (2020-22)
**Md. Shams Raza**
(E) s_raza2000@yahoo.com

Region-III (2019-21)
**Mr. Jayant Bhide**
(E) jayantbhide2000@gmail.com

Region-IV (2020-22)
**Mr. Mukesh Kumar**
(E) Mukesh.kumar@sailcet.co.in

Region-V (2019-21)
**Mr. M. S. Prasadbabu**
(E) profmspbabu@gmail.com

Region-VI (2020-22)
**Mr. Pradeep Rathi**
(E) pjrathi61@gmail.com

Region-VII (2019-21)
**Prof. Dr. M Sundaresan**
(E) bu.sundaresan@gmail.com

### Division Chairpersons

Vacant

Division-I (2019-21)

Division-II (2020-22)
**Dr. P. Sakthivel**
(E) sakthi_velp@yahoo.com

Division-III (2019-21)
**Prof. Suresh Chand Tyagi**
(E) sct_35_2000@yahoo.com
sctyagi1963@gmail.com

Division-IV (2020-22)
**Dr. B. Chidhambararajan**
(E) dbcrajan@gmail.com

Division-V (2019-21)
**Dr. Subhash Chandra Yadav**
(E) scy123@gmail.com

### Nomination Committee (2020-2021)

**Mr. Anand Rao**
(E) nanandrao@yahoo.com

**Mr. Anil Ji Garg**
(E) aniljigarg@gmail.com

**Dr. Sharvari C Tamane**
(E) hodit@jnec.ac.in,
sharvaree73@yahoo.com

## About Computer Society of India

The seed for the Computer Society of India (CSI) was first shown in the year 1965 with a handful of IT enthusiasts who were a computer user group and felt the need to organize their activities. They also wanted to share their knowledge and exchange ideas on what they felt was a fast emerging sector. Today the CSI takes pride in being the largest and most professionally managed association of and for IT professionals in India. The purposes of the Society are scientific and educational directed towards the advancement of the theory and practice of computer science and IT. The organisation has grown to an enviable size of 100,000 strong members consisting of professionals with varied backgrounds including Software developers, Scientists, Academicians, Project Managers, CIO's , CTO's & IT vendors to just name a few. It has spread its branches all over the country. Currently having more than 500 student branches and rooted firmly at 80 different locations, CSI has plans of opening many more chapters & activity centres in smaller towns and cities of the country. The idea is to spread the knowledge, and provide opportunities to as many interested as possible.


The CSI Vision: "IT for Masses"

Keeping in mind the interest of the IT professionals & computer users CSI works towards making the profession an area of choice amongst all sections of the society. The promotion of Information Technology as a profession is the top priority of CSI today. To fulfill this objective, the CSI regularly organizes conferences, conventions, lectures, projects, awards. And at the same time it also ensures that regular training and skill updating are organized for the IT professionals. Education Directorate. CSI also works towards a global approach, by seeking out alliances with organizations overseas who may be willing to come forward and participate in such activities. CSI also helps governments in formulating IT strategy & planning.

# Table of Contents

# An Introduction to GPU Accelerated Machine Learning Algorithms using cuML

**Prathamesh V. Barve, Aarushi S. Phade, Amit D. Joshi**

Abstract

Developing machine learning deep learning products in Python has been facilitated by the availability of powerful machine learning libraries, such as scikit-learn. It becomes time consuming to prototyping tools in python with huge dataset is to be processed. Graphic Processing Units (GPUs), consisting thousands of light-weight cores, become an ideal library accelerator. Here comes the cuML library. In this article, a study that optimizes a set of machine learning algorithms on a GPU has been emphasized. Report on the performance of a popular library - cuML is given in this article. Utilizing the parallel performance advantages of a GPU, better speedup can be achieved over scikit-learn.

## Introduction:

Prototyping machine learning tools in Python is flexible and can be very efficient. Most machine learning libraries are developed to run on a CPU. When working with large data sets, given the high computational demands of many machine learning algorithms, application throughput can be limited when running on traditional (i.e., CPU-based) platforms. Designed as massively parallel architectures, GPUs have become popular when running many machine learning applications, including deep learning and computer vision. One such popular Python-based framework that support high performance computing on GPUs is cuML[3].

cuML is a suite of fast, GPU-accelerated, open source machine learning algorithms designed for data science and analytical tasks. Its API mirrors popular Sklearn library and provides practitioners with the easy fit-predict-transform paradigm without ever having to program on a GPU[2].

Algorithms running on a CPU becomes slow and cumbersome while processing large datasets. RAPIDS provides users a streamlined approach where data is initially loaded in the GPU, and compute tasks can be performed on it directly[1].

## Installation steps:

Anaconda + Ubuntu 18.04[1]:

```
conda install -c rapidsai -c nvidia -c conda-forge \

-c defaults cuml=0.13 python=3.7 cudatoolkit=10.0
```

Docker + examples[1]:

```
docker pull rapidsai/rapidsai:cuda10.0-runtime-ubuntu18.04-py3.7

docker run --gpus all --rm -it -p 8888:8888 -p 8787:8787 -p 8786:8786
\rapidsai/rapidsai:cuda10.0-runtime-ubuntu18.04-py3.7
```

Google colab[1]:

```
!git clone https://github.com/rapidsai/rapidsai-csp-utils.git

!bash rapidsai-csp-utils/colab/rapids-colab.sh stable

import sys, os

dist_package_index =sys.path.index('/usr/local/lib/python3.6/dist-packages')

sys.path = sys.path[:dist_package_index] +['/usr/local/lib/python3.6/site-packages']
+sys.path[dist_package_index:]

sys.path

exec(open('rapidsai-csp-utils/colab/update_modules.py').read(), globals())
```

# Related work:

As an example of a machine learning algorithm 'random forest' classification algorithm are used to compare CPU based version and GPU based version.

Building individual decision trees is where the heavy lifting of Random Forest is done. Individual trees are built using a list of bootstrapped samples. Many algorithms use a top down approach, proceeding with depth-first splits of each node then each newly-created child node. In a GPU context, this can lead to launching an enormous number of CUDA kernels — one per node. These small kernels quickly get queued up as launch time begins to dominate the processing. To remove this bottleneck, cuML uses a breadth-first algorithm, building a full layer of the tree at a time. This makes the runtime of the algorithm scale roughly linearly with depth[5].

As individual Decision Trees are completely independent, building multiple decision trees is embarrassingly parallel. In some cases, the work needed to build a single tree may be too small to fully occupy a large GPU with thousands of CUDA cores. To take advantage of the whole processor, the cuML algorithm can build several trees in parallel on a single GPU. Each tree is built in its own CUDA stream controlled by an OpenMP thread on the CPU[4].

To analyze the performance in a real-world scenario, models are trained on the make_classification() dataset available from sklearn, which has 100 columns and 1M rows. 95% of the total rows for training are randomly picked.

# Code:

| cuML Code | SKlearn code |
|---|---|
| from cuml import RandomForestClassifier as cuRF<br><br># cuml Random Forest params<br>cu_rf_params = {<br>'n_estimators': 100,<br>'max_depth': 8,<br>'n_bins': 15,<br>'n_streams': 8 }<br>cu_rf = cuRF(**cu_rf_params)<br>cu_rf.fit(X_train, y_train)<br><br>print("cuml RF Accuracy Score: "<br>accuracy_score(<br>cu_rf.predict(X_test), y_test)) | from sklearn.ensemble import RandomForestClassifier as sklRF<br>import multiprocessing as mp<br><br>#sklearn Random Forest params<br>skl_rf_params = {<br>n_estimators': 100,<br>'max_depth': 88,<br>'n_jobs': mp.cpu_count() }<br><br>skl_rf = sklRF(**skl_rf_params)<br>skl_rf.fit(X_train, y_train)<br><br>print("sklearn RF Accuracy Score: "<br>accuracy_score(<br>skl_rf.predict(X_test), y_test)) |

# Result:

Speed up of 40x is achieved for 100 tress of depth 8. The difference in accuracy between sklearn(64%) and cuML(63%) was minimal.

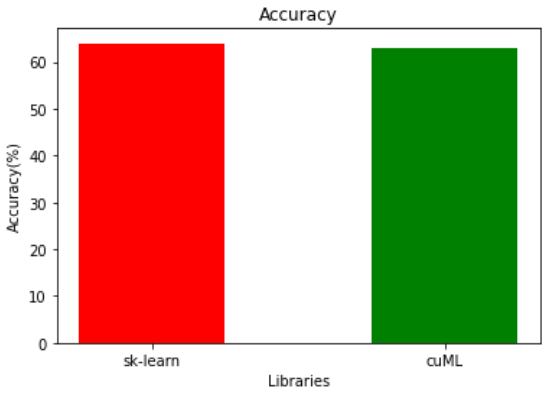



**Fig 1: Accuracy comparison**

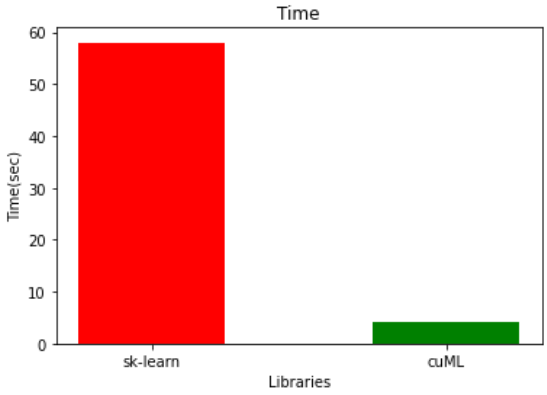



**Fig 2: Time comparison**

# Conclusion:

GPUs are attractive accelerator for machine learning algorithms especially for huge datasets. Equipped with optimized GPU kernels, cuML, high-level Python code can enjoy the same performance benefits as the native GPU implementations. The focus is on optimizing machine learning algorithms on a GPU and developing the convenient drop-in GPU accelerated machine learning libraries in Python.

*References*

[1] https://rapids.ai/start.html
[2] https://docs.rapids.ai/api/cuml/stable/
[3] A. B. L. Larsen. Cudarray: Cuda-based numpy. Technical report, Technical University of Denmark (DTU), 2014.
[4] https://github.com/rapidsai/notebookscontrib/blob/master/bolg_notebooks/randomforest
[5] https://github.com/rapidsai/notebooks/blob/branch0.11/cuml

*Compiled by:*

Mr. Prathamesh V. Barve (121922003) is a student of MTech (Computer Engineering) at College of Engineering, Pune, Maharashtra. He can be reached at : barvepv19.comp@coep.ac.in

Miss. Aarushi S. Phade (121922013) is a student of MTech (Computer Engineering) at College of Engineering, Pune, Maharashtra. She can be reached at : phadeas19.comp@coep.ac.in

Under guidance of:

Mr. Amit D. Joshi (CSI: 01180301) is an Asst. Professor and coordinator of CSI COEP Student branch at College of Engineering, Pune, Maharashtra.
He can be reached at adj.comp@coep.ac.in

# Big Data: As I See

## Sai Varun

***Abstract:***

The systems of data processing have been continuously changing, leading to acquisition of volumes of data. Therefore business processes and objectives need to undergo a change. We need to find alternative ways of processing these data. In the past, we were only concerned with structured data Unstructured data as we see in Whatsapp, Facebook, Twitter etc. were of no consequence.

In this article, we present some of the concepts of big data – the application of which will help us to draw meaningful inferences.

## Introduction:

The systems of data processing have been continuously changing, leading to acquisition of volumes of data. Therefore business processes and objectives need to undergo a change. We need to look for alternative ways of processing these data.

The real challenge to big data lies in both volume and variety. Big data is data that exceeds the processing capabilities of conventional database systems. We have techniques such as OLTP and OLAP in Data Mining and these refer to the mode of processing. A common use of big data processing is to make unstructured data as structured for meaningful processing.

Just as 'Cloud' in cloud computing covers diverse technologies, big data also covers varied technologies.

## The input for Big Data Systems:

The data could come from social networks, web server logs, traffic flow sensors, satellite imagery, broadcast audio streams, banking transactions, financial market data etc, in fact a vast variety of sources. The special feature of Big Data is that it has volume, velocity and variety.

We can employ massively parallel processing architectures such as Green palms and Apache Hadoop based solutions. Facebook, LinkedIn are another sources of information. A MySQL database stores the core data where core computations take place. A common feature in big data is that the source data is diverse and does not fall into neat relational structures. It could be text from social networks, image data or raw data taken from sensor source.

In Big Data processing, we take unstructured data and extract an ordered one fit as a structured input. An example is entity resolution - the process of determining exactly what a name refers to. This process of moving from unstructured source data to a structured one may involve loss of information. We may lose some signals which may be in the form of 'bits' of information. Here, we can use signal processing

techniques to extract signals from back ground noise. A detailed discussion on Big Data could be seen in [1] and [2].

# Big Data: Evolution and some features

Big Data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze. This is different from large data which can be analyzed by traditional statistical techniques.

Big Data is expanding on three fronts at an increasing rate:

1. Data Velocity 2. Data Volume 3. Data Variety

Until 2000's the solution to data explosion was to scale-up to super computers, but the cost effectively increases non-linearly.

Instead of scale-up the Web giants companies like Google & Yahoo started scaling-out. Clusters of commodity hardware computers were used in shelf in data centers. In the past, the solution was to the program, but now it is to send the program to the data which is much easier.

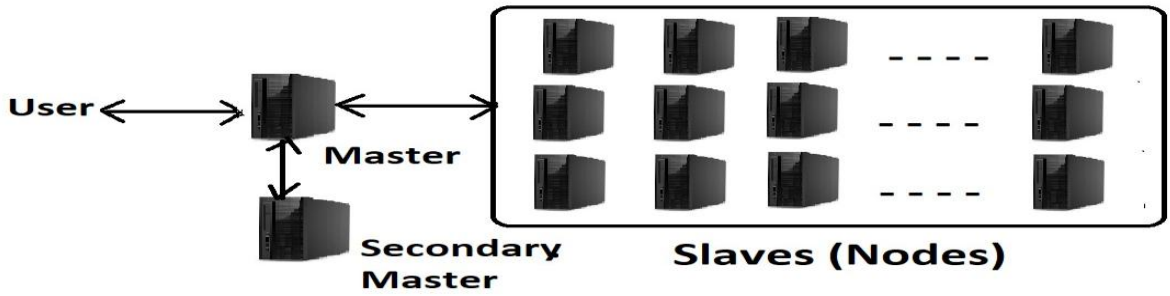



*Fig. User to Node (Adapted from Google)*

# Challenges of Big Data:

Data Storage and quality of Data – The data is growing at a fast pace as the number of companies and organizations are growing. Proper storage of this data has become a challenge but these data may be inconsistent and also may not be accurate.

Lack of big data analysts and quality analysts – There is a huge demand for data scientists and analysts who can understand and analyze this data. But there are not enough trained people and quality analysts and absence of proper tools.

Security and Privacy of Data – Security, and privacy are the biggest risks in Big Data. Proper toolsfor analyzing, storing, managing and use of data from different sources need to be developed. Otherwise data is subject to vulnerable exposure. It increases security and privacy concerns.

# Fields under Big Data:

Big Data is a vast field and there are a number of topics and fields under it on which we can work, for dissertation as well as for research.

*Search Engine Data* – It refers to the data stored in the search engines like Google, Bing and is retrieved from different databases.

*Social Media Data* – It is a collection of data from social media platforms like Facebook, Twitter etc.

*Stock Exchange Data* – It is a data from companies indulged in shares and stock market.

*Black box Data* – Black Box is a component of airplanes, helicopters for voice recording of flight crew and for other metrics.

# Big Data Techniques:

Big Data techniques are required for more detailed analysis, accuracy and concrete decision making. For this, a powerful infrastructure is required to manage and process huge volumes of data. The data can be analyzed with techniques like A/B Testing, Machine Learning, and Natural Language Processing.

The Big Data technologies include business intelligence, cloud computing, and databases.

The visualization of data can be done through the medium of charts and graphs.

Multi-dimensional Big Data can be handled through tensor-based computation. Tensor-based computation makes use of linear relations in the form of scalars and vectors.

# Big Data-Hadoop:

Hadoop is an open-source framework provided to process and store data. Hadoop makes use of simple programming models to process Big Data in a distributed environment across clusters of computers. Hadoop provides storage for a large volume of data along with advanced processing power. It also gives the ability to handle multiple tasks and jobs

# Importance of Hadoop in Big Data:

Hadoop is essential in terms of Big Data. The importance of Hadoop is highlighted as follows:

*a. Processing of huge chunks of data* – With Hadoop, we can process and store huge amount of data mainly the data from social media and IoT(Internet of Things) applications.

*b. Computation power* – The computation power of Hadoop is high as it can process Big Data pretty fast. Hadoop makes use of distributed models for processing of data.

*c. Fault tolerance* – Hadoop provide protection against any form of malware as well as from hardware failure. If a node in the distributed model goes down, then other nodes continue to function. Copies of data are also stored.

*d. Flexibility* – As much data as you require can be stored using Hadoop. There is no requirement of preprocessing the data.

*e. Low Cost* – Hadoop is an open-source framework and free to use. It provides additional hardware to store the large quantities of data.

*f. Scalability* – The system can be grown easily just by adding nodes in the system according to the requirements. Minimal administration is required.

## Challenges/limitations of Hadoop:

No doubt Hadoop is a very good platform for big data solution, still, there are certain challenges in this.

*All problems cannot be solved* – It is not suitable for iteration and interaction tasks. Instead, it is efficient for simple problems for which division into independent units can be made.

*Talent gap* – There is a lack of talented and skilled programmers in the field of Map Reduce in Big Data especially at entry level.

*Security of data* – Another challenge is the security of data. Kerberos authentication protocol has been *developed to provide a solution to data security issues.*

*Lack* of tools – Lack of tools for data cleaning, management, and governance is a concern. Tools for data quality and standardization are also lacking.

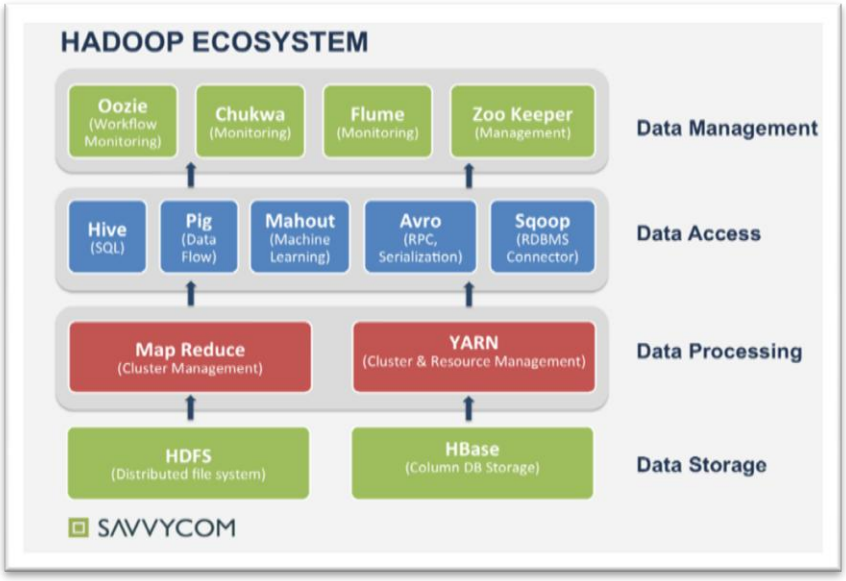
## How the data is stored in Hadoop:

Partition: Divide Data into different parts. Each Part stored in different places.

Replication: Make multiple copies of same data and store in different places.

Hadoop = Partitioning + Replication

## Different Components of Hadoop:




*User to Node Adapted from Google*

# HDFS (Hadoop Distributed File System):

File system for storing Big Datasets in cluster. Uses Partitioning and Replication to store data.
Map Reduce:
Programming language framework for parallel processing of data.
YARN – Yet Another Resource Negotiator
Resource Manager and a Scheduler.
Allows other Applications other than Map Reduce Applications. (Eg Spark)
PIG (Developed by Yahoo!!!):
Provides data flow language (PigLatin).
Can write scripts to do different transformation (LOAD, JOIN, FILTER, GROUP, STORE)
HIVE(Developed by Facebook):
One of oldest "SQL on Hadoop" technology.
Allows to use Hadoop as relational database with Schema defined at run time.
PIG and Hive Translates commands into Map Reduce jobs.
FLUME:
Designed to store Streaming data like weblogs, application logs from distributed systems into HDFS.
SQOOP:
Tool for transferring data from Relational databases to HDFS and back.
SPARK:
Tool for interactive and multi-step analysis on Hadoop Cluster.

Hadoop takes time to store output of each step into disk for every Map Reduce Steps. Spark solves this problem by keeping data in-memory as much as possible, thereby performing several Map Reduce steps in rapid sequence.

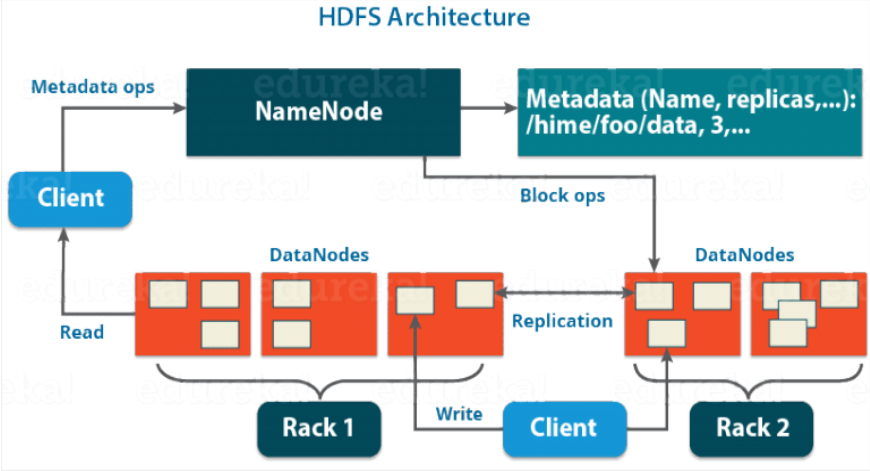



*Fig. User to Node Adapted from Google*

# Conclusions:

Big Data is different from large data. Big Data provides lots of useful information. The availability of Big Data, low-cost commodity hardware, and new information management and analytic software have produces a unique moment in the history of Data Analysis. The convergence of these trends means that we have the capabilities required to analyze astonishing data sets quickly and cost-effectively for the first time in history.

***References:***

[1] Bernard Marr,2016: "Big Data in Practice", John Wiley & Sons U.K, pp. 323.
[2] Tom White,2015: "Hadoop- The Definitive Guide", O Reilly Media, California, pp. 235.
[3] https://www.sas.com/en_us/insights/big-data/what-is-big-data.html
[4] https://www.ntnu.no/iie/fag/big/lessons/lesson2.pdf

***Compiled by :***

Mr. A. Sai Varun pursing IVth year-II Semester in ICFAI Foundation for Higher Education (IFHE-Hyderabad).

Under the guidance of

Dr. D. D. Sarma, Chief Scientist, CSIR-NGRI(R) Hyderabad.

# Big Data Analytics Using Deep Learning

Poornima .S ,Chandhana M and R.Geetha

Abstract:
The collection of big data has been enabled due to the recent development of sensor networks and communication technologies. Limitations in processing huge amount of data were overcome by using big data. Big data analytics requires emerging sophisticated algorithms. Big data offers great opportunities in the fields of industrial control, e-commerce, smart medical on the other hand it also faces so many issues on data mining and information processing due its characteristics of large volume, large velocity, large veracity, large variety. Deep learning plays a vital role in big data analytic solutions. The emerging researches of deep learning models for big data feature learning are discussed in this paper.

## 1.Introduction:

The cyber-physical-social systems, together with sensor networks and communication technologies enhanced the collection of big data [1,2]. Big data is actually called as 4V's model, defined by its four characteristics such as large volume, large velocity, large variety, large veracity [3,4]. The main characteristic feature of big data is its large volume which implies in the huge data amount. Example: Google processes up to 20,000 TB data per day.

## 2.Big Data Analytics:

In this digital world, huge amount of data is generated from various platforms which gave rise to Big data .It handles large amount of raw data. Nowadays, big data is employed in all walks of life such as public administration, scientific research, healthcare, business, national security, internet of things (IoT), stock exchanges, commercial recommendations, etc. Online social network such as Google, Amazon, Facebook, Twitter, etc. allows billions of customers to share multimedia like text, video, audio, images which are increasing day by day. The main challenge lies not only in collecting such huge size of data but to manage it properly and to utilize it in efficient manner to make decision or prediction. Several tools are designed and developed to deal with big data problems.

## 3.Challenges In Big Data Analytics:

In addition to huge amount of data, big data also suffers from some other issues usually known as four V's, Volume, Velocity, Variety and Veracity. Big data deals with unstructured and un-labeled large amount of data. The main concept of big data is to deal with complex raw, dissimilar input data, that consist of unsupervised data of different size and might have a small amount of supervised data. Variety of data representation leads to various challenges in big data to extract useful and structured data from

unstructured data. Data is produced and stored makes data velocity equally important as volume and variety. It is must for Big data tools to timely process the data and to reform it in to effective information .Loss of data will happen if the data is not processed on time. Organizations such as Facebook and Twitter have designed products to deal with data streaming. Veracity is nothing but accuracy, trust and validity of results obtained from analysis. Data from different sources and accuracy becomes a massive challenge in Big Data Analytics.

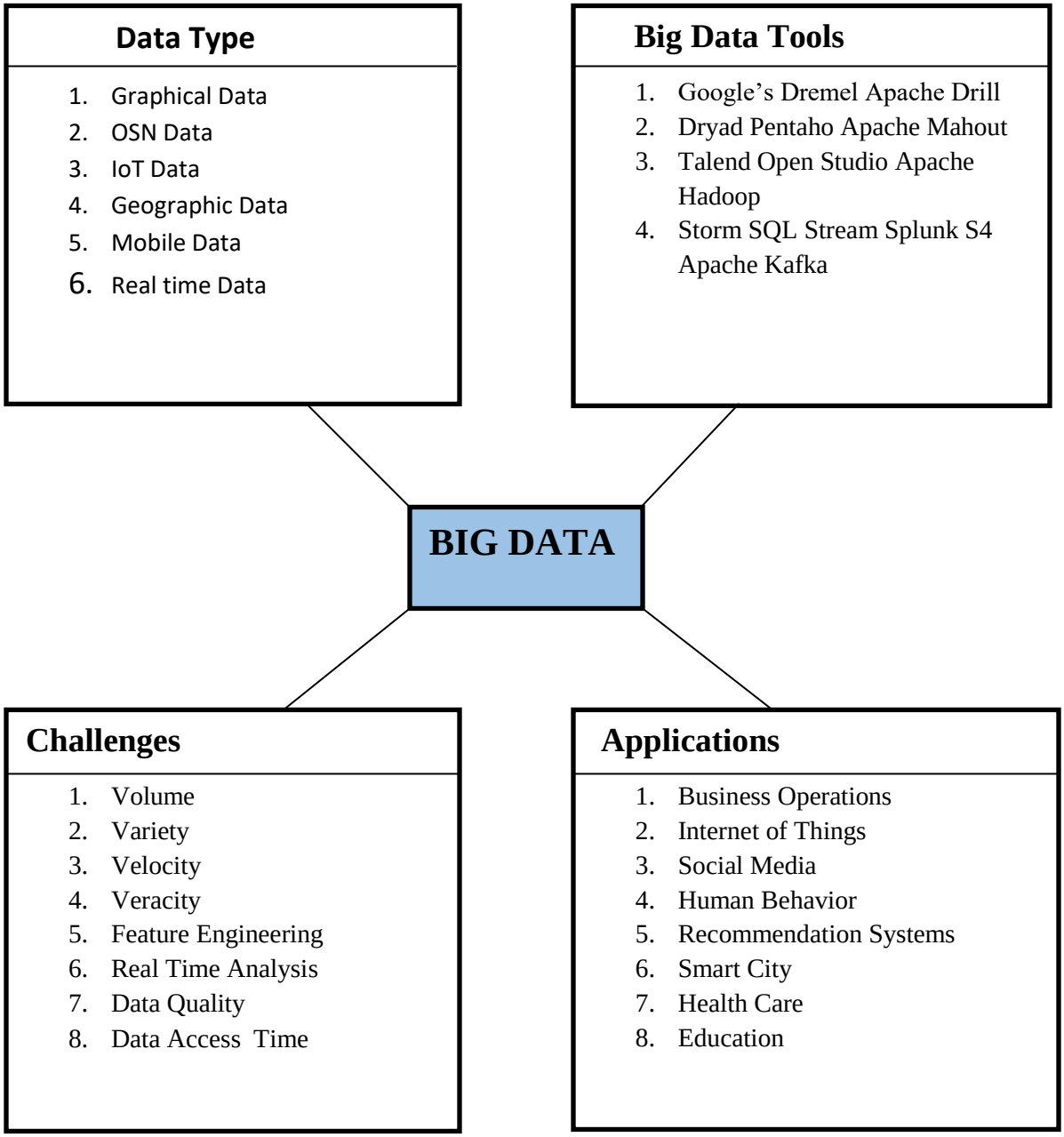



*Figure: Big Data Tools, Applications and Challenges*

### 3.1. Data Type

Number of distinct datasets are developed due to emergence of big data, most of the data sets are domain specific having distinct representation of data, density, distribution and size of the data. To obtain knowledge from these datasets is of high significance in big data research that makes it distinguish from conventional data mining, few of the data types listed below are dealt with big data.

### 3.2. Big Data Tools

To deal with big data, we need to have specialized tools that can handle the big complex form of data in an efficient and effective way. As traditional tools cannot handle the analytics of bigdata, the big data tools are mainly classified into batch processing, stream processing and interactive analysis.

### 3.3. Big Data Applications

The amount of data available in distinct field of life has increased the essential of bigdata to extract useful information for decision making. The traditional methods are not feasible for information extraction from large volume of data. Therefore deep learning plays a significant role in the bigdata analytics and is widely used in different areas such as business operation, smart city, health care, internet of things, social media, recommendation system, human behavior and many more. For instance, to handle the amount of data generated from social media and mobile phones, deep learning techniques are used to extract information related to the user activities

# 4.Deep Learning In Big Data Analytics:

The concept of deep learning is to dig the large volume of data to automatically identify the patterns and to obtain features from complex unsupervised of data without the involvement of the human, which makes it as a very essential tool for big data[5]. In today's fast growing world there are excessive amount of data with distinct formats and sizes are dealt with, different machine learning techniques along with the computationally strong machines are required to deal with volume and variety of data. Deep learning uses supervised/unsupervised techniques to automatically learn and extract data representations. It can be used to solve the big data problems in a more efficient way, which is impossible in traditional method. Here we discuss two different architectures of deep learning with respective to use in big data application.

### 4.1 Deep Belief Networks and Big Data

The deep belief network(DBN) has the capability to learn feature representations from labeled and unlabeled data. It construct model by using supervised and unsupervised techniques. a DBN[6] architecture consist of input layer, hidden layers and output layers.

The first deep learning successful model is DBN. Different from stacked auto-encoder, DBN is stacked by several restricted Boltzmann machines.

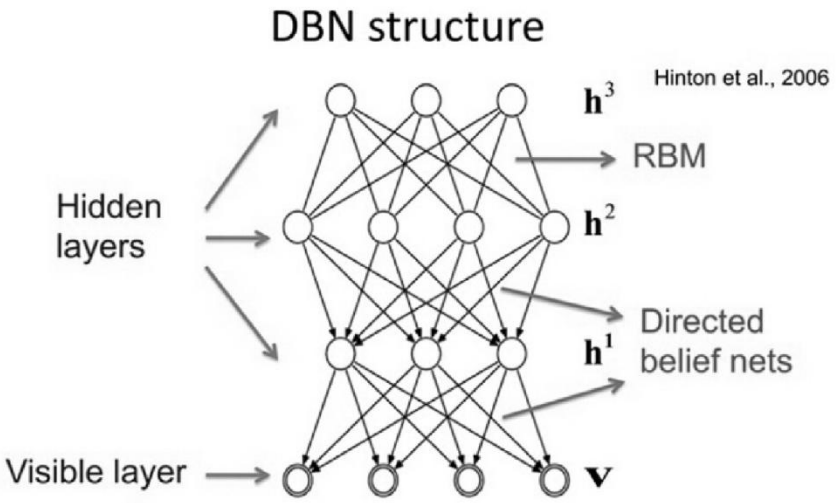



*Figure : Deep Belief Network*

## 4.2. Convolutional Neural Networks and Big Data:

The convolutional neural network (CNN) [7]has many hierarchical layers, consisting of feature maps layers and classification layers. Typically CNN starts with convolutional layer that accepts data from input layer. The convolutional layer is responsible for convolution operations having few filter maps of same size.

# 5. Conclusion:

The explosion of big data offers an enough training objects, which helps to improve the performance of deep learning. Furthermore high performance computing devices and architectures such as graphic processing units and CPU clusters cluster enable the training of large scale deep learning models for bigdata feature learning. Today deep learning models enjoy the success with a great many parameters, typically millions of parameters, together with the large number of training objects. While bigdata bring enough training objects, it also poses some challenges of deep learning. Therefore in the past few years, many deep learning models have been developed for bigdata learning. In big data typically defined as four v's model: volume, variety, velocity and veracity, which implies excessive amount of data, various types of data, real time data and low quality data respectively. Therefore we summarized deep learning models for big data learning from four aspects accordingly.

***REFERENCES***

[1] Chen X-W , Lin X . Big data deep learning: challenges and perspectives. IEEE Access 2014;2:514–25 .
[2] Zhou L , Pan S , Wang J , Vasilakos AV . Machine Learning on Big Data: Opportunities and Challenges. Neurocomputing 2017 .
[3] Najafabadi MM , Villanustre F , Khoshgoftaar TM , Seliya N , Wald R , Muharemagic E . Deep learning applications and challenges in big data analytics. J Big Data 2015;2:1 .
[4] Khumoyun A , Cui Y , Hanku L . Spark based distributed Deep Learning framework for Big Data applications. In: Information Science and Communications Technologies (ICISCT), International Conference on; 2016. p. 1–5 .
[5] Bengio Y , Courville A , Vincent P . Representation learning: A review and new perspectives. IEEE Trans Pattern Analysis Mach Intell 2013;35:1798–828 .
[6] G.E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural Comput. 18(7) 1527−1554.
[7] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, F. Li, Large-scale video classification with convolutional neural networks, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, (2014), pp. 1725−1732. IEEE

*Compiled by :*

Ms. S. Poornima (CSI 01442267) is a student of BE IV year (CSE) at S.A. Engineering College, Chennai, Tamil Nadu. She can be reached at poornimasrinivasan22@gmail.com

Ms. M.Chandhana (CSI 01442244) is a student of BE IV year (CSE) at S.A. Engineering College, Chennai, Tamil Nadu. She can be reached at chandhanamahesh27@gmail.com

Under the guidance of :

Dr. R. Geetha (CSI N1259352) is a Professor and Head of the Department, CSE, S.A. Engineering College, Chennai, Tamil Nadu. She can be reached at geetha@saec.ac.in

# An Efficient technique to detect Soil and Leaf Nutrients using Image Processing and Internet of Things

Rajiv R. Bhandari, Vaibhav Shelar,
Meghna Brahmankar, Jignesh Thosare

***Abstract:***

There are many problems faced by farmers in India for yielding good crops. Nowadays with the help of advanced technologies like IOT and Image processing, we can solve these problems. Nutrients play an important role in the quality as well as quantity of good crops and deficiency of nutrients causes a reduction in the quality of crops and productivity and sometimes even causes destruction of plants. There are total 13 essential nutrients required for plants. Among these 13 nutrients, NPK i.e. Nitrogen, Phosphorus, and Potassium are the 3 most essential and important nutrients. Nutrient deficiency can be detected by examining soil and leaf. So, the proposed system detects deficiency of NPK (Nitrogen, Phosphorus and Potassium) in soil as well as leaf. Farmers generally choose an approximate amount of fertilizers for soil. The first module in the system detects NPK deficiency in the soil through the method of optical transducer using IOT and the second module is a diagnostic system using digital image processing which diagnoses the deficiency symptoms much earlier than human eyes could recognize using image processing. It can detect the amount of nutrients and is capable of deciding which specific nutrient is absent or is in less quantity. The system uses a Raspberry Pi. The result of these tests will suggest farmers that which fertilizers (nutrients) should be provided to plants. Hence it will increase the quality and productivity of crops and will reduce the economic loss.

## 1. Introduction:

In the world of advanced technology now various types of technology have been created to facilitate the daily activities of man. As well as in agricultural technology, a variety of tools that have been created to help farmers make their agricultural activities and get a good crop [1]. Soil fertility is an important factor to measure the quality of the soil as it indicates the extent to which it can support plant life. The fertility of soil is measured by the amount of macro and micronutrients, water, pH, etc. Soil nutrients are depleted after every harvest and hence must be replenished. To maintain nutrient levels in the soil in case of deficiency, fertilizers are added to soil. Most of the farmers choose to approximate the amount of fertilizers and add them manually. However, the addition of fertilizers in the right amount is a matter of great importance as excess or insufficient addition can harm the plant life and reduce the yield. The nutrition problem in farming is very crucial. So there is a need for an efficient system that will detect the nutrients present as well as the nutrient deficiency in crops or plants. Plants extract nutrients that they need for their growth from the soil [2].

# 2. Literature Survey:

### 2.1 Analysis of Nutritional Deficiency in Citrus Species Tree Leaf using Image Processing:

Citrus trees are the nutrition food for humans as well as animals. However, due to the uncertain climatic conditions, it will prone to different pathological disorders because of the nutritional deficiency. In Vidarbha regions, citrus suffer from certain deficiencies of essential elements, in which plants gain from the soil. The segmentation of disease symptoms in citrus leaf images can be a valuable aid for the detection of nutritional deficiencies and disorders. In this research, different digital image segmentation techniques have been employed which analyses the regions of the citrus leaf caused by some diseases such as spots and wavy structure. This paper investigates the abnormalities in citrus leaf caused by the diseases by the segmentation methodologies. The nutritional deficiency of the citrus tree is directly reflected on its plants. If any temporal part of the symptom is disconnected then, it can be segmented to its original part using the clustering technique. The disease spots are identified by the clustering where the wav disorders are segmented by the Kirsch Operator. This proposed system analyzes the disorder of the citrus tree by analyzing its leaf using the segmentation technique with the integrated use of clustering feature and the Kirsch operator. [4]

### 2.2 Automatic Soil Nutrient Detection and Fertilizer Dispensary System

The fertility of soil is measured by the amount of macro and micronutrients, water, pH etc. Soil nutrients are depleted after every harvest and hence must be replenished. To maintain nutrient levels in the soil in case of deficiency, fertilizers are added to soil. Most of the farmers choose to approximate the amount of fertilizers and add them manually. Use of modern trends and technology promises to provide a solution to the above problem. Though automated techniques for seeding, weeding, harvesting the crops etc. have been proposed and implemented, none of the techniques target at maintaining soil fertility. The proposed research aims are storing the levels of Nitrogen, phosphorous, potassium in the soil by the measuring the amount of nutrients present. The presence of nutrients is determined by chemical processes and quantified using sensors. An automated system has been developed for the controlled addition of fertilizers in order to avoid excess/ deficient fertilizers in the soil0. [5]

# 3. Proposed System:

3.1 The proposed system consists of two separate modules as follows:

1) Soil nutrient deficiency detection

2) Leaf nutrient deficiency detection

In the first module, soil nutrients and soil contents will be detected. The results will show which nutrient is absent or in very less quantity whereas, in the second module leaf nutrients will be detected with the help of image processing algorithms, we have proposed a model to detect the type of deficiencies in the leaves. The color and texture features are used to recognize and classify the deficiencies. The combinations of features prove to be very effective in deficiency detection. The combination of these two modules will give the exact deficiency results which will be useful in agriculture or farming, Soil

testing, Nutrient level detection, Detection of NPK levels in soil and leaf, Detection of nutrient deficiency.

There are many different techniques available for the nutrition detection in crops and plants. But there is no system or a device which will detect the nutrient deficiency in soil as well as leaf. The systems available work either for soil or for a leaf. The proposed system is an integration of both the modules namely soil nutrient deficiency detection and leaf nutrient deficiency detection. There is a single device in proposed system which will detect the deficiency in both soil and leaf. The soil module uses IOT and leaf module uses Image Processing. The results of both the modules will be combined to get exact and correct result.

## Soil nutrient deficiency detection

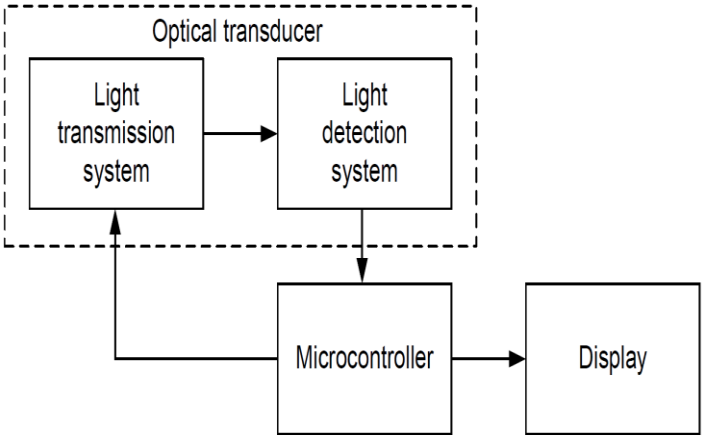
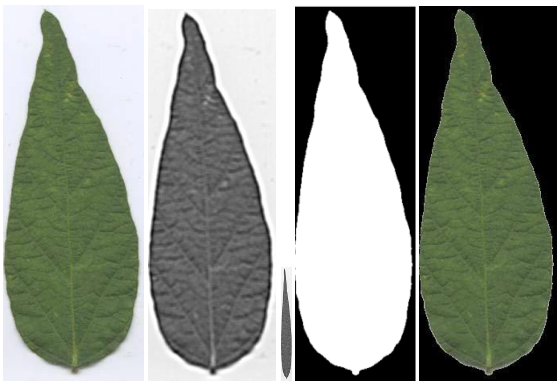



*Fig.1 Block diagram of integrated optical transducer with microcontroller*

In the optical transducer method, LED's are utilized and used as the source of direct light and by absorbing the light, soil interacts. The photodiode has the capability of converting the light into current. So, photodiode detects the remaining light as well as it evaluates the absorption rate. The output from the photodiode is manipulated using Raspberry pi as a result the output current is converted and displayed as output voltages.

| Nutrient | Absorption wavelength (nm) | LED type | Wavelength (nm) |
|---|---|---|---|
| Nitrogen (N) | 438-490 | LED 1 | 460-485 |
| Phosphorus (P) | 528-579 | LED 2 | 500-574 |
| Potassium (K) | 605-650 | LED 3 | 635-660 |

## *Leaf nutrient deficiency detection*



In this module, the pi camera captures the image. Further the image gets processed and the NPK values get displayed on the LCD screen.

# 4 .Algorithm :

The camera captures the image of a leaf placed on white paper.

Calculate RGB values of a leaf

Calculate average of RGB values and threshold value is set to 170

Edge detection :

if avg >170 :

img.putpixel ((x,y), (255, 255, 255))

else :

img.putpixel ((x,y), (R, G, B))

Convert image into histogram. Open out.csv ( out.csv file contains training data and the npk values of leaves with deficiency of NPK)

The histogram of testing leaf is compared with the histograms in database i.e. the training data.

Find the exact or nearly similar histogram and it's values.

Get the NPK values of histogram

Display NPK values on LCD screen

# 5. Data Flow Diagram:

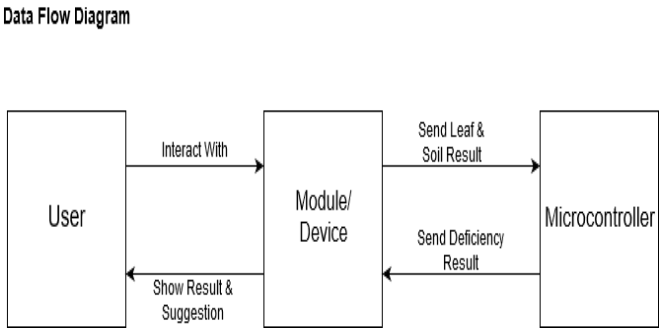



*Fig. 2 Data Flow Diagram*

# 6. System Architecture:

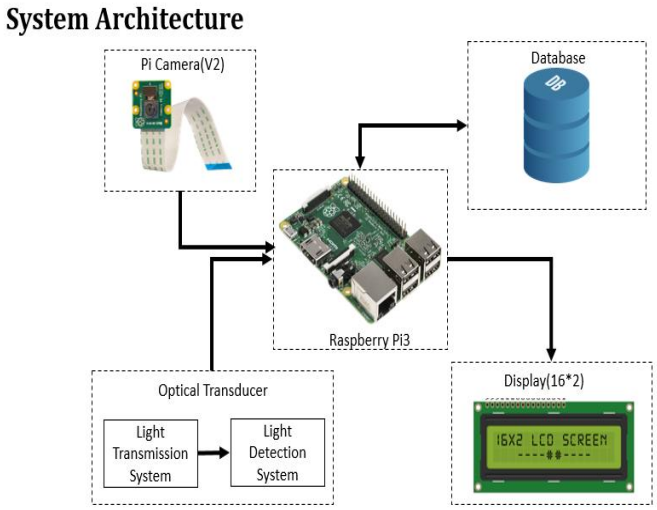
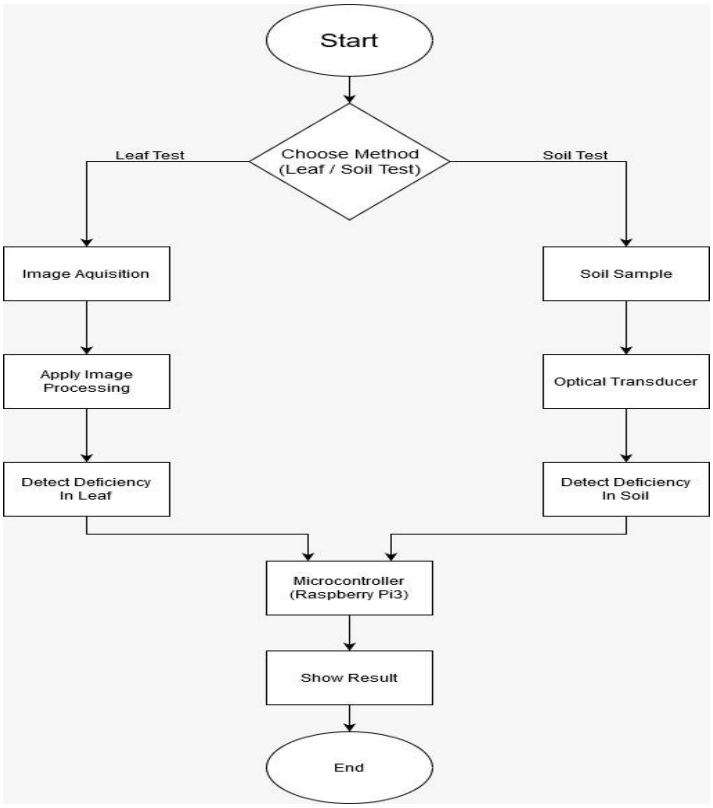



*Fig. 3 System Architecture*

## Overall System Flow




# 7. Hardware Requirements:

1) Raspberry Pi3

2) Pi Camera

3) Led Light (RGB)

4) Photodiode Sensor

5) LED Display (16*2)

# 8. Results:

### 8.1 For Soil :

| Nutrient | Low (V) | Medium (V) | High(V) |
|---|---|---|---|
| Nitrogen | 3.3<x<3.7 | 3.8<x<4.1 | x>4.2 |

| | | | |
|---|---|---|---|
| Phosphorus | 2.5<x<2.7 | 2.8<x<3.3 | x>3.4 |
| Potassium | 1.5<x<2.1 | 2.2<x<2.6 | x>2.7 |

TABLE 1 shows the threshold values for NPK soils in three voltage levels; High, Medium and Low. The value x indicates the voltage absorption for each nutrient.

### 8.2 For Leaf :

After the image is captured by camera, it gets converted into histogram and compared with the database and available values. The following table shows the deficiency values.

| Nitrogen (N) | Phosphorus (P) | Potassium (K) | Deficiency |
|---|---|---|---|
| 85 | 42 | 32 | K |
| 80 | 18 | 60 | P |
| 20 | 69 | 96 | N |

*Table 2. Leaf Nutrient Deficiency Values*

If the value generated is less than the threshold value which is set to 45 then the deficiency of that particular nutrient is detected as shown in Table 2.

# 9. Conclusions:

Nutrients are one of the very important factors that affect the quality and quantity of crops. Nitrogen, Phosphorus and Potassium i.e. NPK are most essential nutrients among all the nutrients. Using IOT, we can detect nutrient deficiency in soil and using Image Processing, we can detect leaf nutrient deficiency. The test results of soil and leaf will detect which nutrient is absent or in less quantity. Farmers will be able to provide appropriate fertilizers to crops which will result in better productivity.

### *REFERENCES*

[1] Marianah Masrie, Mohamad Syamim Aizuddin Rosman,Rosidah Sam and Zuriati Janin, “Detection of Nitrogen, Phosphorus and Potassium (NPK ) nutrients of soil using Optical Transduser” Proc. of the 4th IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA) 28-30 November 2017, Putrajaya, Malaysia

[2] Amrutha A, Lekha R, A Sreedevi, “Automatic Soil Nutrient Detection and Fertilizer Dispensary System”, 2016 International Conference on Robotics Current Trends and Future Challenges (RCTFC)

[3] Akshay Badhe, Sandeep Kharadkar, Rushikesh Ware, Pratik Kamble, Prof. Shilpa Chavan, "IOT Based Smart Agriculture And Soil Nutrient Detection System" International Journal on Future Revolution in Computer Science & Communication Engineering ISSN: 2454-4248 Volume: 4 Issue: 4 774 – 777

[4] Muhammad Asraf Hairuddin, Nooritawati Md Tahir and Shah Rizam Shah Baki, "Overview of Image Processing Approach for Nutrient Deficiencies Detection in Elaeis Guineensis", 2011 IEEE International Conference on System Engineering and Technology (ICSET)

[5] Lili Ma, Yuehua Chen, Junlong Fang, Shuipeng Gong, "Color Analysis of Leaf Images of Deficiencies and Excess Nitrogen Content in Soybean Leaves", 2010 IEEE

***Compiled by:***

Ms. Meghana Bramhankar ( 01449073 ) is a student of BE final year (Computer Engineering) at SNJB's College Of Engineering, Chandwad ( Nashik ) Maharashtra. She can be reached at bramhankarmeghana519@gmail.com

Mr. Jignesh Thosare ( 01449081 ) is a student of BE final year (Computer Engineering) at SNJB's College Of Engineering, Chandwad (Nashik) Maharashtra. He can be reached at thosarejignesh@gmail.com

Mr. Vaibhav Shelar ( 01449082 ) is a student of BE final year (Computer Engineering) at SNJB's College Of Engineering, Chandwad (Nashik) Maharashtra. He can be reached at vaibhavshelar3005@gmail.com

***Guided By:***

Rajiv R. Bhandari (F8001720), working as Assistant Professor in Department of Computer Engineering at SNJB's Late Sau. K. B. Jain College of Engineering, His area of Research is Wireless Sensor Network, Internet of Things.

# Cyber Crime And Cyber Security In India

P. Kiruthika, M. Rajakumar

## Introduction

Huge quantity of data is present in the world. But the security is not ensured. The two important terms involved are cyber-crime and cyber security.

***Cybercrime:***

Cybercrime is a criminal activity that is networked computers. The two main intentions for the cyber-crime is to obtain any kind of profit or damage any devices.

***Cyber security***:

Cyber security is a task which is performed to protect our data. Cyber security is an upcoming trend that ensures the data security.

Many technologies are there in this world for data storage and data analysis. Some of them are Cloud computing, Data analytics, Big data etc. The need for data privacy is still high. To achieve this, we go for cyber security, which enhances the data security.

***Cyber Crime***

Various types of cybercrime are:

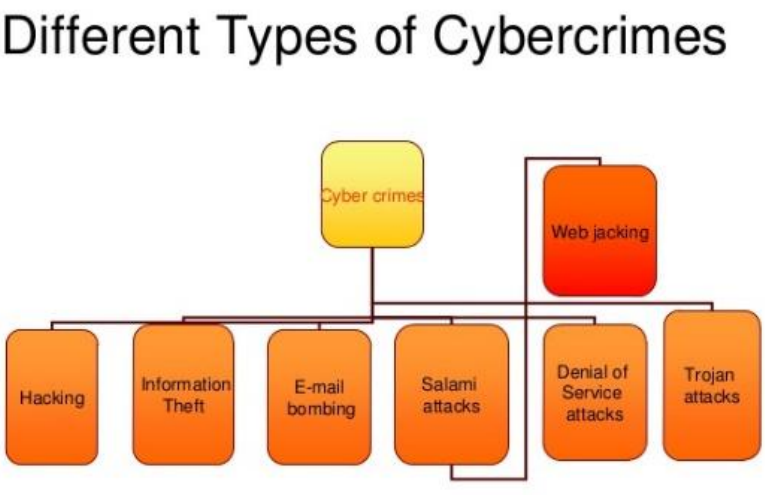



*Fig. 1: Types of Cybercrime*

*Identity Theft:*

This involves the access to personal information of the people by Cybercriminals by the techniques phishing, hacking, etc, …

*Botnets:*

This method is hacking of networked computer externally by remote hackers.

*Phishing:*

Phishing is sending of unwanted malicious messages or URL to client's computers.

*DDOS Attacks:*

DDOS Attacks are infecting the large network by traffic sources.

# Cyber Attacks In India:

A report says that the four major metropolitan cities that were most affected due to cyber-attacks in 2019 are Kolkata, Bangalore, Delhi and Mumbai, while reporting the states the 3 major states that were affected are West Bengal, Delhi and Maharashtra. Most of the cyber-attacks are focussing are Windows and Android. In the past year, approximately 973 million instances of attacks were registered which means that 1852 Windows machines are getting affected every minute in one year. The main cause of these attacks is Trojans. Trojans are available as software copies in illegitimate format. The second cause is worms and the third is infectors. Ransomware is also a kind of Cyber-attack which is the minimal affected attack in India 14 minutes per PC were getting attacked. The number of Crypto-theft attacks registered is 11 in the country for one minute. In Android the tool used for attacking is through applications. At least, everyone was getting infected by malicious content applications for every 3 minutes. Letting every application without proper security check in the Google Play store is taken as a loophole for the cyber criminals.

Other than this, many kinds of Cybercrimes present. They are, Cyber stalking, Social Engineering, Potentially unwanted programs (PUPs), Online Scams, Child Pornography etc, …


*Fig. 2: Cyber-attacks in India*

### *Cyber Security*

### *Techniques Used for Cyber security*

Various techniques of cyber security are:

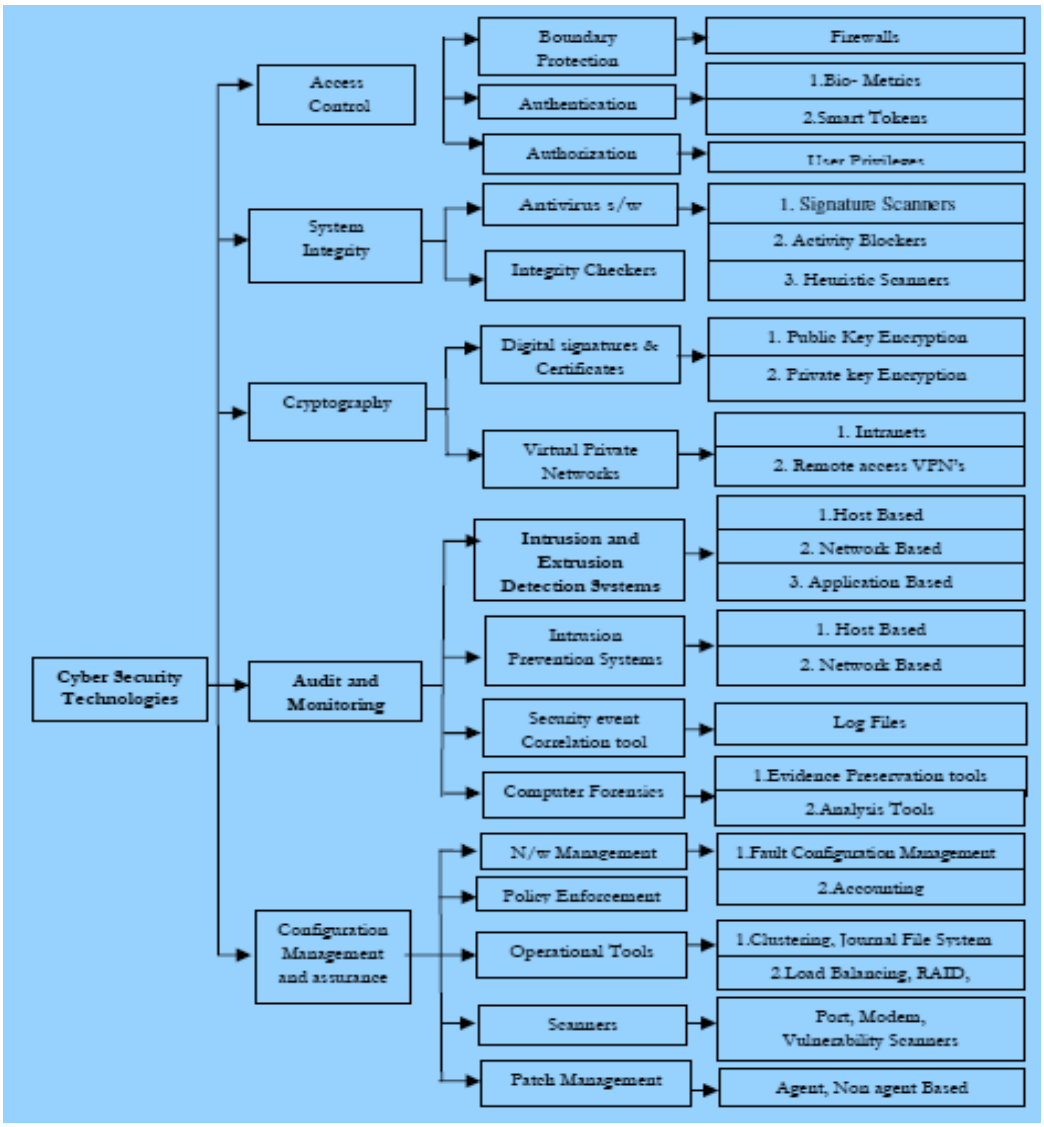



*Fig. 3: Techniques used in Cybersecurity*

Some of the major techniques are described below.

*Authentication:*

Authentication refers to preventing the data from Eavesdropping. The important information such as password while transmitting through internet is intercepted by the middleman. This can be avoided by authentication.

*Encryption:*

Encryption comes under cryptographic technique in which the conversion of ciphered text is done. Initial, plaintext is converted as ciphered trees.

*Anti-Virus:*

Antivirus is created to protect our computer from a threat called virus. The antivirus is installed in a computer to scan the of other viruses in the computer.

*Firewall:*

Firewall acts as a wall and stops the unauthorized access to a computer over the internet.

# Cyber Security In India

According to the Data security council of India. India is the second cyber affected country in the world between 2016 and 2018. Cyber security works towards 3 major principles. They are as follows,

Confidentiality

Integrity

Availability

Confidentiality refers to sharing of sensitive information only with specific persons.

Integrity refers to protecting the sensitive information from being altered.

Availability refers to the ability to access the sensitive information to those who rely on it.

The top 3 states are Maharashtra, Uttar Pradesh and Karnataka. In 5 years, a total of about 5900 cases were registered in Maharashtra. Secondly, Uttar Pradesh with about 5000 cases and then Karnataka holds 3rd position with about 3500 cases.

The government took the following measures and announced in Lok Sabha. They are:

The ministry of Home affairs ordered to take necessary steps in improving trained manpower count for detection and investigation and to improve technical facilities.

The police and officials of judiciary were initiated for a training session in handling tools and detection of cyberattacks.

Central Bureau of Investigation arranged and initiated to set up Cyber Forensics training laboratory for providing advanced training to police officials in states such as Tripura, Assam, Jammu and Kashmir and so on.

In association with Data security council of India and NASSCOM Cyber Forensics laboratory were setup in some major states such as Pune, Mumbai, Bangalore and Kolkata to provide awareness.

An online portal to register Cyber-crime queries and complaints is planned Ministry of Home Affairs.

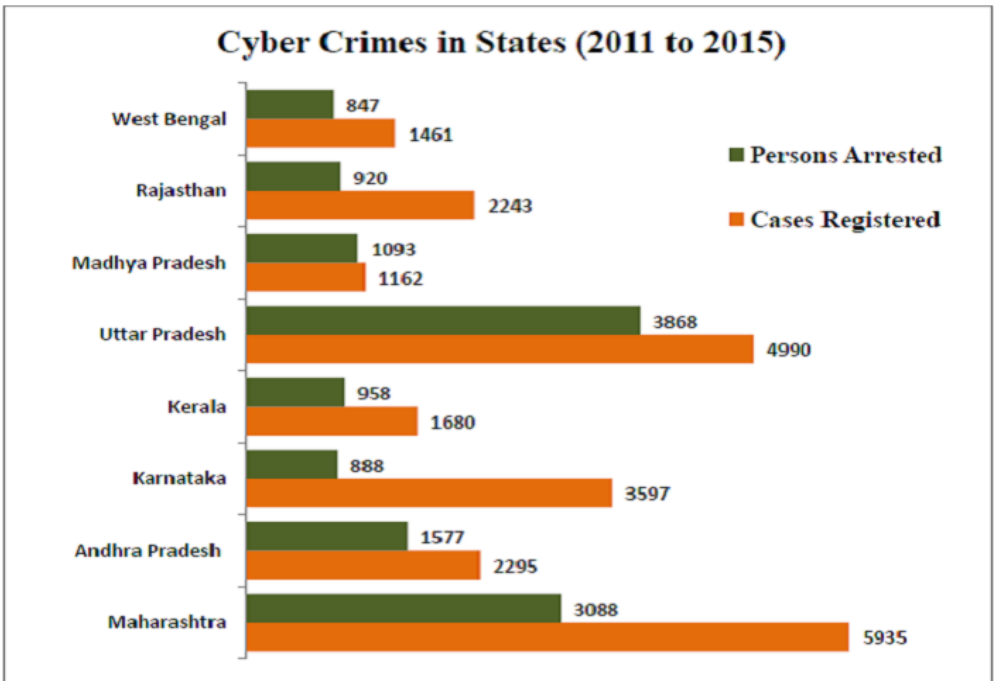



*Fig. 5: Cybercrimes rate in Indian states*

# Phishing Analysis

Among the various types of cybercrimes, one of the most common cyber-attack is Phishing. And in this section two phishing incident happened in real time are discussed.

*1: Website Phishing Analysis*

This is an attempt made to attack the users by sending emails. By believing the e-mail as true, the users would submit their banking credentials such as username and password.

In this, 120 people were targeted. The authorizations of officials working in the HSBC bank were taken. The customers of HSBC bank were sent a fake e-mail with their own user name and password. The e-mail structured as follows:

The cyber wing of HSBC bank noted that there is a third-party access to your account. To confirm that it is you, kindly login and reactivate your account.

Table 1: Phishing through E-mail

| Answer to phishing mail | Exact target of number of employees |
|---|---|
| Positive interaction (Income Tax Department) | 8 |
| Positive interaction (Other Departments) | 44 |

| | |
|---|---|
| Negative interaction (With incorrect info) | 28 |
| Negative interaction (With no Response | 40 |

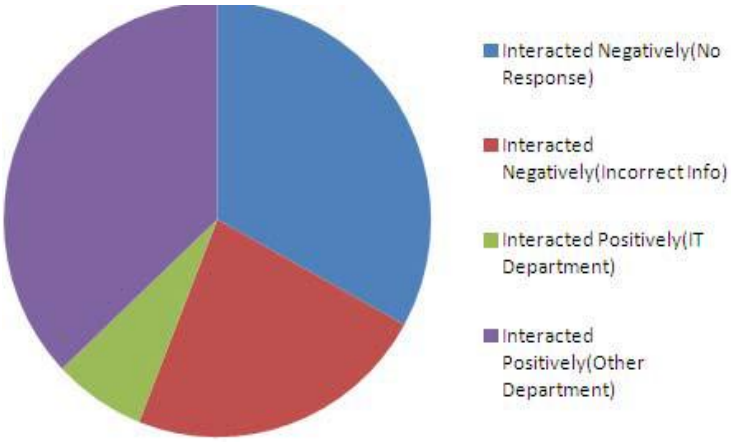



*Fig. 6: Phishing through E-mail*

The results were shocking to us. 44% gave us a positive interaction. They logged in and shared their credentials. 8% of employees from the income tax department revealed their details to us. The Income tax officials are more expected to know about the cybercrime attacks but they were not and they shared their credentials. 28% of people gave negative information which is sharing of incorrect information. They seem to be more alert in cyber-crime attacks. 40% of employees did not response to e-mail.

This shows the Phishing is more dangerous and the people should be more careful while sharing any credentials.

*2: Phone Phishing Analysis*

This analysis involves 50 employees. Those 50 employees were tempted by their female colleagues and attained the user credentials such as username and password. This attack executed as follows:

To show this as authentic phone call, the user was contacted for about 3-4 times. And they were cheated with fake reasons such as account security, connectivity issues, accessibility etc, …

Table 2: Phishing Through Mobile Phones

| Phone Phishing Experiment Response | Exact Number Of Employees |
|---|---|
| Sharing full credentials | 16 |

| | |
|---|---|
| Sharing only user name | 8 |
| Refused for any information | 26 |
| Sub Total | 50 |

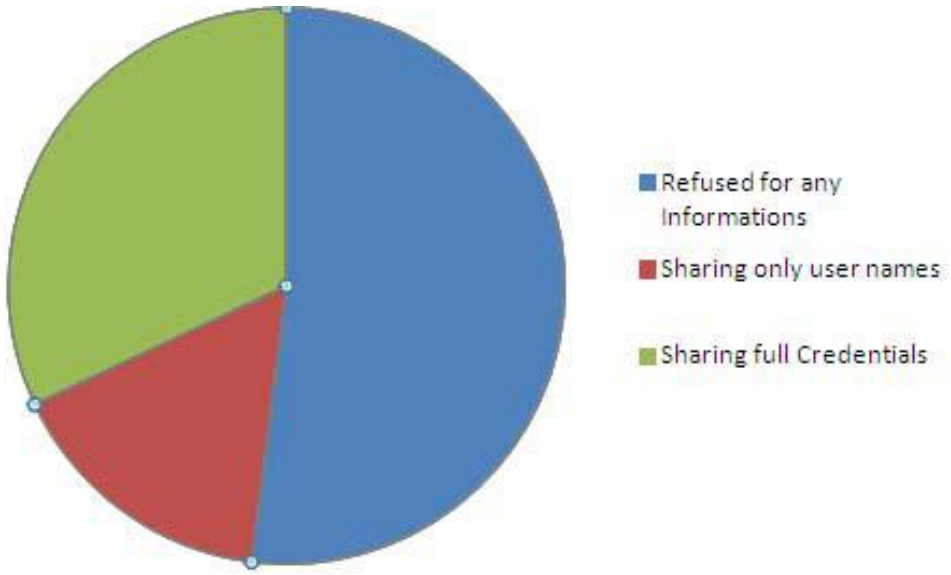



*Fig. 7: Phishing Through Mobile Phones*

The result is to shocking because the bank employees itself are not aware of this kind of cyberattack. The result is 16% of employees shared their full banking credential such as username and password. 8% of the employees shared their username alone and 26% employees did not provide any kind of information.

Thus, this is also an incident that shows the awareness among people towards cybercrime.

# Conclusion

Let of Cybercrime activities are threatening us in day-to-day life. Many Cyber security measures are being adopted to control the cybercrime activities. Further researches in the field of cyber security would ensure the data privacy. Thus, privacy which is considered as an impossible thing in this world, is made possible by the technology 'named' Cybersecurity.

***REFERENCES***

[1] https://searchsecuriyu.target.com/defintion/cybercrime
[2] https://digitalguardian.com/blog/what-cyber-security
[3] https://www.news18.com/news/tech/1852-cyber-attacks-hit-indua-each-minute-last-year-mumbai delhi-most-affected-2295963.html
[4] https://inc42.com/buzz/cyber-attacks-india/
[5] https://www.transunion.com/blog/identity-protection/why-is-cyber-security-important
[6] https://www.pandasecurity.com/mediacenter/panda-security/types-of-cybercrime/
[7] http://www.crossdomainsolutions.com/cyber-security/tools-techniques/
[8] https://factly.in/cyber-crimes-in-india-which-state-tops-the-chart/#prettyPhoto
[9] Vinit Kumar Gunjan, Shrada Avdhanam, Amit Kumar "A Survey of Cybercrime in India", 2013.

***Compiled by:***

Ms P. Kiruthika [01491613] is a UG student of SRM Valliammai Engineering College Pursuing degree in B. TECH INFORMATION TECHNOLOGY. Her area of interests includes Cyber security and Blockchain.

Mr. M. Rajakumar (01491535) is a UG student of SRM Valliammai Engineering College in Computer Science and Engineering department. His areas of interests include Cyber security and IoT.

Under the Guidance of:

Dr. M. Senthil Kumar (LM- I1504760) is currently working as an Associate Professor in Computer Science and Engineering department at SRM Valliammai Engineering College of Tamil Nadu. He is a CSI-Student Branch Counsellor of the College. His research interests are in IOT, Big Data, Software Engineering and development of new tools for effort estimation.

# Artificial Intelligence And Its Impact On Human Lives

Y. Dikshita

## Introduction:

Artificial Intelligence (AI) is a branch of Science which deals with helping machines find solutions to complex problems in a more human-like fashion. This generally involves borrowing characteristics from human intelligence, and applying them as algorithms in a computer friendly way. A more or less flexible or efficient approach can be taken depending on the requirements established, which influences how artificial the intelligent behavior appears.

From a business perspective AI is a set of very powerful tools, and methodologies for using those tools to solve business problems. From a programming perspective, AI includes the study of symbolic programming, problem solving, and search.

Search may include a variety of techniques.Typically AI programs focus on symbols rather than on numeric processing.

### LISP

LISP, is the early programming language developed in the 1950s, strongly associated with AI. LISP is a functional programming language with procedural extensions. LISP (LISt Processor) was specifically designed for processing heterogeneous lists - typically a list of symbols.

## Importance Of Artificial Intelligence (AI)

It might be easier to state what part of our modern society artificial intelligence (AI) hasn't touched to show how important it is to our daily lives, business operations and society. Intelligence machines are influencing nearly every facet of our lives to help improve efficiencies and augment our human capabilities. AI is so intertwined in all that we do; it's hard to imagine living life without it. As a simple application in India AI can be used to identify garbage accumulations, water seepage in municipal area to monitor.

AI is a central tenet for the disruptive changes of the 4th Industrial Revolution; a revolution that will likely challenge our ideas about what it means to be a human and just might be more transformative than any other industrial revolution we have seen yet.

AI analyzes more and deeper data using neural networks that have many hidden layers. Building a fraud detection system with five hidden layers was almost impossible a few years ago. All that has changed with incredible computer power and big data. We need lots of data to train deep learning models

because they learn directly from the data. The more data we can feed them, the more accurate they become.

## The Working Mechanism Of Artificial Intelligence

Computers are good in following established processes, i.e., sequences of steps to execute a task. If we give a computer steps to execute a task, it should easily be able to complete it. The steps are nothing but algorithms. An algorithm can be as simple as printing two numbers or as difficult as predicting who will win an election in the coming year! So, how can this be accomplished ?

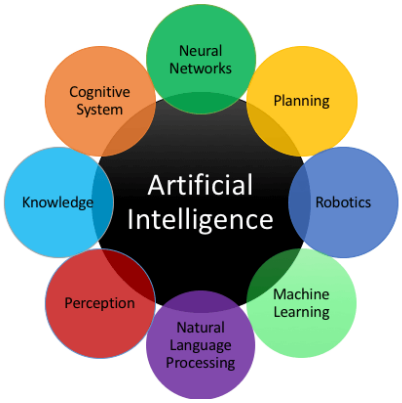
Let's take an example of predicting the weather forecasting in 2021. For this, we need a lot of data. Let's take the data from 2006 to 2019. Now, we will divide this data in an 80:20 ratio. 80 percent of the data is going to be our labeled data, and the rest 20 percent will be our test data. Thus, we have the output for the entire 100 percent of the data that has been acquired from 2006 to 2019.

These data, which are labeled, are fed into the computer. This labeled data constitute 80 percent of data. Here, the algorithm is learning from the data which has been fed into it.

Next, we need to test the algorithm. Here, we feed the test data, i.e., the remaining 20 percent of the data, to the machine. The machine gives us the output. Now, we cross verify the output given by the machine with the actual output of the data and check for its accuracy. While checking for accuracy if we are not satisfied with the model, we tweak the algorithm to give us the precise output or at least somewhere close to the actual output. Once we are satisfied with the model, we then feed the data to the model so that it can predict the weather forecast for the year 2021.

Artificial Intelligence works with large amounts of data which are first combined with fast, iterative processing and smart algorithms that allow the system to learn from the patterns within the data. This way, the system would be able to deliver accurate or close to accurate outputs. As it sounds, AI is a vast subject, which involves much-advanced and complex processes, and hence its field of study includes many theories, methods, and technologies. The major subfields under AI are explained below:

## Major Subfields of Artificial Intelligence:



(https://www.scoro.com/blog/artifical-intelligence-everything-you-want-to-know)

Cognitive Computing: The ultimate goal of cognitive computing is to imitate the human thought process in a computer model. How can this be achieved? Using self-learning algorithms, pattern recognition by neural networks, and natural language processing, a computer can mimic the human way of thinking. Here, computerized models are deployed to simulate the human cognition process.

Computer Vision: Computer vision works on allowing computers to see, recognize, and process images, the same way as the human vision does, and then it provides an appropriate output. Computer vision is closely related to Artificial Intelligence. Here, the computer understands what it sees, and then analyze it, accordingly.

Machine Learning: Machine Learning is the learning in which a machine can learn by its own from examples and previous experiences. The program developed for it need not be specific and is not static. The machine tends to change or correct its algorithm as and when required.See Andrieu,C. et al.,[1].

Neural Networks: Artificial Neural Networks (ANNs) were developed getting inspired by the biological neural network, i.e., the brain. ANNs are one of the most important tools in Machine Learning to find patterns within the data, which are far too complex for a human to figure out and teach the machine to recognize. {See Samuel, A.L. [6]

Deep Learning: In Deep Learning, a large amount of data is analyzed, and here the algorithm would perform the task repeatedly, each time twisting/editing a little to improve the outcome.

Natural Language Processing: Natural language processing means developing methods that help us communicate with machines using natural human languages like English.

# Applications Of Artificial Intelligence

Now, it is time for us to know various real-life applications of AI across industry vertical. AI based computer systems can execute several tasks just like humans, more often in an efficient way than humans saving in time and energy. AI was proposed by McCarthay in[4]. Future activities such as visual understanding, words understanding and paraphrasing etc., wil be done by AI. AI is being used in almost all sectors of economy. In agriculture, medical and health care services, education, legal , energy, banking sectors, social media platforms etc., AI finds its applications. In the present scenario, AI can be profitably used in detecting Carona cases given the symptoms given, constitutional structure of the patient and the environment he lives /or moved in. See Kurmude et al., [3].

### Fraud Detection

Every time you make a transaction online/offline, using your credit or debit card, you receive a message from your bank asking if you have made that transaction. The bank also asks you to report if you haven't made the transaction. The bank feeds its Artificial Intelligence system with data regarding both fraudulent and non-fraudulent transactions. The AI system learns from this data and then predicts which transactions are fraudulent and which are not based on this huge training set.

### Retail Business

The market size of AI software is expected to reach up to $36 million by 2025. This hype in the market has caused retailers to pay attention to Artificial Intelligence. Thus, the majority of big- and small-scale industries are adopting AI tools in novel ways across the entire product life cycle- right from the assembling stage to the post-sale customer service interactions.

### Autopilot

With the AI technology, the pilot only needs to put the system on the autopilot mode and then the majority operations of the flight will be taken care of by AI itself. It is reported by The New York Times that only seven minutes of human intervention (which mostly relates takeoff and landing) is required for the average flight of a Boeing plane.

Autonomous cars are now here. FORD and other companies are fast imple- menting AI in their organizations. Autonomous cars are self driving cars without the need of a human interaction. They consist of precise computer systems connected to motion sensors and cameras. These are special computers which are trained to recognize other cars and human beings. [See Bacchus,F.,et al.,[2].

## Limitations:

There is a danger in the implementing/usage of AI in that humanity could extinct because of excessive dependency. People may lose their jobs and this will lead to a vicious circle.

## Challenges:

Artificial intelligence is going to change every industry, but we have to understand its limits. The principle limitation of AI is that it learns from the data. There is no other way in which knowledge can be incorporated. That means any inaccuracies in the data will be reflected in the results. And any additional layers of prediction or analysis have to be added separately.

Today's AI systems are trained to do a clearly defined task. The system that plays poker cannot play solitaire or chess. The system that detects fraud cannot drive a car or give you legal advice. In fact, an AI system that detects health care fraud cannot accurately detect tax fraud or warranty claims fraud.

In other words, these systems are very, very specialized. They are focused on a single task and are far from behaving like humans.

Likewise, self-learning systems are not autonomous systems. The imagined AI technologies that you see in movies and TV are still science fiction. But computers that can probe complex data to learn and perfect specific tasks are becoming quite common.

# Conclusions:

AI is gaining a lot of ground and its impact is felt in many sectors of our economy. Yet there is a growing fear that the widespread implementation of AI will erode human jobs. There is a fear that AI systems may pave a way for large-scale violence in the world. But that is a very myopic way of looking at things!

In recent decades, technology has grown rapidly and massively economy leading to a lot of growth in economy. AI finds its applications in many sectors. But it has its limitations.

***REFERENCES:***


[1] Andrieu, C., de Freitas, N., Doucet, A., and Jordan, M.I. (2003). An introduction to MCMC for machine learning. Machine Learning, 50(1-2)

[2] Bacchus, F., Grove, A.J., Halpern, J.Y., and Koller, D. (1996). From statistical knowledge bases to degrees of belief. Artificial Intelligence, 87(1-2): 75-143. http://www.cs.toronto.edu/fbacchus/Papers/BGHKAIJ 96.pp

[3] Kurmude D.V., Kakarwal S,N., Deshmukh, R.R (2019). Artificial Intelligence : Simplifying Human Life, CSI Communications, Vol 42, Issue No. 10 p 6-7.

[4] McCarthay John (1977)."Epistemological problems of artifical intelligence".

[5] Samuel, A. L. (1959). "Some Studies in Machine Learning Using the Game of Checkers". IBM Journal of Research and Development. 3 (3): pp 210–229.

[6] Schank, Roger C. (1991). "Where's the AI". AI magazine. Vol. 12 no. 4, p. 38.


***Compiled by***

Ms. Dikshita Yerram is pursuing IV year II semester (ECE) at Anurag Group of Educational Intuitions, Hyderabad.

***Under the guidance of***

Dr. D. D. Sarma, Chief Scientist (R), CSIR-NGRI- Hyderabad.

# A Framework for Distributed Machine Learning

Eesha Kurode, Pranav Joglekar,
Shivam Hire, Amit Joshi

Abstract:
Deep learning models have high time-complexity for training. A lot of power is wasted in labs of universities and colleges where the computer systems are kept running for the whole day and used only for a short duration. This article proposes a framework to reduce the training time of deep learning models by harnessing the power of idle machines. This is done by using cluster computing. The scope of this article extends to deep neural networks only.

## Introduction:

Cluster computing is a form of computing in which a huge task is divided into smaller subtasks and distributed to nodes in a cluster. The nodes work parallelly to increase the efficiency of the task. The nodes are set of machines or computers which are interconnected via LAN. The communication between the nodes is carried out using Message Passing Interface (MPI), a standardized and portable message-passing standard [1].

This article focuses on an image classification task using supervised learning on a neural network. A neural network is a network of neurons grouped together into a set of layers. Each layer performs a set of functions and the final layer i.e. the output layer gives the results. All edges connecting the neurons have weights which determine the computation between the corresponding neurons. A supervised learning algorithm takes data as a pair of input and its expected output. Its aim is to learn the classification model/weights in the neural network by backpropagating the error. The gradient descent is an optimization algorithm which is used to update the weights of the neural network according to the backpropagated error.

The demonstration is that of classification of the MNIST dataset (handwritten digital database) using the mini-batch variant of gradient descent algorithm [2]. Mini-batch gradient descent splits the training dataset into small batches that are then used to calculate model error and update model coefficients.

## Related Work:

In recent times, due to the exponential increase in data and complexity of models, parallel machine learning algorithms have been extensively researched. For instance, in “Large Scale Distributed Deep Networks” a framework for parallel distributed training of two gradient descent algorithms has been

introduced [3]. In “Efficient Mini-batch Training for Stochastic Optimization” a variant of mini-batch SGD whose convergence rate does not degrade when the batch size increases has been proposed. [4]

Cloud-based services such as Google Colab’s distributed training with TensorFlow and Amazon Web Services also provide a framework for distributed machine learning [5] [6]. Kuo Zhang, Salem Alqahtani, Murat Demirbas in their paper have described a comparison between all such frameworks [7]. Our aim is to create a framework that is easy to use locally, is very intuitive and enhances speed of training.

Framework Design

The frontend provides the following functions via a graphical interface:

- Adding/removing nodes from cluster
- Viewing the details of the cluster
- Uploading a pre-processed dataset for training
- Tuning parameters for training
- Training the model and viewing results
- Downloading the trained/pre-trained model

The backend performs two broad tasks, elaborated below:

Setting up a cluster is a difficult and tedious task which involves checking the system requirements of available machines and then running several commands to connect them together. This is automated in our framework. Nodes will be added or removed whenever the corresponding command is raised by the user.

The second and the most important task is to train the model in a distributed manner. This can be done in three ways:

Distributing the layers of neural network amongst different nodes

Although the inter-node communication required in this approach is less, a set of nodes remain idle for most of the training time. This is because the processing in neural networks proceeds layer-wise; that is one layer cannot start processing before all the previous layers have completed their computations. In addition to this, if there are disproportionate number of neurons in layers, then the load of processing won’t be distributed equally. This will cause a bottleneck of the set of nodes which have been assigned with the layers having the most neurons.

Distributing neurons within each layer between different computers

This approach ensures less idle nodes and thus more parallelism. Equal distribution of computation is also possible. However, more message passing will occur in this case as each node will have to communicate its computed values to all other nodes.

### A combined approach

In this, the first approach is applied on a broader level but multiple layers are assigned (such that the total number of nodes which is assigned to the set of nodes remain constant). It is ensured that the

layers are at least not consecutive. This results in lesser inter-node communication as also; nodes do not remain idle as multiple non-consecutive layers are assigned to them.

Using the mini-batch gradient descent, the entire pipeline of computers remains non-idle. As soon as one sample of a batch passes the first layer and goes to the next layer, another sample is given as input to the first layer and so on. Once all the samples in one mini-batch are processed, all layers wait for the backpropagated error values and then update parameters. This waiting time is the only time when nodes may remain idle. The idea of this approach is depicted with an example in Figure 1.

Our framework implements the combined approach to parallelize the training process.

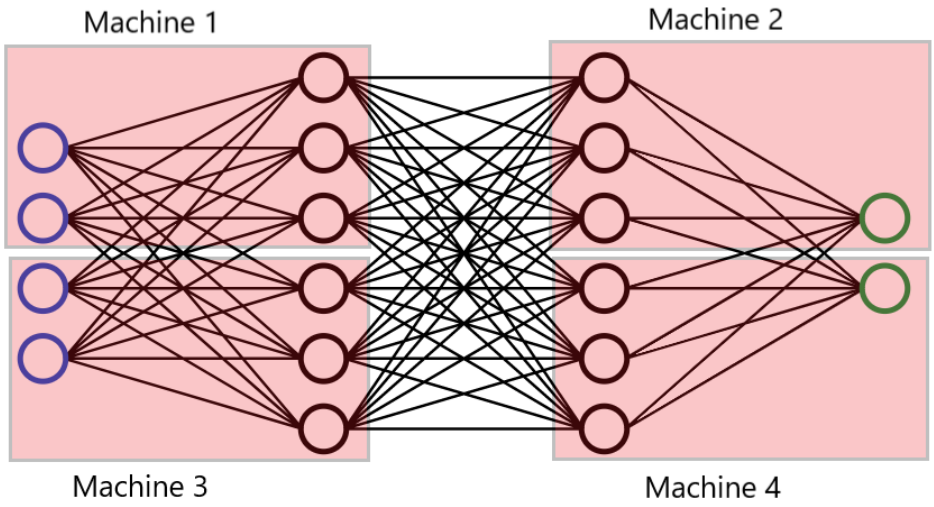



Figure 1: A neural network distributed amongst 4 nodes in a cluster

# Conclusion:

In this article, a heuristic algorithm for training a neural network for an image classification task using cluster computing is proposed. The solution proposed is very user-friendly and automates the task of cluster setup. Our framework is expected to significantly reduce the training speed as compared to a single-machine based training.

For future work, we intend to extend the algorithm to various other applications of neural networks as well as to other deep learning models like Decision Trees, SVMs, Autoencoders, etc.

*References*

[1] http://www.mpich.org.

[2] http://yann.lecun.com/exdb/mnist/.

[3] Jeffrey Dean, Greg S. Corrado, Rajat Monga, Kai Chen, Matthieu Devin, Quoc V. Le, Mark Z. Mao, Marc'Aurelio Ranzato, Andrew Senior, Paul Tucker, Ke Yang, Andrew Y. Ng, "Large Scale Distributed Deep Networks," in Neural Information Processing Systems, 2012.

[4] Mu Li, Tong Zhang, Yuqiang Chen, Alexander J. Smola, "Efficient Mini-batch Training for Stochastic Optimization," in ACM, 2014.

[5] https://www.tensorflow.org/guide/distributed_training.

[6] https://www.tensorflow.org/guide/distributed_training.

[7] K. Zhang, S. Alqahtani and M. Demirbas, "A Comparison of Distributed Machine Learning Platforms," in International Conference on Computer Communication and Networks (ICCCN), 2017.

[9] https://docs.aws.amazon.com/dlami/latest/devguide/distributed-training.html.

## Compiled by:

Eesha Kurode is a third year B.Tech. (Information Technology) student at College of Engineering, Pune, Maharashtra, India. She can be reached at kurodeer17.it@coep.ac.in

Membership ID: 01475185

Pranav Joglekar is a third year B.Tech. (Information Technology) student at College of Engineering, Pune, Maharashtra, India. He can be reached at joglekarp17.it@coep.ac.in

Shivam Hire is a third year B.Tech. (Information Technology) student at College of Engineering, Pune, Maharashtra, India. He can be reached at hiresc17.it@coep.ac.in

## Under the guidance of:

Mr. Amit Joshi (CSI: 01180301) is a professor at Department of Computer Engineering and Information Technology & Coordinator of Computer Society of India, College of Engineering Pune Chapter. He can be reached at adj.comp@coep.ac.in

# Generating 2D Julia Set Fractals using CUDA

Jugal Rawlani, Amit Joshi

Abstract:
CUDA is a parallel computing platform that enables developers to use CUDA-enabled GPUs for general-purpose programming. CUDA programming model allows developers to use popular programming languages along with CUDA APIs to optimize complex computation using GPU level parallelism. Julia Set is a set of points over a complex plane for which the series ZN+1 = ZN2 + C does not tend to infinity. This article illustrates the usage of the CUDA platform to reduce the time consumed for generating 2D Julia Set Fractals. It also covers the environmental setup required to run CUDA based applications.

## 1. Introduction:

CUDA (Compute Unified Device Architecture) is a general-purpose parallel computing platform and programming model, introduced by NVIDIA, that enables developers to leverage the immense parallel processing power of many-core CUDA-enabled GPUs to compute complex problems. The CUDA platform is devised to work with popular programming languages such as C, C++, Python, FORTRAN, and MATLAB. [1, 3]

CUDA platform reduces the runtime of applications involving various extensive and complex computations using the CUDA-enabled many-core NVIDIA GPUs [2]. A performance comparison between CUDA and OpenCL revealed that CUDA performed better both in transferring data to and from the device (GPU) to host (CPU), and the kernel runtime for CUDA's kernel was consistently faster than the rivalling OpenCL's kernel [5]. Soon after its release, a variety of industries, such as Medical Imaging, Computational Fluid Dynamics, migrated to CUDA based applications enjoying an impressive deal of increase in performance as compared to their previous state-of-the-art implementations. [4].

Julia Set is a set of complex numbers of the form Z = X + iY, all of which do not tend to infinity (or do not diverge) for the series ZN+1 = ZN2 + C, where C is a complex constant [7].

## 2. Setting up the Environment:

This section will walk you through the basic setup of NVIDIA drivers and SDKs necessary for a system to run CUDA based applications. This section will also cover how you can compile a CUDA based application on your system, after the initial setup. Following steps will guide you to get your system ready to compile and run CUDA based applications.

*Pre-installation Actions:*

Pre-installation Actions basically consists of checking if your system is CUDA-enabled, and you have the correct version of all the required packages.

*CUDA-enabled GPU:* To verify that your GPU is CUDA-capable, you can check the list of CUDA-capable GPUs at CUDA-enabled GPUs [8].

Command to check what GPU you have: $ lspci | grep -i nvidia

*Linux Version Support*: CUDA toolkit is available for some specific distributions of Linux, you can check the list of supported Linux distributions at CUDA Toolkit Release Notes [3]. Command to check the Linux distribution that you are using:
$ uname -m && cat /etc/*release

*gcc:* gcc is a necessary requirement for CUDA toolkit, and a specific version of gcc is required as per the CUDA toolkit version, check the supported gcc version at CUDA Installation Guide [7]. Command to check the version of gcc:
$ gcc –version

*Downloading the CUDA toolkit:* CUDA toolkit can be downloaded from CUDA Downloads [9]. Make sure you download the runfile (local) version.
*Installing CUDA toolkit*: Once CUDA toolkit is downloaded, it can be installed using the following command: $ sudo sh cuda_<version>_linux.run

CUDA toolkit will then be installed.

*Reboot the system:*

This is an important step, as it allows the system to load NVIDIA drivers on boot-up.

*Post-Installation Actions:*

*Update PATH variable*: Add the following line at the end of ~/.bashrc
export

PATH=/usr/local/cuda-10.2/bin:/usr/local\/cuda-10.2/NsightCompute-2019.1${PATH:+:${PATH}}

*Update Environment PATH:* For 64-bit Operating Systems add the following line at the end of ~/.bashrc
export LD_LIBRARY_PATH=/usr/local/cuda-10.2/lib64\
${LD_LIBRARY_PATH:+:${LD_LIBRARY_PATH}}

For 32-bit Operating Systems add the following line at the end of ~/.bashrc
export LD_LIBRARY_PATH=/usr/local/cuda-10.2/lib\
${LD_LIBRARY_PATH:+:${LD_LIBRARY_PATH}}

*Verifying Installation:* You can verify the CUDA toolkit installation by running the following command:
$ nvcc --version

For a successful installation, it will display the CUDA toolkit version.

For more detailed steps, refer CUDA Installation Guide [7].

# 3. The Problem Statement:

To visualize Julia Set Fractals, consider a plane where each point (x, y) represents a pixel, and for each point, we calculate the following series:

Zn+1 = Zn2 + C

(where Z0 = x + iy)

Based on whether the series for a particular point converges or diverges, we assign a colour to the respective pixel.

Considering an image of 1000 x 1000 pixels, and a threshold of 'n' as 200, we have to perform a total of 1000 * 1000 * 200 operations to get the value of all the pixels in the image, i.e. 2 * 108 operations for this particular example, which can require quite some time to execute.

Hence, coming to the problem statement:

Using CUDA technology to reduce the time required to generate 2D Julia Set Fractals.

# 4. The Program:

CPU

```
void kernelCPU(Pixel *pixelPtr) {
   int i = 0, j = 0;
   for(j = 0; j < DIM; j++) {
      for(i = 0; i < DIM; i++) {
         int offset = i + j * DIM;
         int juliaValue = checkPointForJuliaSet(i, j);
         // set value for pixels
         pixelPtr[offset].r = 255 * RED * (juliaValue ==
               2 ? 1 : 0);
         pixelPtr[offset].g = 255 * GREEN * juliaValue;
         pixelPtr[offset].b = 255 * BLUE * (juliaValue
          == 1 ? 1 : 0);
      }
   }
}
```

// The CPU kernel iterates over all DIM x DIM pixels serially and computes whether the point belongs to Julia Set or not.

GPU

```
__global__ void kernelGPU(Pixel *pixelPtr) {
   int x = blockIdx.x;
   int y = blockIdx.y;
   int offset = x + y * gridDim.x;
   int juliaValue = checkPointForJuliaSet(x, y);
   pixelPtr[offset].r = 255 * RED * (juliaValue ==
                                        2 ? 1 : 0);
   pixelPtr[offset].g = 255 * GREEN * juliaValue;
   pixelPtr[offset].b = 255 * BLUE * (juliaValue ==
                                        1 ? 1 : 0);
}
```

// The GPU kernel splits the work, of calculating if a pixel belongs to Julia Set or not, to blocks, which do this parallely, and each block effectively checks value for only one pixel.

# 5. Results and Discussion:

Test System Specifications:

CPU: Intel Core i5-7200U @ 2.5GHz

GPU: NVIDIA GeForce 940MX (4GB Memory)

For an Image of dimension 1000 x 1000 pixels:

Time taken by CPU: 1.0491742 seconds

Time taken by GPU: 0.000020 seconds or 20 microseconds

For an Image of dimension 10000 x 10000 pixels:

Time taken by CPU: 104.785264 seconds

Time taken by GPU: 0.0000226 seconds or 22.6 microseconds

The above readings have been averaged over 5 times for accurate results.

Following attached Images are some of the generated Julia Set Fractals each of the image has a dimension of 10000 x 10000 pixels:

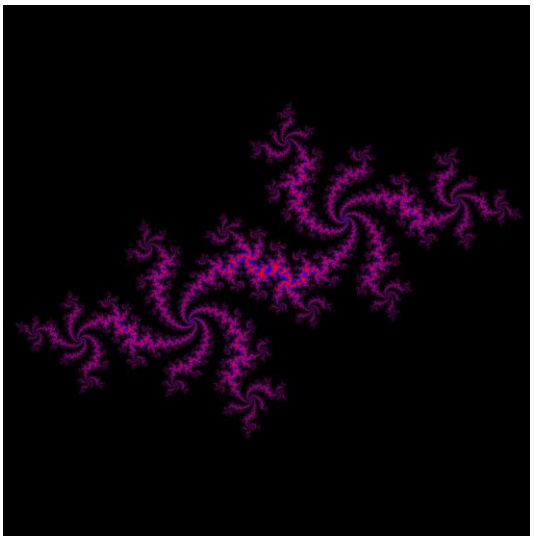


*Figure 1: C = -0.54 + 0.54i*

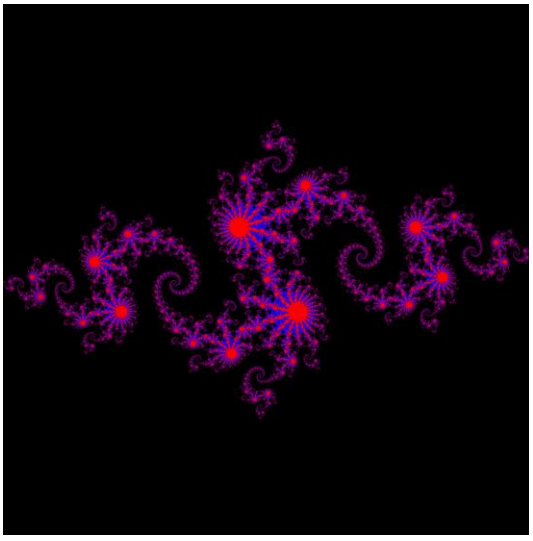


*Figure 2: C = -0.8 + 0.156i*

# 6. Conclusion:

CUDA enables developers to leverage the power of highly parallel many-core GPUs to solve many complex problems in a time-efficient manner. When comparing the time required to generate 2D Julia Set Fractals by CPU vs CUDA-enabled GPU, we can see that the CPU calculates the value for one pixel at

a time, and hence takes a substantial amount of time to complete, on the other hand, CUDA-enabled GPU takes a few microseconds to complete the calculation by taking advantage of the parallelism. In essence, we can employ CUDA technology to compute complex problems faster than the CPU.

*References:*

[1] An Overview of CUDA platform https://en.wikipedia.org/wiki/CUDA
[2] Official CUDA Documentation by NVIDIA https://docs.nvidia.com/cuda/
[3] CUDA toolkit Release Notes https://docs.nvidia.com/cuda/cuda-toolkit-release-notes/index.html
[4] CUDA by Example, A GPU Programming Book https://developer.nvidia.com/cuda-example
[5] A Performance Comparison CUDA v/s OpenCL https://arxiv.org/abs/1005.2581 (173 Citations).
[6] Julia Set Fractal http://paulbourke.net/fractals/juliaset/
[7] An Exploration of Julia Sets https://www.mcgoodwin.net/julia/juliajewels.html
[8] CUDA Installation Guide https://docs.nvidia.com/cuda/cuda-installation-guide-linux/index.html
[9] CUDA-enabled GPUs https://developer.nvidia.com/cuda-gpus
[10] CUDA Downloads https://developer.nvidia.com/cuda-downloads

Compiled by:

Mr. Jugal Rawlani, student of Final Year BTech (Computer Engineering) at College of Engineering, Pune (COEP), Maharashtra.

Email: rawlanijr16.comp@coep.ac.in

Under the guidance of:

Mr. Amit Joshp [CSI : 01180301], is working as Assistant Professor in Dept. of Computer Engineering and IT, College of Engineering, Pune (COEP), Maharashtra. He can be reached at: adj.comp@coep.ac.in

# Reasons To Switch To Linux

## Aravind A

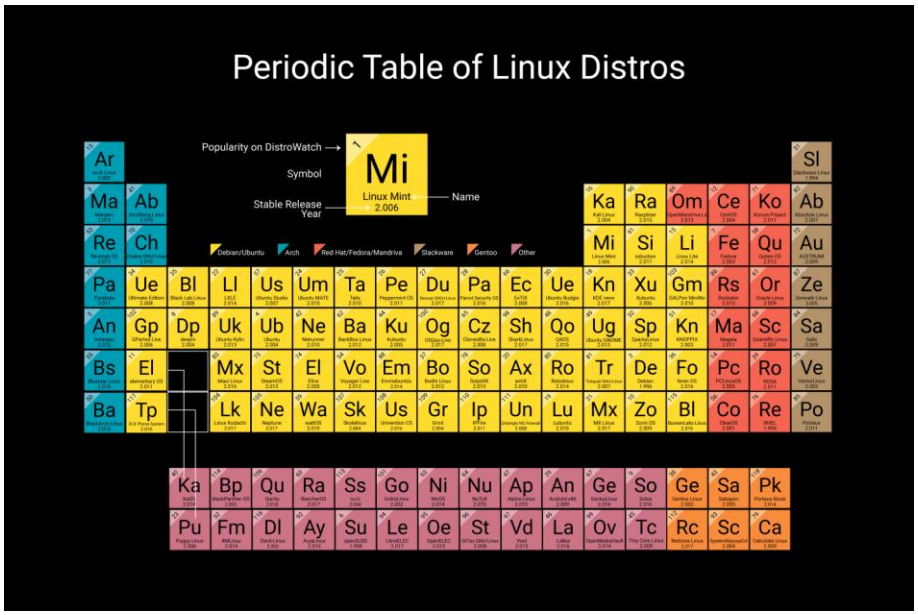

We can all agree that Windows is the most ubiquitous operating system around. I mean, if you are reading this on a PC, chances are that it will be running Windows. But since Windows discontinued support for Windows 7 and Dr. Piracy got locked down in the recent version of Windows, people are searching for alternatives (barring the extremely committed pirates) are mostly landing on the Linux ecosystem. In fact, the adoption of the Linux ecosystem has considerably increased in the past year or two. So in this article I will show you the main reasons that you should consider switching to good old tux.

So on with the show

### 1.It's FREE……

This is the most obvious reason Linux is a good alternative for Windows. You don't have to pay a license fee, you just download it and it's yours. That's it. Its this sense of freedom that is translated into the whole ecosystem. This is due to the GNU General Public Licensing scheme . Basically it means that you basically own part of the system, not… specifically you the reader, but the developers who contribute to the operating system like bug fixes, adding new features or optimizing the ecosystem. And because they helped in the development of the operating system, the technically own that piece of code. This level of openness ensures that Linux as a platform is a stable and well optimized platform. This is also helped by organizations like the Linux foundation whose main goal is to standardize Linux for a wide range of applications(backed by IBM, Microsoft, Intel etc..). This is a win for both the producers who deploy services using Linux and the consumers using these services.

### 2.Options




The other thing that Linux has going for is the variety of options or distributions available. If you want privacy, there is a Linux for you. If you want to do content creation, there is a Linux for you. If you want to be a hacker, there is a Linux for you. Chances are if you have any kind of needs, there is a Linux distro specifically for you. And if you still cant find what you need, you can apt-get anything and make your tux just the way you want it. And since these distros can be run off a USB stick or a live CD, you can give it a test run without affecting your existing system. Its this level of choice that makes it a viable choice to switch to.

### 3.Performance

The next thing is the performance Just like the options available for every use-case, there are a variety of distributions that can run on a variety of hardware ranging from a supercomputer to a microwave oven. This is why Linux has a long standing tradition of breathing new life to old computers.

This is possible thanks to the modular approach of the platform allowing the kernel to be juiced up or shaved down depending on needs. This allows Linux to run on a variety of devices ranging from old desktop computers to workstations to IoT devices (which is the largest platform using Linux by the way), to even your mobile phone (ehm…….Android). Distros like tiny core can run on just 16MB of RAM, with a GUI !! You can even run a Linux distro on a your system RAM(tails linux). No hard drive required. And distros like raspbian allow you to use a low powered device like a raspberry pie as a pretty good general purpose desktop(You can even play games in it !).

### 4.The App Ecosystem

We Linux users have trolled windows users for going to the developers website and downloading the required drivers and software and painstakingly installing all of them.

Well there is a reason for that. In Linux, once you install the operating system, all the package dependencies and device drivers will be automatically installed for you. No fumbling around required. And when you want to install an app, centralized repositories like pacman, apt, rpm etc.(Commonly

referred as package managers) let users have a one-stop place to seamlessly download and install all the required packages. Just one command, and all the required packages are installed and optimized. And if you still don't find it there, external repositories like flatpak and snap allow a wide variety of software for almost any distribution you are running. Now operating systems like Windows are now implementing these app repositories. But Linux was using this system since the early 90's (Google it !).

### 5.Stability & Transparency


Now if you have been following the windows train, you might have come across the various bugs and the forcing of windows update is also a pain in the a$$.

But good news…. In Linux, you don't have to worry about any of that. Since most of the Linux software and drivers are open source, it means that if a company releases a piece of software, the community most likely have tinkered with it and ironed out all of the issues. This means that the drivers and kernel are rock stable.

Also, because of its server background, Linux can be run indefinitely without the need of rebooting in between. This makes it ideal for enterprise usage. No wonder Microsoft are using Linux for their Azure platform (Even though they have a Windows version specifically for the job) !

Also you can choose to update the system and apps (emphasis on the “choose” you can do it when you want, how you want, or not do it at all (Windows stopped working after hearing this..;)). This level of control also allows you to monitor everything in your system (And I mean EVERYTHING!).

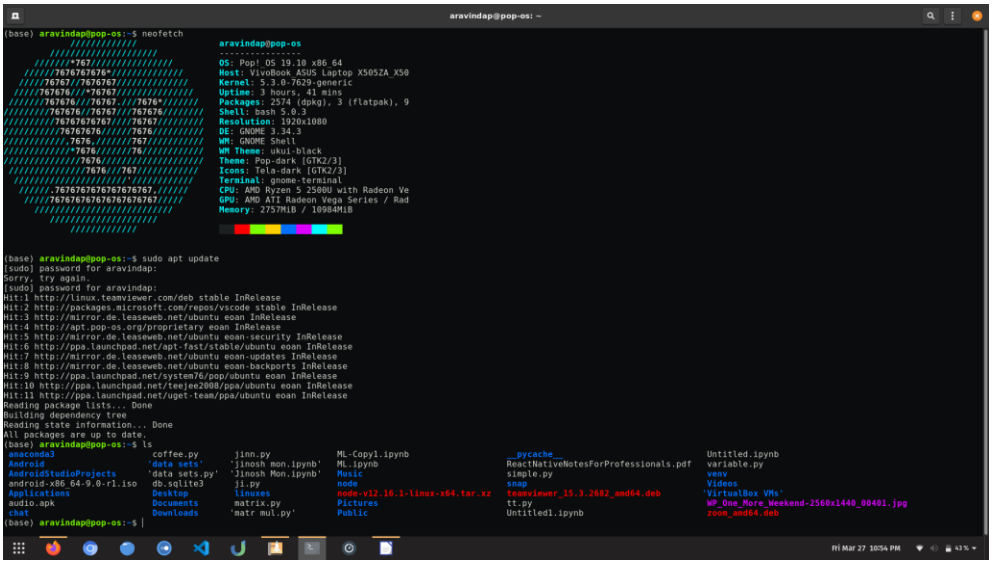
Another thing is that because of this level of control, Linux is extremely transparent. A good example is Ubuntu’s latest data collection policy. They say it upfront and you can monitor what data gets sent or just opt out of it as you choose.

In addition to that, the open-source nature means that since there are thousands of eyes in the code, there is very little chance of any fault or malware creeping in to your system. This level of transparency is also a thing that sets Linux apart.

## 6.Customization

This is a field where Linux truly shines. Not only do they have a wide variety of desktop styles to choose from and they were the first to implement virtual desktops, which are basically multiple desktops with various apps which you can easily switch to helping to keep things organized, but interfaces like KDE and Gnome allows users a variety of extensions and theming options (apart from the usual wallpapers and icons). You can even make custom interfaces for specific purposes or heck you can go full old school and use the command line. Since Linux is mostly based around server and enterprise usage, most if not all of the Linux functions can be done through the command line. Again, the community is there and you can make some pretty sweet setups suiting your needs.

## 7.The Command Line/ Terminal




The terminal is what comes to mind when we think of Linux. Its not something to be intimidated of. Its just old fashioned. Once you get the hang of it, you can use the terminal to control basically anything. Just type a command and there you go! The objective of the terminal is that its the fastest way to get stuff done. Its basically a lean, mean, coding machine. In fact you can even use “one-liners” which are commands that execute multiple commands at once making the process much, much more efficient.

There are even scripts that you can use to automate tasks via the terminal. So if you are using Linux or are intending to switch to it (Which is the whole purpose of this article ;) ) its worth a shot looking into it.

### 8.Portability

Linux uses something called a live kernel. This means that the OS can be run on a number of different systems with just a USB drive without affecting the system its being run on. This is a very valuable feature for software testers and people who don't want to take their system with them. Just flash the ISO on a flash drive, plonk it into a system and there you go. And because of this live nature, crucial software and driver updates can be done while the system is running, without a reboot.

Also since all the data is stored in a root folder rather than in specific registers, distro hopping is extremely easy. Just backup your root directory, install a distro and transfer your root directory to it and you are up and running. In fact there is a script in github you can download and run and it will download all the required packages, install all apps and data and basically make your system up and running. With just one click!!! Its this level of portability that makes Linux ideal for enterprise usage (like deploying a full suite of workstations) and helpful for many users like me who jump from distribution to distribution a lot.

### 9.Gaming

A rather interesting use case, Linux has come a long way in terms of gaming. Valve's SteamOS was a Linux distribution made specifically for gaming. Plus there are a number of distros targeted specifically at gamers (Manjaro, Nitrux just to name a few). There are thousands of Linux specific games in the steam store, and thanks to valve's proton API and game stores like lutris, games that are being ported from other platforms are growing by the day. To add icing on the cake, Google's game streaming service, Stadia specifically use an enterprise version of Debian to run their games in the cloud. So gaming on Linux is a reality…. And its here.

# Conclusion

So these reasons might have been able to convince you to jump to the Linux bandwagon. If you do so, we, the community are here ready to hold your hand through this endless rabbit hole. If you don't, well….. who am I to judge (you're just a moron). Anyway I hope this might be helpful for you or to your friends.

So have I missed anything ? Do let me know if I did. Thanks for taking your time reading it and stay safe.

Compiled by :

Mr. Aravind A (01534892) is a student of B.Tech III year (CSE) at Thejus Engineering College, Vellarakkad (Thrissur), Kerala. He can be reached at aravindap@protonmail.com

Under the guidance of:

Mr. Shine P Xavier (F8002773) is a Assistant Professor / CSE, at Thejus Engineering College, Vellarakkad (Thrissur), Kerala. He can be reached at shinepxavier@gmail.com.

# Engagement Analysis in MOOCs

Anamika Gupta, Saksham Jain,
Prashi Goyal, Abdal Lalit

Abstract

Massive Open Online Courses (MOOCs) have enabled the most peculiar changes in the learning environment allowing students across the globe to learn, practice and explore myriad domains. Paced by the boost in Internet Connectivity across the world, MOOC have turned up as degree and professional courses. Driven by online video lectures, discussion forums, quizzes, practice assignments and reading materials, MOOC provides flexibility in learning.

The writing aims to study engagements of students with the online material and presents analysis done by professionals in various engagement styles viz. cognitive, behavioral and emotional engagements on different material types viz. discussion forums, quizzes and assignments. Additionally, the article presents the clustering of learners based on their engagement levels.

## Introduction

Massive Open Online Course (MOOC) is an online course which aims at a large number of participation of the students and providing open access to a course [2][4]. Some MOOCs also provide discussion forums where students can interact with each other [4]. MOOCs provide problem-centric learning with clear expositions, instructor accessibility and passion, active learning, peer interaction and helpful course resources [5]. There are various MOOC platforms available on the web like Coursera offering 5600+ courses, edX offering 2500+ courses, FutureLearn offering 1500+ courses, Swayam offering 600+ courses and many other platforms which are offering hundreds of courses, some of which have thousands of students enrolled [3][6]. According to the 2018 survey, a list of five top MOOCs providers by registered users was launched which contains the number of registered users on the various platforms: Coursera (37 million), edX (18 million), XuetangX (14 million), Udacity (10 million) and FutureLearn (8.7 million) [14]. There are a range of activities in which students can engage themselves in these courses like watching lectures, taking quizzes to check their understanding, giving assignments or exams, and participating in discussion forums. There is a need to analyse the engagement of students with MOOCs because it will help us in designing future online courses and evaluating their effectiveness [2][3][4].

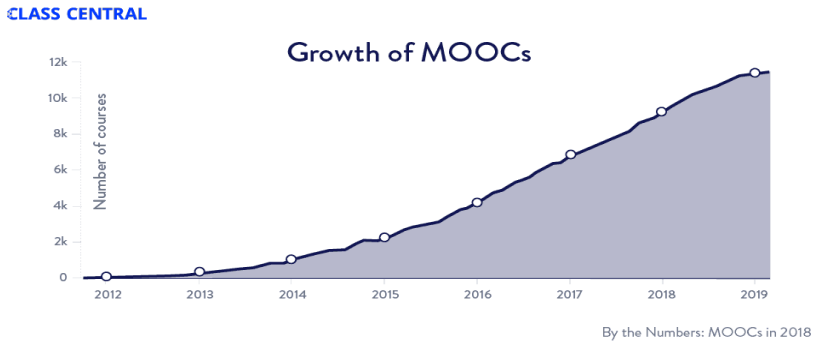



*Figure 1: Growth of MOOCs over the years[14]*

Engagement in MOOCs is considered as an important aspect in the teaching and learning context as it may influence student's retention, their learning, and their achievements in the course. A Student's engagement can be viewed as an observable display or a manifestation of motivation. Generally, engagement is measured by checking whether the students have completed their course or not. Scholars have identified the engagement of the students as a construct that contains three components [2][5]:

Behavioural engagement: This term refers to the participation of students in social, academic and extracurricular activities. It is the student's displayed behaviour that is related to his/her effort towards the learning process for a single activity/assignment or for their overall learning process [1][2][5].

Cognitive engagement: This term refers to psychological investment in learning that ranges from memorizing to using self-regulated strategies to promote one's understanding. It can be analysed by using the clickstream data of video lectures - the student click-event records. The number of pauses is a good indicator of cognitive engagement [1][2][5].

Emotional engagement: It refers to the students' affective responses or feelings towards teachers, peers, the course and learning [2][3][5]. It can be easily measured by the reviews of the courses given by the students using sentiment analysis.

Some researchers wanted to derive the relationship between the different types of engagements, specially between behavioural and cognitive engagements. They derived the degree of association between these engagements. In a study, they took a sample of 50,676 students who had registered for some courses on the Coursera platform, out of which 19,548 students had watched at least one lecture after the registration. In this study, behavioral engagement was measured by the number of lectures a student watched per week while cognitive engagement was measured by considering the number of pauses taken during lectures by the student per week. Then, they applied a standard clustering technique to derive engagement patterns of the students. This helped the researchers in inferring if the students who were behaviorally engaged were cognitively engaged as well. As the final result, they concluded that there were some students who had a high behavioural engagement but a low cognitive engagement. In addition to this, cognitive engagement has a unique contribution in predicting academic

achievement, which can give some extra information to the instructors about the students' performance in the registered courses [2][3].

# Classification of Engagement Styles

Researchers studied the engagement pattern of students on the basis of two fundamental activities: viewing a lecture and handing in the assignments for credit [3]. They used a concept of assignment fraction. It was defined as the fraction of assignments submitted out of the total number of lectures viewed and assignments submitted. It generally denotes that a student with the assignment fraction of 0 has only viewed lectures, while a student with the assignment fraction of 1 has only submitted the assignments without viewing any lecture. Thus, they have characterised the engagement styles into five types [2][3]:

***Viewers:*** Their primary focus is to view the lectures only, with handling in few if any assignments.

***Solvers:*** Their primary focus is to handle the assignments only for a grade, with viewing few if any lectures.

***All-rounders***: There is a balance between both viewing the lectures as well as handling the assignments.

***Collectors:*** They primarily download the lectures and unlike viewers, may or may not actually watch the lectures.

***Bystanders:*** They only register for the course and hardly do they perform any of the activity. Their total activity contribution is below the threshold [3].

These categories of engagement styles can now be used to study engagement in discussion forums. Discussion forums, on any MOOC website, are platforms for learners to engage in peer-to-peer discussion about the course. These forums were built in an effort to make online courses feel similar to offline teaching. Interaction in discussion boards can take many forms - from answering a question to engaging in a rating scheme.

# Analysing Engagement in Discussion Forums

After analysing a data set of students engaged in a course on Coursera, some researchers observed a pattern wherein the average grade and activity level of a student making the initial post was lower than that of students making the subsequent posts. From this, it could be inferred that high-achieving students are helping other students in their class by participating in discussion forums. The researchers calculated values for P(S|F), the probability of engagement style given the forum presence and P(F|S), the probability of forum presence given the engagement style. It was observed that All-rounders were heavily engaged on the forum, while Bystanders comprised only 10 % of the forum participants. [3]

| | Bystander | Viewer | Collector | All-rounder | Solver |
|---|---|---|---|---|---|
| $P(S\|F)$ | 0.106 | 0.277 | 0.192 | 0.408 | 0.017 |
| $P(F\|S)$ | 0.050 | 0.334 | 0.369 | 0.894 | 0.648 |

*Table 1 : showing how engagement styles are distributed [3]*

The researchers also found that the thread length varies almost linearly with the number of distinct contributors to the thread. They observed that a thread with k posts has approximately 2k/3 distinct contributors. [3]

# Engagement in Assignments

Assignments in a MOOC serve as the basis for awarding certificates and counts in the final grade. There are two types of assignments - graded and ungraded. Graded assignments include weekly quizzes, peer review assignments and the ungraded assignments usually include practice assignments and in-video quizzes which are meant to retain the interest and concentration of an individual.

Researchers at Stanford University observed that most students received a grade 0 but they were also watching a sufficient number of lectures. Hence, receiving a zero-grade cannot be attributed to the failure to put in efforts. As shown in Fig 3, the grade received by a student is not a perfect indicator of her engagement. [3]

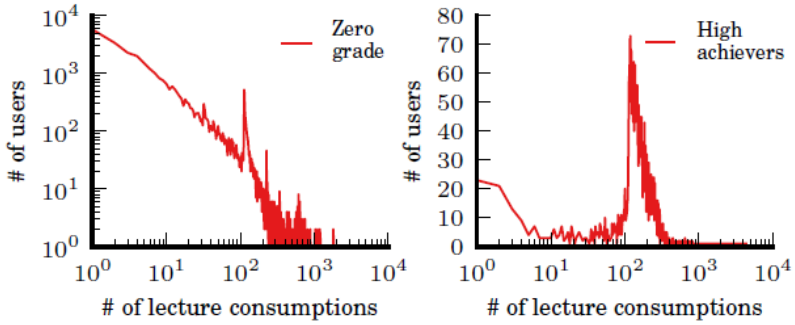
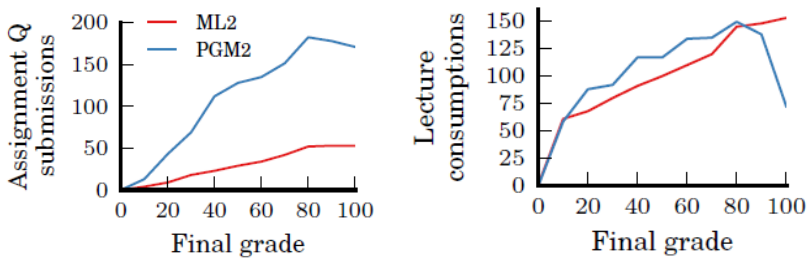



*Figure 2: Comparative analysis of zero-grade receivers and high-achievers [3]*

They analysed the activity of students vs. their grades. As shown in Fig 4, the grades of the students increased almost linear with their activity in the ML2 course. However, in the PGM2 course, this linear relationship was only true for students who achieved a grade of less than 80%. After that, activity somewhat fell. This was because PGM2 was a technical and more challenging course than ML2. This elucidates the difference between MOOCs where the primary indicator of one's grade is the efforts expended (ML2) versus the quality of work handed in (PGM2). [3]




(a) (b)

*Figure 3: (a) and (b) showing the analysis of final grade vs. activity in two course ML2 and PGM2 on Coursera [3]*

# Completion rates in MOOCs

According to a study by Katy Jordan (2013), there were 50 investigated MOOCs which have generated 50,000 enrollments on average, with the completion rate below 10%. If we consider it as 7.5%, it means there are 3700 completions per 50,000 enrollments. Similarly, the dropout rates of MOOCs offered by Stanford, MIT and UC Berkeley were 80-95%. Therefore, it is important to find out why students drop out of courses and suggest strategies like accommodation of students to different time tables, promoting student course completion , increasing student to instructor interaction as well as online learning skills, supplemental tutoring, peer-based assessment, that can be implemented to increase the online retention rate in the future to reduce this drop out rate[12].

The graph below showing the completion rates of the courses in the previous years[14]:

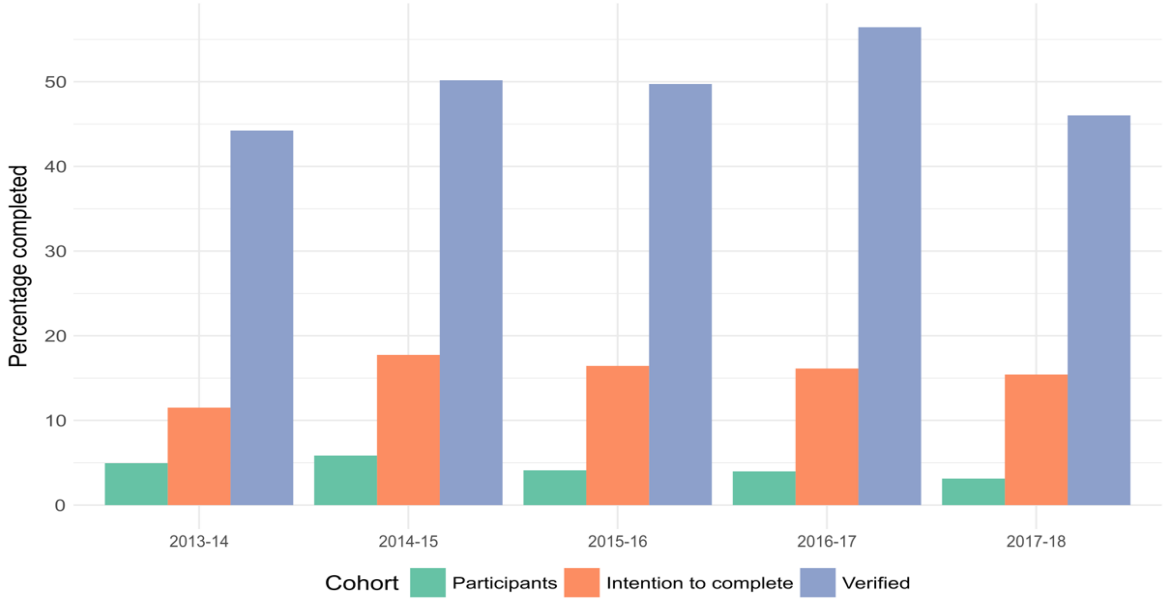



Figure 4: The graph showing completion rates of MOOCs in previous years [14]

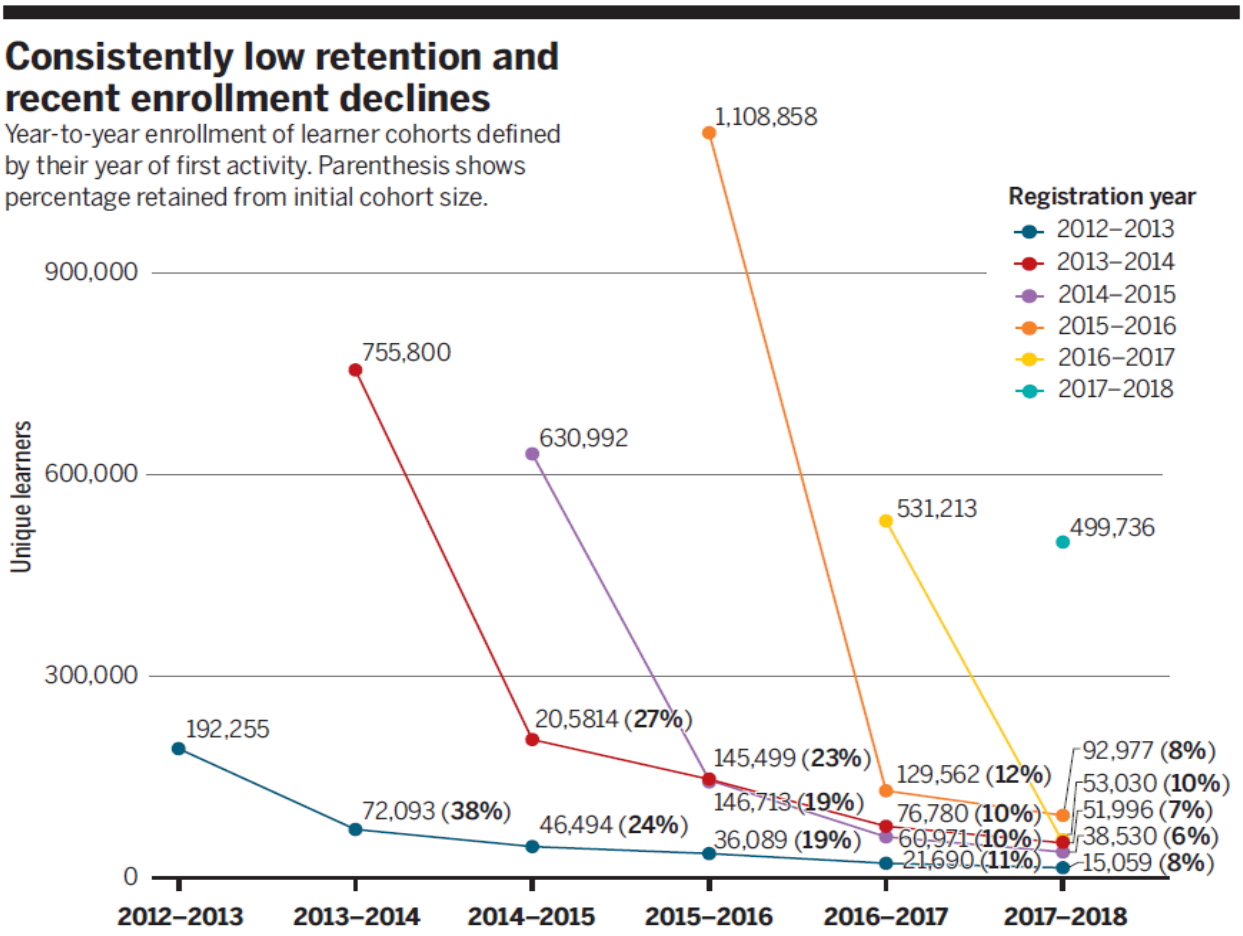



*Figure 5: showing year wise enrollment of learners and percentage of their retention rates [14]*

The table shows below the completion rate date according to research by Jordan(2013)[12]:

| MOOC | Students enrolled | Students completed | Percentage of Completion | Platform |
|---|---|---|---|---|
| Equine Nutrition | 23322 | 8416 | 36.1% | Coursera |
| Astrobiology | 39556 | 7707 | 19.5% | Coursera |
| Functional Programming Principles in Scala | 50000 | 9593 | 19.2% | Coursera |
| Computing for Data Analysis | 50899 | 6420 | 12.6% | Coursera |
| introduction to Machine Learning | 104000 | 13000 | 12.5% | Coursera |
| Databases | 60000 | 6500 | 10.8% | Coursera |
| Image and video processing - From Mars to Hollywood with a stop at the hospital | 40000 | 4069 | 10.2% | Coursera |
| Internet History, Technology and Security | 46000 | 4595 | 10.1% | Coursera |
| Gamification | 81600 | 8280 | 10.1% | Coursera |
| Introduction to Philosophy | 98128 | 9445 | 9.6% | Coursera |
| Critical Thinking | 75884 | 6909 | 9.1% | Coursera |
| Mathematical Biostatistics Bootcamp | 21916 | 2087 | 9.5% | Coursera |
| Sports and Society | 19281 | 1626 | 8.4% | Coursera |
| Introduction to Mathematical Thinking | 27930 | 1950 | 7% | Coursera |
| Software Engineering for SaaS | 50000 | 3500 | 7% | Coursera |
| Introduction to International Criminal Law | 21000 | 1432 | 6.8% | Coursera |
| Drugs and the Brain | 66800 | 4400 | 6.6% | Coursera |
| Listening to World Music | 36295 | 2191 | 6% | Coursera |
| Data Analysis | 102000 | 5500 | 5.4% | Coursera |
| Pattern-Oriented Software Architectures for Concurrent and Networked Software | 30979 | 1643 | 5.3% | Coursera |
| Introduction to Genetics and Evolution | 33000 | 1705 | 5.2% | Coursera |
| Computational Investing, Part 1 | 53205 | 2554 | 4.8% | Coursera |
| An Introduction to Operations Management | 87000 | 4000 | 4.6% | Coursera |
| Greek and Roman Mythology | 55000 | 2500 | 4.5% | Coursera |
| E-learning and Digital Cultures | 42844 | 1719 | 4% | Coursera |
| Introduction to Astronomy | 60000 | 2141 | 3.6% | Coursera |
| Introduction to Sociology | 40000 | 1283 | 3.2% | Coursera |
| Social Network Analysis | 61285 | 1410 | 2.3% | Coursera |
| Human-Computer Interaction (studio track) | 29105 | 791 | 2.7% | Coursera |
| Bioelectricity - a quantitative approach | 12000 | 313 | 2.6% | Coursera |
| Human-Computer Interaction (studio track) | 29105 | 791 | 2.7% | Coursera |
| A Beginner's Guide to Irrational Behaviour | 142839 | 3892 | 2.7% | Coursera |
| Think Again: How to Reason and Argue | 226652 | 5322 | 2.3% | Coursera |
| Social Network Analysis | 61285 | 1410 | 2.3% | Coursera |
| Medical Neuroscience | 44980 | 756 | 1.7% | Coursera |
| A History of the World since 1300 | 83000 | 605 | 0.7% | Coursera |
| Stat2.1x Introduction to Statistics - Descriptive Statistics | 52661 | 8181 | 15.5% | EdX |
| 3.091x Introduction to solid state chemistry | 28512 | 2082 | 7.3% | EdX |
| 6.002x Circuits and Electronics | 46000 | 3008 | 6.5% | EdX |
| CS50x - Introduction to Computer Science I | 150349 | 1388 | 0.9% | EdX |
| Introduction to Inforgraphics and Data Visualization | 2000 | 140 | 7% | Moodle |
| Introduction to Artificial Intelligence | 160000 | 20000 | 12.5% | Udacity |

*Table 2: completion rate date according to research by Jordan(2013) [12]*

Thus, from the table above, we can infer that Coursera is one of the most preferred platforms of MOOCs with maximum number of enrollments of the students but the retention rate is very low.

# Conclusion

The article analysed the existing works in engagement analysis of MOOCs, strategies of increasing the completion rate of MOOCs and have explored the engagement styles and indicators. For instance, the number of pauses during the online lecture serves as a good indicator of the cognitive engagement of the user. The clustering of users in various categories had been based on factors like viewing an online video and submission of the assignment. The analysis of activity in the discussion forums has served as a parameter for grading the MOOC by various researchers. The article also focused on the completion rates of MOOCs on various MOOCs platforms by gathering the data from various sources and suggested some strategies to increase this completion rate.

### *References*


[1] https://www.igi-global.com/dictionary/k-12-online-student-engagement/55712
[2] Li, Q., & Baker, R. (2018). The different relationships between engagement and outcomes across participant subgroups in massive open online courses. Computers & Education, 127, 41-65.
[3] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec.. Engaging with massive online courses. In Proceedings of the 23rd international conference on the World wide web (WWW '14). Association for Computing Machinery, 2014, New York, NY, USA, 687–698.
[4] https://en.wikipedia.org/wiki/Massive_open_online_course
[5] http://www.ifres.ulg.ac.be/MOOC-blog/wp-content/uploads/2016/05/engagement-in-moocs-lu.pdf
[6] https://www.classcentral.com/provider/coursera
[7] Ruiqi Deng, Pierre Benckendorff, Deanne Gannaway, Linking learner factors, teaching context, and engagement patterns with MOOC learning outcomes, Journal of Computer Assisted Learning, 10.1111/jcal.12437, 0, 0, (2020).
[8] Lan, M., Hew, K.F. Examining learning engagement in MOOCs: a self-determination theoretical perspective using a mixed method. Int J Educ Technol High Educ 17, 7 (2020).
[9] R. Wes Crues, Nigel Bosch, Michelle Perry, Lawrence Angrave, Najmuddin Shaik, and Suma Bhat. 2018. Refocusing the lens on engagement in MOOCs. In Proceedings of the Fifth Annual ACM Conference on Learning at Scale (L@S '18). Association for Computing Machinery, New York, NY, USA, Article 11, 1–10.
[10] Ilona Nawrot and Antoine Doucet. 2014. Building engagement for MOOC students: introducing support for time management on online learning platforms. In Proceedings of the 23rd International Conference on World Wide Web (WWW '14 Companion). Association for Computing Machinery, New York, NY, USA, 1077–1082.
[11] Miguel L. Bote-Lorenzo and Eduardo Gómez-Sánchez. 2017. Predicting the decrease of engagement indicators in a MOOC. In Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK '17). Association for Computing Machinery, New York, NY, USA, 143–147.
[12] Khalil, Hanan & Ebner, Martin. (2014). MOOCs Completion Rates and Possible Methods to Improve Retention - A Literature Review.
[13] https://www.classcentral.com/report/mooc-stats-2018/
[14] https://www.insidehighered.com/digital-learning/article/2019/01/16/study-offers-data-show-moocs-didnt-achieve-their-goals

***Compiled by***

Mr. Abdal Lalit is a student of B.Sc. (H) Computer Science II year at Shaheed Sukhdev College of Business Studies, University of Delhi. He can be reached at alalit@acm.org

Ms. Prashi Goyal is a student of B.Sc. (H) Computer Science II year at Shaheed Sukhdev College of Business Studies, University of Delhi. She can be reached at prashig@acm.org

Mr. Saksham Jain is a student of B.Sc. (H) Computer Science II year at Shaheed Sukhdev College of Business Studies, University of Delhi. He can be reached at sakshamjain2@acm.org.

***Under the guidance of***

Dr. Anamika Gupta, Associate life member of CSI (CSI: 00118185), Sr. Assistant Professor, S.S. College of Business Studies, University of Delhi. She has more than two decades of teaching and research experience in Machine learning, image processing, data science with several research publications in reputed conferences and journals. She can be reached at anamikargupta@sscbsdu.ac.in.

# Online Canteen Management System - powered by Digital Transactions

Dhanush Rajashekar, Prajwal H.C.,
Chandan.M.Bharadwaj, Anand Panduranga

## Introduction:

During breaks, there is a massive crowd in the college canteen. Much time is getting wasted standing in the queue at the coupon counter and service counter. Students and faculty get late for their lecture sessions, which is not conducive. We intend to provide a solution to this problem by the development of an online canteen management system, backed by digital transactions.

## Benefits:

Using the Canteen Management System, we can avoid the time typically wasted at the serving counter in the form of taking a limited number of orders at a time, while keeping the rest of members waiting for their turn. Also, one can have a facility for placing orders in advance so that his/her order is kept ready just at the particular time he/she chooses. The facility of payments via e-wallet saves the time typically spent in tendering exact change.

The main benefits of the project are:

1) Digital and Paper-free: The concept of paper cash is non-existent, and the payment method simplifies at both the customer and management end. The entire process becomes paper-free as no bills or tokens need to be issued.

2) Waiting Time: The variable time spent between ordering the item and receiving reduces with the issue of order numbers and slot concept.

3) Real-Time menu: The process of placing the order is digitalised, which simplifies the canteen management personnel task.

4) Notifications: The customers get an alert once their order is ready. They help in reducing the crowd at the food counter.

5) Tamper-proof: As unique order ID's get generated to each customer, tampering gets avoided. Also, accounting errors get eliminated with the entire billing getting digital.

6) Social Distancing: The entire process promotes social distancing as there is no room for casual crowd gathering to order their food.

## Software Requirements:

Languages used: HTML, CSS, Javascript, PHP

Applications used: 1) XAMPP server(MySQL and apache)

2)IDE's: Bracket and Atom

3) Web Browser: Google Chrome or Mozilla Firefox

The management system is observed from the Main-site and Admin-site perspectives for a well-rounded understanding.

*Main-site:*

1) Category Area: States the categories/cuisines available in the canteen.

2) Food Area: States the available food/menu for a given day.

3) Login/Register Area: Enables the user to either register or login(if already registered) for ordering the food.

*Admin-site:*

1) Category Area: Admin updates/changes the categories of the food available in the canteen, on a given day

2) Food Area: Admin updates the daily menu here. An additional option for adding/deleting items to/from the menu is available for the admin.

3) Orders dashboard: Admin receives the customer's orders, and the system parallelly generates a unique ID for each order.

*Operation:*

1) Login/Register:

The functionality of the website is supported totally by MySQL(PHP) Databases.

On clicking of the login button by the user, the login prompt appears where he/she enters the credentials, which gets stored in the USERS database. It verifies the credentials and allows the user to login.

2) Ordering Process:

A registered customer gets to select the cuisine and food item of his/her choice. Once the user places the order, he/she gets the unique order ID number for reference, which gets stored in the ORDERS database. The admin will also get a confirmation of the order in the Admin site.

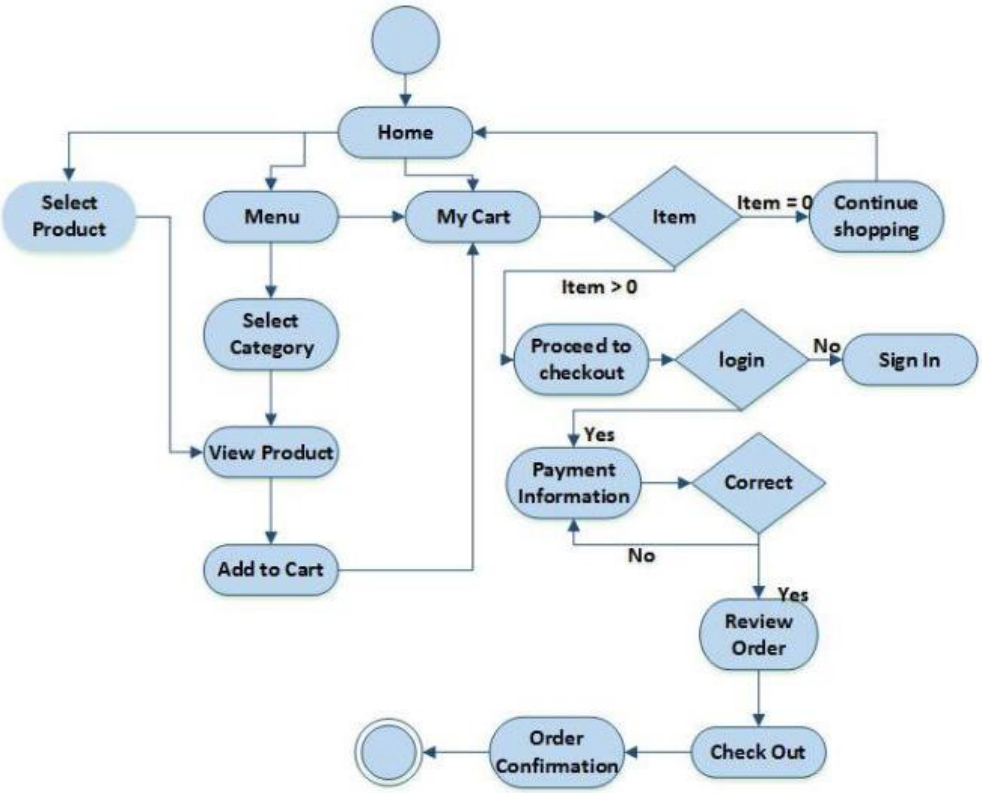



*Figure 1: Operation of Canteen Management System*

# Future Enhancements:

1) The menu automatically updates itself, based on the popular preferences and eating patterns of the customers, achieving a dynamic menu system.

2) Integration of the system onto an Android/iOS application.

3) Rewarding of the regular customers with points, which are redeemable on their future orders.

4) An effective real-time reviewing system, with provision to provide feedback on the quality, quantity, and taste of the food.

# Conclusion:

As the system optimises the overall process of canteen operation, an effective long-term solution is realised. In the current pandemic situation, this would be of valuable help to avoid the congregation of large crowds and also help maintain hygiene in the restaurant/canteen. Digital transactions also help students in the proper management of money through storage at a single location. This project is a step taken by the students for the benefit of their fellow student community.

*References:*

1) Swiggy Business Model | How does Swiggymake money - https://www.bstrategyhub.com
2) How Does Online Food Ordering System Work? - https://www.nibblematrix.com
3) How to Set Up Online Food Ordering for Restaurants in WordPress - https://www.youtube.com/watch?v=hBi_TZeOkW8
4) Web Development Roadmaps - https://www.w3schools.com
5) Galbo, N. R. (2019). Canteen service quality and student satisfaction. International Journal of Scientific & Technology Research, 8(6), 114-26.

*Compiled by:*

Dhanush Rajashekar is a 2nd-year student at Global Academy of Technology, Bengaluru. He can be reached at dhanushrajashekar1315@gmail.com.

Prajwal.H.C. is a 2nd-year student at Global Academy of Technology, Bengaluru. He He can be reached at hcprajwal43@gmail.com.

Chandan.M.Bharadwaj is a 2nd-year student at Global Academy of Technology, Bengaluru. He He can be reached at chandanbharadwaj007@gmail.com.

Under the guidance of:

Anand Panduranga (CSI: I1505715) is currently working as an Assistant Professor in the Department of Computer Science Engineering at Global Academy of Technology, Bengaluru. He can be reached at anand.panduranga@gat.ac.in.

# Accelerate code with OpenACC

**Aniket Zope, Sahil Sawant**
**Dhruv Kudale, Amit Joshi**

***Abstract:***
OpenACC is a sect of standardized, high-level pragmas that enables C/C++ and FORTRAN programmers to utilize parallel coprocessors. True to its name, OpenACC serves as an Accelerator Programming API. The term Accelerator Programming describes the general process of offloading functions from the CPU to the specialized hardware such as GPUs, and coprocessors to have better performance. Thus, while using OpenACC the basic approach is to insert special comments (directives) into the code so as to offload computation and parallelize the code at the level of GPU. This article introduces you to the world of OpenACC, important directives, aspects related to OpenMP, CUDA and so on. Further, examples using these directives are presented to guide you readers to kickstart development using OpenACC.

## 1. Introduction:

What is OpenACC? OpenACC was developed initially by PGI, Cray, CAPS enterprise, and NVIDIA with the purpose of providing a standard for accelerator directives [1]. OpenACC is a user-driven directive-based performance-portable parallel programming model. It is designed to port codes to a wide variety of heterogeneous platforms.

One benefit of OpenACC would be that it takes a higher-level approach, similar to that of OpenMP, where acceleration is achieved through a series of programmer directives, or pragmas. This extra level of abstraction brings many benefits, ranging from code- readability to generally faster code due to compiler optimizations.

An advantage of OpenACC over CUDA would be the support for other coprocessors other than NVIDIA GPUs. A wide variety of additional coprocessors, such as AMD GPUs and Intel MICs, are supported, consequently making OpenACC portable to most hosts / coprocessor combinations. As an aside, from the previous points it appears that OpenACC essentially provides a more “general” version of OpenMP.

OpenACC and MPI can favourably be combined with one another.

OpenACC and OpenMP can also be combined but the directives cannot mix, i.e. directives cannot be used in the same loop at the same time.

OpenACC and CUDA are fully interoperable. This has many benefits, as the programmer can develop using the OpenACC higher level directives, and implement the very computational complex functions manually using CUDA.

# 2. Setting up the Environment:

OpenACC compilers, profilers and debuggers are designed and available to download from multiple vendors and academic organizations [2]. The main purpose is to carry out the usage of the OpenACC directives in developing programs by getting the support of OpenACC in form of commercial compilers like from PGI, Cray and many more [3].

There are several ways to compile OpenACC programs. It can be compiled using gcc as well given the proper configuration. When compiling an OpenACC program, there are several dependencies that must be installed and located. You will need MPI, NVCC for compilation, and the location of the shared library lcudart.so for running your program.

For your convenience, the following snippet shows the commands that were used to set up the compilation. Your mileage may vary.

```
setenv  OMPCC /path_to_ompcc_compiler/
setenv  MPI /path_to_mpich_folder/
setenv  CUDA /path_to_cuda-5.5_folder/
setenv  PATH ${PATH}:${MPI}/bin:${OMPCC}/bin:${CUDA}/bin
setenv  LD_LIBRARY_PATH ${CUDA}/targets/x86_64-linux/lib
```

After installing the dependencies, you are good to go!! Depending upon the compiler support for OpenACC which you have, you can write your own programs having OpenACC directives which can be compiled with in different ways and executed on your system.

Basic OpenACC directives and examples of writing and compiling programs using OpenACC are elaborated upon in the further sections.

# 3. Basic OpenACC Directives:

The language of OpenACC is rather a set of descriptive directives which the programmer uses. This is done in order to parallelize a given portion of user written code [4]. After adding the required directives, the compiler figures out how to do it. Thus, with the help of OpenACC, programming on accelerators like GPU is made convenient and readable. Hence, it is possible for programmers to create an efficient parallel OpenACC code with only minor modifications to a serial CPU code [5].

What are compiler directives? They are special statements in specific regions of the code that tell the compiler or runtime to execute or generate parallel code as well as capable of allocating, copying and deallocating memory in and from the GPU.

OpenACC directive syntax: In C, OpenACC directives are specified in the following manner:

```
#pragma acc directive [clause [,] clause …]
//often followed by a structured code block
```

The kernels Construct: OpenACC allows us to use this directive to describe to the compiler a structured block of code to be run on an accelerator (GPU in this case) which the compiler will try to optimize if possible.

The loop Construct: This instruction tells the compiler to break up the loop into many smaller chunks, so that parallelism can be achieved. Nested for loops can also have this pragma applied to multiple loops. An important thing to note is that when used with the kernels construct, this makes the entire section into kernel code of GPU. Following snippet shows its usage:

```
#pragma acc kernels
 {
#pragma acc loop
for (...) {
...
}
}
```

The independent Construct: This construct used in conjunction with the loop construct can tell the compiler that loop iterations are data-independent and can be executed in parallel, overriding dependency analysis. We can apply this as shown in the following snippet:

```
#pragma acc loop independent
{
        for (…) {
//data independent code
}
```

The parallel Construct: This construct (much like from OpenMP's construct of same name) instructs the compiler to optimize a piece of code to produce an arbitrary number of kernels using arbitrary dimensions of grids and blocks. The way of applying this is same as in OpenMP.

The Memory Constructs: You can also specify the memory copying from device to host and vice-versa in an abstracted fashion. You can also choose to move memory independent of parallel sections by using the data construct. What's interesting is that you can also specify which elements to copy as well.

*A] Copy:* Upon executing the data copy pragma, we can handle memory operations with respect to GPU. Following snippet shows its usage followed by the explanation:

```
int matrix[10]
int n = 5;
//Fill the matrix
#pragma acc data copy (matrix [0: n]) {
...
}
```

After encountering the pragma, memory space for sliced number of ints is created on the device. The matrix, despite being 10 elements long, only has the first five elements (because of the slicing) copied into the device (GPU) from the host (CPU). At the end, the device copies the memory back into the host. The space is then freed from the device. Along with data, this can be used with kernel and parallel sections also!

*B] Create:* If you were to replace all instances of copy with create in the previous example, the program would allocate space on the device only. No memory gets copied in or out.

*C] Copyin*: If you were to replace all instances of copy with copyin in the previous example, the program would allocate space on the device, and like the name suggests, copy in the memory from host to device.

*D] Copyout*: Allocates space on the device, and transfers the data in the space to the host after the end of the pragma block. This is useful for when you know the dimensions of the output, and can write the final values of the memory immediately.

Along with these constructs, many others can be combined as well in order to write and execute desired parallel code on the accelerators. Similarly, NVIDIA support is available too [6].

# 4. An OpenACC Program:

This section introduces you to few examples with which you can start programming using OpenACC.

Vector Addition: The example shows a simple vector addition program to be executed in a parallel manner on a GPU using OpenACC clauses [7]. In the following snippet, a function addVectors is shown.

```
void addVectors (int n, int *a, int *b, int *c) {
// sum component wise and save result into vector c
   #pragma acc kernels copyin (a [0: n], b [0: n]), copyout (c [0: n])
   for (int i = 0; i < n; i++) {
      c[i] = a[i] + b[i];
   }
}
```

As studied previously, the pragma clause defined in OpenACC has the kernels construct followed by copyin and copyout clauses for memory management. The kernels keyword tells the compiler that the following block of code will be executed in the GPU. The copyin will copy input vectors i.e. the data of a and b in the device. After parallel addition, the result vector will be copied back to the host form the device in vector c.

Vector SAXPY: Following is the function that computes SAXPY i.e. Single Precision a.X + Y where ‘a’ is a scalar and X and Y are vectors [8]. Following is the function that will carry out the task:

```
void saxpy (int n, float a, float *x, float *restrict y)
  {
    #pragma acc parallel loop
    for (int i = 0; i < n; i++)
      y[i] = a * x[i] + y[i];
  }
```

As seen in the above example, the OpenACC parallel directive tells the compiler to mark the parallel region and optimize the execution. The loop clause will parallelize the iterations in the for loop to run across parallel gangs as there are no data dependencies in the loop body. Thus, with a single line of OpenACC pragma, we can run the same code and enhance portability on many GPUs without much effort.

*Compiling and execution:* Once the C/C++ program with desired OpenACC extension is saved, you can compile and execute the program. Depending upon the compiler support available for OpenACC, the command for compilation differs. For PGI support (CUDA toolkit from NVIDIA is a dependency required for it as mentioned in previous section), you can compile and execute in the following manner:

```
//Compile
cc -acc <program name> -o <executable name>
//Execute
./<executable name>
```

Thus, with this you will be successfully able to carry out vector addition on an accelerator like GPU by incorporating the use of OpenACC.

# 5. Conclusion:

Accelerated computing is providing one of the most exciting discoveries today. For researchers who seek faster application performance, OpenACC is a directive-based programming model designed to provide a simple yet powerful approach to accelerators. OpenACC helps to achieve performance goals without much programming effort. Apart from that, OpenACC will assist single source code to deliver performance portability across various platforms.

*References:*

[1] OpenACC Overview: https://www.openacc.org/
[2] OpenACC Download and Tools https://www.openacc.org/tools
[3] OpenACC Compiler Supports https://en.wikipedia.org/wiki/OpenACC
[4] OpenACC Tutorial for programming https://docs.computecanada.ca/wiki/OpenACC_Tutorial
[5] Introduction to OpenACC https://www.bu.edu/tech/files/2017/04/OpenACC-2017Spring.pdf
[6] OpenACC Constructs https://developer.nvidia.com/openacc
[7] OpenACC for Vector Addition https://www.olcf.ornl.gov/tutorials/openacc-vector-addition/
[8] OpenACC in practice https://homepages.dcc.ufmg.br/~douglas/research/openacc/index.html

***Compiled by:***

Mr. Aniket Zope, student of First Year MTech (Computer Engineering) at College of Engineering, Pune (COEP), Maharashtra. He can be reached at : zopeas19.comp@coep.ac.in

Mr. Sahil Sawant, student of First Year MTech (Computer Engineering) at College of Engineering, Pune (COEP), Maharashtra. He can be reached at : sawantsj19.comp@coep.ac.in

Mr. Dhruv Kudale, student of Final Year BTech (Computer Engineering) at College of Engineering, Pune (COEP), Maharashtra. He can be reached at: kudaledd16.extc@coep.ac.in

***Under the guidance of:***

Mr. Amit Joshi [CSI: 01180301], is working as Assistant Professor in Dept. of Computer Engineering and IT, College of Engineering, Pune (COEP), Maharashtra. He can be reached at : adj.comp@coep.ac.in

# GPU Computing using OpenCL

Dhruv Kudale, Amit Joshi

Abstract:
OpenCL is a framework designed to write programs that enable heterogeneous computing. It defines a C like language which helps to execute programs on different platforms including but not limited to GPUs. OpenCL also provides a standard interface for parallel computing. This article helps to assist students who intend to begin programming with OpenCL in Windows OS to use their AMD GPUs supported by Visual Studio as a standard IDE for developing relevant projects. Along with the guidance to get your Windows machine ready for OpenCL applications, it also covers basic kernel writing and sample OpenCL program snippets for vector addition in order to get started.

## 1. Introduction:

The Open Computing Language aka OpenCL is a widely deployed and open standard used for parallel and cross platform programming. It is implemented by the Khronos Group [1]. It is a framework designed as a C/C++ like language used to involve computing with different multicore CPUs, GPUs, cloud servers, DSPs, mobile devices, supercomputers, etc. OpenCL helps to unite diverse platforms and is thus evolved to meet the growing standards of computing industry.

OpenCL greatly improves the speed and responsiveness of applications involving different platforms and parallel computing. The benefits of OpenCL computing can be seen in a wide range of industries including scientific research, medical imaging, digital signal processing as well as deep learning. The simplicity of the model and the fact that OpenCL provides systematic and well-defined API calls to compile and load programs on other hardware devices encourages a wide class of users to use OpenCL. Consideration of problem sizes, compatibility for different kinds of devices and host device optimization paves a way for the usage of OpenCL.

OpenCL in Windows Having necessary drivers and SDK installed for the required platform like GPU along with carrying out proper linking and project configuration on a convenient IDE or Editor makes using OpenCL in Windows pretty easy and straightforward. The usage of OpenCL with Visual Studio in Windows along with preliminary coding required to kickstart your interests is what we will explore in the following sections of the article.

## 2. Setting up the Environment:

There is a depiction of the basic steps which involve getting the necessary drivers and SDKs, setting up the IDE or Editor and further configuration of the project to link and detect OpenCL library in the application [2, 3]. Following steps will guide you to get your system ready to work with OpenCL:

Drivers and SDK: Considering the fact that you have to start AMD GPU programming using OpenCL, most of the stuff already works! You will have AMD drivers preinstalled which you can check from AMD Radeon Settings System Overview [4]. If not available the drivers can be manually installed from [5, 6, 7]. Apart from that you will need OpenCL SDK which is available as a Zip folder where necessary library functions are included as a part of OCL_SDK_Light [8]. With these drivers installed, the first step is cleared.

Setting up the Editor or IDE: Install the latest version of Visual Studio Code with systematic steps, compatible for your Windows system and get acquainted to the interface of the IDE [9]. Start a new project as C/C++ application and proceed with the next step.

Project Configuration In this last step we need to tell the compiler and linker about the whereabouts of the OpenCL SDK. From the project explorer view, right click on the project and select project properties and make the following changes highlighted in orange:

Specify the include file path in the C/C++ compiler as a part of additional include directories up to the include folder from the SDK installed

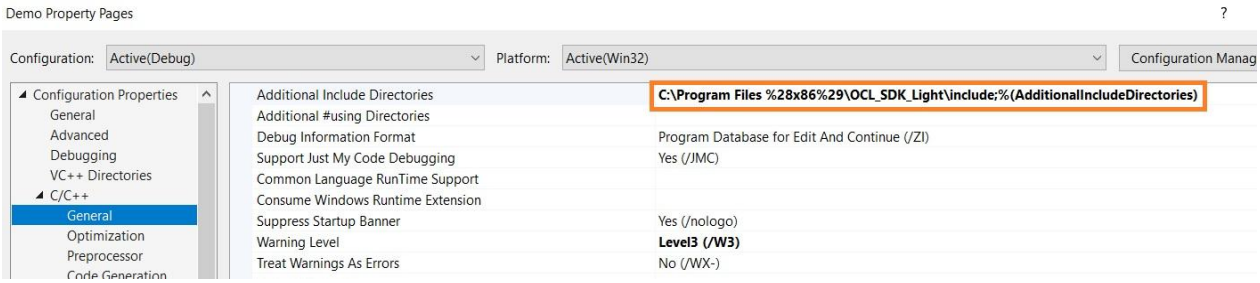



Fig 1: Project Properties Compiler Settings

Specify the necessary file paths (x86 or x86_64) from the installed SDK for the linker tab Mention “opencl.lib” as the additional dependencies in the linker tab of properties.

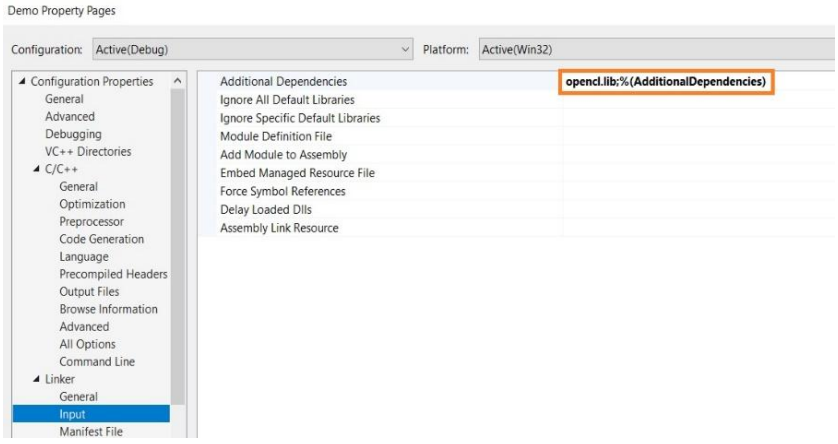



Fig 2: Project Properties Linker Settings

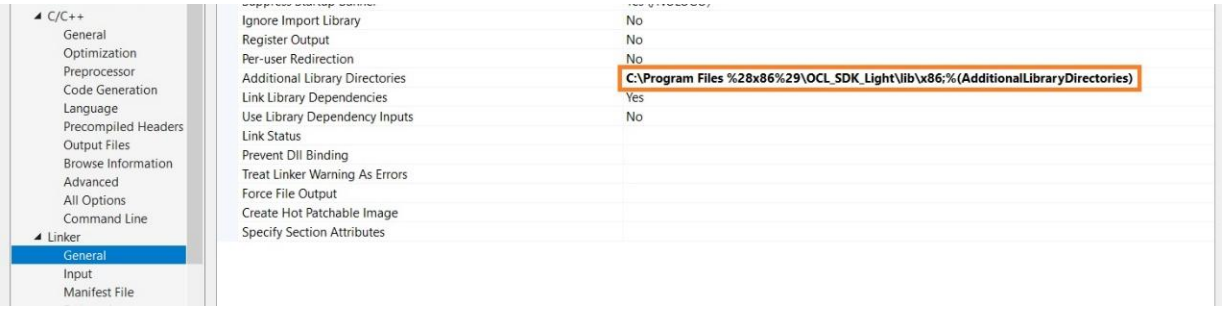



Fig 3: Project Properties Additional Dependencies

And with this setup you are good to go!! You will be able to write OpenCL programs with having the functions and usages detected by the IDE. You can build and run your own OpenCL programs. The next section helps you to get started on an OpenCL program of Vector addition to carry out on an AMD GPU and compare the stats with the corresponding CPU execution.

# 3. The Problem Statement

After carrying out above configuration steps, you are free to write detectable OpenCL code. The problem statement of vector addition is chosen to get a clarity on the OpenCL programming model [10]. Adjacent Fig 4 gives us the schematic idea for carrying out vector addition. The CPU is the host and the GPU is the device where we execute the desired functionality. In order for the GPU (device) to work in a parallel manner on data, the host data needs to be copied to the device buffers. The problem addressed is to carry out parallel addition of corresponding elements of two large vectors which is done by the kernel in the device. Simultaneous additions, computing the result and improving the performance is the main goal.

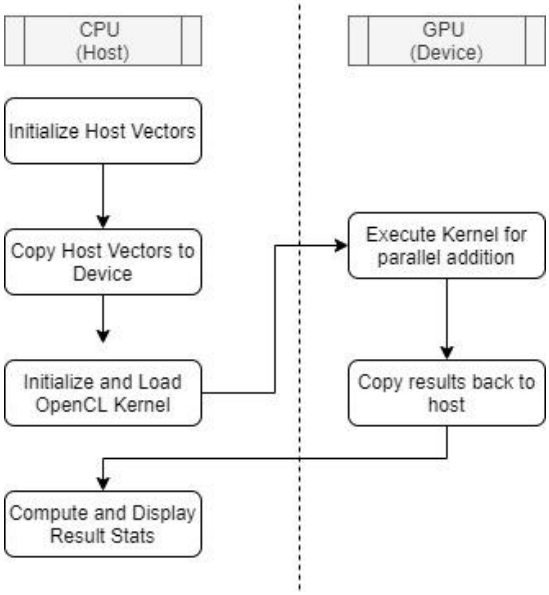



Fig 4: Schematic Flow of OpenCL Program

# 4 The Program:

The program follows the schematic flow as depicted in Fig 4 to carry out Vector Addition in a parallel manner on a GPU to enhance the performance [11].

Declaring host vectors of size ‘n’:

```
size_t bytes = n * sizeof(int);
h_a = (int*)malloc(bytes);
h_b = (int*)malloc(bytes);
h_c = (int*)malloc(bytes);
```

Creating device buffers by the clCreateBuffer () function and copying the data of host vectors into the device buffer by clEnqueueWriteBuffer () function:

```
    d_a = clCreateBuffer(context, CL_MEM_READ_ONLY, bytes, NULL, NULL);
        d_b = clCreateBuffer(context, CL_MEM_READ_ONLY, bytes, NULL, NULL);
        d_c = clCreateBuffer(context, CL_MEM_WRITE_ONLY, bytes, NULL, NULL);
        //Copying data
    err = clEnqueueWriteBuffer(queue, d_a, CL_TRUE, 0, bytes, h_a, 0, NULL, NULL);
        err = clEnqueueWriteBuffer(queue, d_b, CL_TRUE, 0, bytes, h_b, 0, NULL, NULL);
```

The Kernel: This is a function written and called by the host and executed in the device. In this case, the parallel addition of vector elements take place. Each element is identified by the thread id which is returned by the get_global_id (0) function as seen in the following snippet:

```
__kernel void vecAdd(  __global double *a, __global double *b, __global double *c,
                const unsigned int n)   {
  //Get our global thread ID
  int id = get_global_id(0);

  //Make sure we do not go out of bounds
  if (id < n)
    c[id] = a[id] + b[id];
}
```

The parameters of the kernels are loaded and kernel is executed in the following manner:

```
    err = clSetKernelArg(kernel, 0, sizeof(cl_mem), &d_a);
        err = clSetKernelArg(kernel, 1, sizeof(cl_mem), &d_b);
        err = clSetKernelArg(kernel, 2, sizeof(cl_mem), &d_c);
        err = clSetKernelArg(kernel, 3, sizeof(unsigned int), &n);

        // Execute the kernel
        err = clEnqueueNDRangeKernel(queue, kernel, 1, NULL, &globalSize, &localSize, 0,    NULL,
NULL);
```

After execution of the kernel, the contents of the result vector are copied from device to the host by usage of the clEnqueueReadBuffer () function:

```
clEnqueueReadBuffer(queue, d_c, CL_TRUE, 0, bytes, h_c, 0, NULL, NULL);
```

The details of all different kinds of inbuilt OpenCL functions are mentioned in the OpenCL reference pages [12].

# 5. The Output and Stats:

Both GPU and CPU computation is carried out and timed. Following is the snapshot of the output for vector addition, the size of each vector being 1 million.

The speedup in this case is 3. The speedup increases and more benefit is seen when length of vector increases!

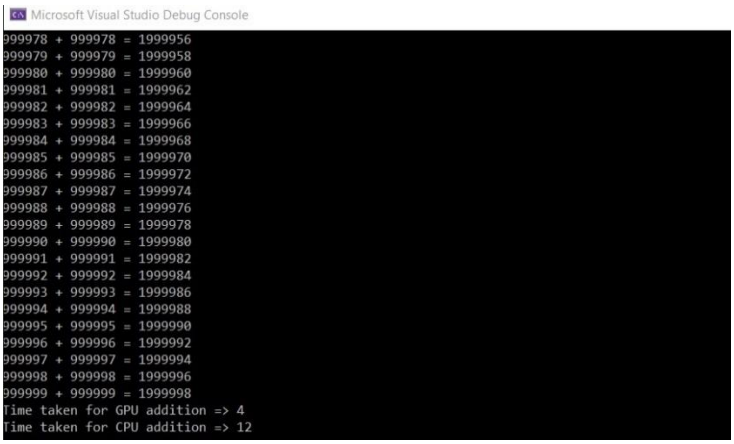



*Fig 5: Snapshot of the Output of OpenCL Vector Addition Program*

### OpenCL vs CUDA:

Framework Type CUDA is a parallel computing proprietary framework created and implemented by NVIDIA whereas OpenCL as the name suggests is an open framework.

Support and Compatibility CUDA works on limited hardware as CUDA enabled GPUs are only available from NVIDIA. On the other hand, OpenCL works on wide variety of heterogeneous platforms including but not limited to multicore CPUs, GPUs, DSPs, etc of NVIDIA, AMD, etc.

Performance The observation of comparing performances of similar kernels of CUDA and OpenCL shows that CUDA performs better and faster by a small edge in data transfer and kernel execution than OpenCL [13]. However, OpenCL does remain a fair and considerable alternative to CUDA.

Both CUDA and OpenCL have a similarly designed programming model and both continue to grow. They have a fairly a comparable performance with respect to one another with their share of pros and cons. CUDA must be the choice for faster and tighter performance for CUDA enabled hardware, whereas OpenCL should be used for portability and better integration support.

# 6. Conclusion:

Compatible drivers, OpenCL SDK, convenient IDE and proper linking and configuration can make OpenCL programming in Windows pretty easy. A systematic approach as mentioned above can assist you in making phenomenal OpenCL applications that address complex problem statements. There is a lot of support and a collection of reference pages that enable to proceed with OpenCL programming in a suitable and comfortable environment in Windows. OpenCL has better integration support and is considerably fast and responsive. As compared with CUDA, the OpenCL framework spans across a large number of platforms. The framework API and the programming model is intuitively easy to understand and work on.

*References:*

[1] OpenCL Overview: The Open Standard for Parallel Programming of Heterogeneous Systems https://www.khronos.org/opencl/
[2] OpenCL on Visual Studio : Configuration tutorial for the confused https://medium.com/@pratikone/opencl-on-visual-studio-configuration-tutorial-for-the-confused-3ec1c2b5f0ca
[3] Installing OpenCL on Windows https://streamhpc.com/blog/2015-03-16/how-to-install-opencl-on-windows/
[4] How to Find Driver Information and Check for Updates Within Radeon Settings https://www.amd.com/en/support/kb/faq/dh-016
[5] OpenCL Runtimes for Intel Processors https://software.intel.com/en-us/articles/opencl-drivers
[6] AMD Drivers and Support https://www.amd.com/en/support
[7] NVIDIA Driver Downloads https://www.nvidia.com/Download/index.aspx
[8] OCL_SDK_Light installation https://github.com/GPUOpen-LibrariesAndSDKs/OCL-SDK/releases
[9] Microsoft Visual Studio Full-featured integrated development environment (IDE) for Android, iOS, Windows, web, and cloud, https://visualstudio.microsoft.com/downloads/
[10] Programming model https://leonardoaraujosantos.gitbooks.io/opencl/content/chapter1.html
[11] OpenCL Vector Addition Program https://www.olcf.ornl.gov/tutorials/opencl-vector-addition/
[12] OpenCL reference pages https://www.khronos.org/registry/OpenCL/sdk/1.0/docs/man/xhtml/
[13] A Comprehensive Performance Comparison of CUDA and OpenCL, Jianbin Fang, Ana Lucia Varbanescu and Henk Sips, Delft University of Technology, Delft, the Netherlands

Compiled by:

Mr. Dhruv Kudale, student of Final Year BTech (Computer Engineering) at College of Engineering, Pune (COEP), Maharashtra. He can be reached at : kudaledd16.extc@coep.ac.in

Under the guidance of:

Mr. Amit Joshi [CSI :01180301] is working as Assistant Professor in Dept. of Computer Engineering and IT, College of Engineering, Pune (COEP), Maharashtra. He can be reached at adj.comp@coep.ac.in

# A Prospective Study on Ethical Hacking

**Shaik. Azeem, P.Vijay Kumar**
**Shaik.Aftab, N Md Jubair Basha**

Abstract:
Now-a-days, Ethical Hacking and Ethical Hacker are terms used to describe hacking performed by a company or individual to help identify potential threats on a computer or network. An ethical hacker attempts to bypass system security and search for any weak points that could be exploited by malicious hackers. This information is then used by the organization to improve the system security, to minimize or eliminate any potential attacks .The main reason is to clarify the misunderstanding going around the ethical hacking and to provide a proper ideology about ethical hacking. This article presents the brief overview, respective rules to be followed, various types, different phases and tools for ethical hacking. A broad study on ethical hacking has been carried out in this article.

# Introduction :

Ethical Hacking refers to the act of locating weaknesses and vulnerabilities of computer and information systems by replicating the intent and actions of malicious hackers. It is also known as penetration testing, intrusion testing or red teaming. Lots of people think that ethical hacking and ethical hackers are responsible for all the cyber attacks ,but the ethical hackers are not responsible for this. All the ethical hackers work for the organizations to find bugs and report them to the developers of software. Many organizations faced the problem of data breaches, security problems caused by hackers, in order to be protected from hackers ethical hackers are hired. Latest survey tells that only 38% of global organizations are prepared for cyber attacks, so it becomes important for more and more ethical hackers.

## Rules to be followed :

For hacking to be deemed ethical, the hacker must obey the following rules:

Expressed (often written) permission to probe the network and attempt to identify potential security risks.

You respect the individual's or company's privacy.

You close out your work, not leaving anything open for you or someone else to exploit at a later time.

You let the software developer or hardware manufacturer know of any vulnerabilities you locate in their software or hardware, if not already known by the company.

Ethical hackers should have sound knowledge about computer networks.

We have to follow laws made by different lands.

### Pros of Hacking :

Hacking is quite useful in the following scenarios −

- To recover lost information, especially in case you lost your password.
- To perform penetration testing to strengthen computer and network security.
- To put adequate preventative measures in place to prevent security breaches.
- To have a computer system that prevents malicious hackers from gaining access.

### Cons of Hacking :

Hacking is quite dangerous if it is done with harmful intent. It can cause −

- Massive security breach.
- Unauthorized system access on private information.
- Privacy violation.
- Hampering system operation.
- Denial of service attacks.
- Malicious attack on the system.

## Types of Ethical Hacking:

It is no big secret that any system, process, website, device, etc., can be hacked. In order to understand how the hack might happen and what the damage could be, ethical hackers must know how to think like malicious hackers and know the tools and techniques they are likely to use. Some of the types of ethical hacking are

- Web applications hacking
- System hacking
- Web server hacking
- Hacking wireless networks
- Social engineering

## Types of Hackers:

Hackers can be differentiated based on their intent of hacking the systems. There are generally 10-types of Hackers, they are:

*White Hat Hackers:* White hat hackers are the one who is authorized or the certified hackers who work for the government and organizations by performing penetration testing and identifying loopholes in their cybersecurity. They also ensure the protection from the malicious cyber crimes. They work under the rules and regulations provided by the government, that's why they are called Ethical hackers or Cybersecurity experts.

*Black Hat Hackers:* They are often called Crackers. Black Hat Hackers can gain the unauthorized access of your system and destroy your vital data. The method of attacking they use common hacking practices they have learned earlier. They are considered to be as criminals and can be easily identified because of their malicious actions.

*Gray Hat Hackers*: Gray hat hackers fall somewhere in the category between white hat and black hat hackers. They are not legally authorized hackers. They work with both good and bad intentions; they can use their skills for personal gain. It all depends upon the hacker. If a gray hat hacker uses his skill for his personal gains, he/she is considered as black hat hackers.

*Script Kiddies:* They are the most dangerous people in terms of hackers. A Script kiddie is an unskilled person who uses scripts or downloads tools available for hacking provided by other hackers. They attempt to attack computer systems and networks and deface websites. Their main purpose is to impress their friends and society. Generally, Script Kiddies are juveniles who are unskilled about hacking.

*Green Hat Hackers:* They are also amateurs in the world of hacking but they are bit different from script kiddies. They care about hacking and strive to become full-blown hackers. They are inspired by the hackers and ask them few questions about. While hackers are answering their question they will listen to its novelty.

*Blue Hat Hackers:* They are much like the script kiddies; are beginners in the field of hacking. If anyone makes angry a script kiddie and he/she may take revenge, then they are considered as the blue hat hackers. Blue Hat hackers payback to those who have challenged them or angry them. Like the Script Kiddies, Blue hat hackers also have no desire to learn.

*Red Hat Hackers:* They are also known as the eagle-eyed hackers. Like white hat hackers, red hat hackers also aims to halt the black hat hackers. There is a major difference in the way they operate. They become ruthless while dealing with malware actions of the black hat hackers. Red hat hacker will keep on attacking the hacker aggressively that the hacker may know it as well have to replace the whole system.

*State/Nation Sponsored Hackers*: State or Nation sponsored hackers are those who are appointed by the government to provide them cybersecurity and to gain confidential information from other countries to stay at the top or to avoid any kind of danger to the country. They are highly paid government workers.

*Hacktivist:* These are also called the online versions of the activists. Hacktivist is a hacker or a group of anonymous hackers who gain unauthorized access to government's computer files and networks for further social or political ends.

*Malicious Insider or Whistleblower*: A malicious insider or a whistleblower could be an employee of a company or a government agency with a grudge or a strategic employee who becomes aware of any illegal activities happening within the organization and can blackmail the organization for his/her personal gain.

# Phases of Ethical Hacking:

There are 5 phases of ethical hacking


*Figure 1: Phases of ethical hacking*

*1.Reconnaissance:* Reconnaissance, also known as the preparatory phase, is where the hacker gathers information about a target before launching an attack and is completed in phases prior to exploiting system vulnerabilities. One of the first phases of Reconnaissance is dumpster diving. It is during this phase that the hacker finds valuable information such as old passwords, names of important employees (such as the head of the network department) and performs an active reconnaissance to know how the organization functions. As a next step, the hacker completes a process called foot printing to collect data on the security posture, reduces the focus area such as finding out specific IP addresses, identifies vulnerabilities within the target system, and finally draws a network map to know exactly how the network infrastructure works to break into it easily. Footprinting provides important information such as the domain name, TCP and UDP services, system names and passwords. There are also other ways to do footprinting, including impersonating a website by mirroring it, using search engines to find information about the organization and even using information of current employees for impersonation.

During reconnaissance, an ethical hacker attempts to gather as much information about a target system as possible, following the seven steps listed below −

- Gather initial information
- Determine the network range
- Identify active machines
- Discover open ports and access points
- Fingerprint the operating system
- Uncover services on ports
- Map the network

Reconnaissance takes place in two parts –

a)Active Reconnaissance and b)Passive Reconnaissance

*a)Active Reconnaissance*

In this process, you will directly interact with the computer system to gain information. This information can be relevant and accurate. But there is a risk of getting detected if you are planning active reconnaissance without permission. If you are detected, then system admin can take severe action against you and trail your subsequent activities.

*b)Passive Reconnaissance*

In this process, you will not be directly connected to a computer system. This process is used to gather essential information without ever interacting with the target systems.

*2.Scanning*: In this phase, the hacker identifies a quick way to gain access to the network and look for information. There are three methods of scanning: pre-attack, port scanning/sniffing and information extraction. Each of these phases demonstrates a specific set of vulnerabilities which the hacker can utilize to exploit the system weaknesses. The pre-attack phase is where the hacker scans the network for specific information based on the information gathered during reconnaissance. The port scanner or sniffing phase is where scanning includes the use of diallers, port scanners, vulnerability scanners and other data-gathering equipment. The information extraction phase is where the attackers collect information about ports, live machines and OS details to launch an attack.

The purpose of each scanning process is given below:

- Port Scanning – detecting open ports and services running on the target.
- Network Scanning – IP addresses, Operating system details, Topology details, trustedrouters information etc
- Vulnerability scanning – scanning for known vulnerabilities or weakness in a system.

Scanning Methodology:

Check for Live Systems: Ping scan checks for the live system by sending ICMP echo request packets. If a system is alive, the system responds with ICMP echo reply packet containing details of TTL, packet size etc.

Check for Open Ports: Port scanning helps us to find out open ports, services running on them, their versions etc. Nmap is the powerful tool used mainly for this purpose.

We have various types of scan:

1.Connect scan: Identifies open ports by establishing a TCP handshake with the target.TCP Connect Scan is presented in Figure 2, the source is https://www.safaribooksonline.com

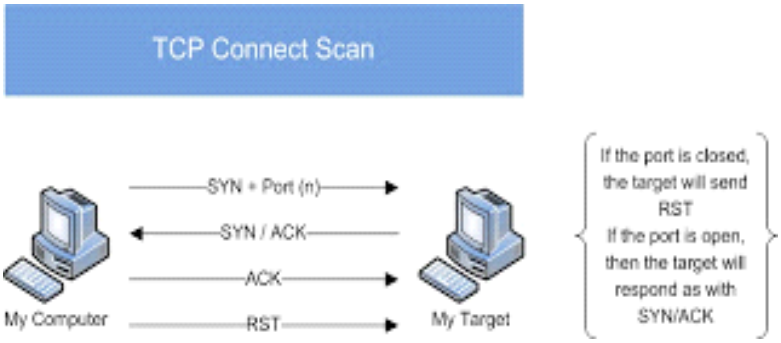



*Figure 2: TCP Connect Scan*

The respective is as follows:

Nmap command: nmap -sT -v -p- <TargetIP>

2.Half-open scan:It is otherwise known as Stealth scan used to scan the target in a stealthy way by not completing the TCP handshake by abruptly resetting the communication.The Half-open Scan is represented in Figure 3 the source is also from https://www.safaribooksonline.com

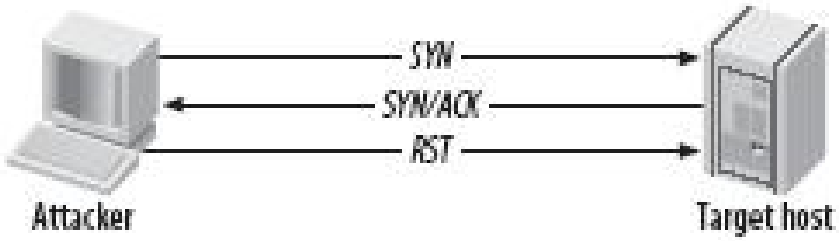



*Figure 3: Half-open Scan*

The respective is as follows:

Nmap command: nmap -sS -v <TargetIp>

3.XMAS scan: This is also called as inverse TCP scanning. This works by sending packets set with PSH, URG, FIN flags. The targets do not respond if the ports are open and send a reset response if ports are closed.Figure 4 presents the XMAS Scan. The source of this figure is from https://www.information-security.fr


*Figure 4: XMAS Scan*

4.FIN scan: Fin flag is set in the TCP packets sent to the target. open ports doe does not respond while closed ports send a reset response. Figure 5 presents the FIN Scanthe source of this figure from https://securitcrs.wordpress.com


*Figure 5: FIN Scan*

The respective is as follows:

Nmap command: nmap -SF <targetIp>

5.ACK scan: Here the attacker sets the ACK flag in the TCP header and the target's port status is gathered based on window size and TTL value of RESET packets received from the target. Figure 6 presents the ACK Scan and source of this figure is from https://www.hackingloops.com


*Figure 6: ACK Scan*

The respective is as follows:

Nmap command: nmap -SA -v <targetip>

6.Null Scan: Works by sending TCP packets with no flags set to the target. Open ports do not respond while closed ports respond with a RESET packet. Figure 7 presents the NullScan.

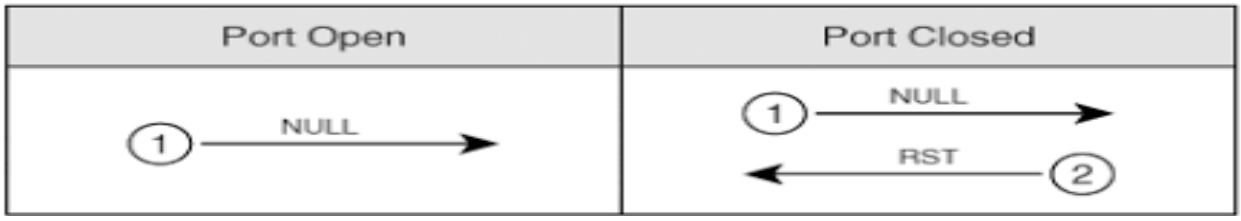



Figure 7: Null Scan

The respective is as follows:

Nmap Command: nmap -sN -p- <targetIP>

7.Idle Scan: Here the attacker tries to mask his identity uses an idle machine on the network to probe the status details of target ports. Figure 8 presents the Idle Scan and the source of this figure is from https://en.wikipedia.org/wiki/Idle_scan

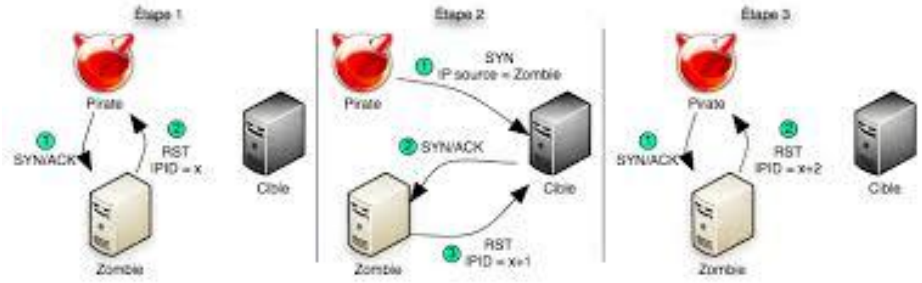



Figure 8: Idle Scan

The respective is as follows:

Nmap command: nmap -Pn -sI ZombieIp TargetIp

Draw Network Diagrams

With the information gathered, the attacker can come up with a network diagram which might give him information about network and architecture of the target organization helping him to identify the target easily,

Tools: Network View, Opmanager etc

Prepare Proxies

Proxies can use to maintain the anonymity of the attacker by masking the IP address. It can capture information passing through it since it acts as an intermediary between client and server and the attacker can access the resources remotely using the proxies.

e.g.: TOR browsers, Onion sites etc, Proxify, Psiphon etc

Countermeasures

- Configure IDS and firewall to block probes.
- Keep firewall, routers, IDS firmware update
- Run port scanners to verify the security of the target.
- Add rules in firewall restricting access to ports.
- Disable ICMP based scanning at firewall

3.Gain Access: The hacker gains access to the system, applications, and network, and escalates their user privileges to control the systems connected to it.

Password Cracking:

There are few basic methods of password cracking:

Brute-force: Trying all possible combinations until the password is cracked.

Dictionary attack: This is a compiled list of meaningful words, compared against the password field till a match is found.

Rule based attack: If some details about the target are known, we can create rules based on the information we know.

Rainbow table: Instead of comparing the passwords directly, taking the hash value of the password, comparing them with a list of pre-computed hash values until a match is found.

Rainbow table method gives an advantage to the attacker since no account lockout is enabled for wrong hashes against the password. To prevent rainbow table attack, salting can be used. Salting is a process of adding random numbers to the password so the attacker will not be able to crack the hash without that salt added.

Types of Password Attacks :

a) Passive online attacks

A passive attack is an attack on a system that does not result in a change to the system in anyway

The attack is to purely monitor or record data.

- Wire Sniffing
- Man in the middle
- Replay attack

b) Active online attacks

An active online attack is the easiest way to gain unauthorized administrator-level access to the system

- Password guessing
- Trojan/spyware/keyloggers

- Hash injection
- Phishing

c)Offline attacks

Offline attacks occur when the intruder checks the validity of the passwords. Offline attacks are often time to consume.

- Pre-computed hashes
- Distributed Network
- Rainbow

d)Non-electronic attacks

Non-electronic attacks are also known as non-technical attacks. This kind of attack doesn't require any technical knowledge about the methods of intruding into another system.

- Social engineering
- Shoulder surfing
- Dumpster Diving
- How to defend against password cracking:
- Don't share your password with anyone
- Do not use the same passwords during password change
- Enable security auditing to help monitor and track password attack
- Do not use clear text protocols and protocols with weak encryption
- Set the password change policy to 30 days
- Monitor the server's logs for brute force attacks on the user's accounts
- Avoid storing passwords in an unsecured location
- Never use passwords such as date of birth, spouse, or child's or pet's name
- Enable SYSKEY with the strong password to encrypt and protect the SAM database
- Lockout an account subjected to too many incorrect password guesses.

## Escalation of Privileges:

An attacker can gain access to the network using a non-admin user account, and the next step would be to gain administrative privilege.

There are two types of Privilege Escalation:

Horizontal Privilege Escalation occurs when a malicious user attempts to access resources and functions that belong to peer users, who have similar access permissions.

Vertical Privilege Escalation occurs when a malicious user attempts to access resources and functions that belong to a user with higher privileges, such as application or site administrators.

### Executing Applications:

Intruder executes malicious applications after gaining administrative privileges so they can run malicious programs remotely, to capture all sensitive data, crack passwords, capture screenshots or to install a backdoor.

Tool: RemoteExec, PDQ Deploy, DameWare NT Utilities

Keylogger: keystroke loggers are programs or hardware devices that monitor each keystroke a user types on a keyboard, logs onto a file, or transmits them to a remote location. keyloggers are placed between the keyboard hardware and the OS

A key logger can

- Record each keystroke
- capture screenshots at regular intervals of time showing user activity such as when he or she types a character or click a mouse button
- Track the activities of users by logging window titles, names of launched applications and other information
- monitor online activity of users by recording addresses of the websites that they are have visited and with the keywords entered by them
- record all the login names, bank and credit card numbers and passwords including hidden passwords or data that are in asterisk or blank spaces
- record online chat conversion

Types of Keylogger

- Hardware Keylogger
- Software Keylogger

Spyware: Spyware is stealthy computer monitoring software that allows you to secretly record all activities of a computer user.

Trojans :It is a malicious modification of a software program or hardware component. The hardware Trojan is occurred at integrated circuit chip of a system.

4.Maintaining Access: Here, the hacker secures access to the organizations, Rootkits, Trojans and uses it to launch additional attacks on the network. By hiding files in attacker system hacker can maintain access to the attacker system when ever hacker wanted.

Rootkits: Rootkits are programs that hackers use in order to evade detection while trying to gain unauthorized access to a computer. Rootkits when installing on a computer, are invisible to the user and also take steps to avoid being detected by security software.

A rootkit is a set of binaries, scripts and configuration files that allows someone to covertly maintain access to a computer so that he can issue commands and scavenge data without alerting the system's owner.

Depending on where they are installed there are various types of rootkits:

- Kernel Level Rootkits
- Hardware/Firmware Rootkits
- Hypervisor (Virtualized) Level Rootkits
- Boot loader Level (Bootkit) Rootkits

NTFS DATA Stream:

Alternative Data Stream support was added to NTFS (Windows NT, Windows 2000 and Windows XP) to help support Macintosh Hierarchical File System (HFS) which uses resource forks to store icons and other information for a file. Using Alternative Data Streams a user can easily hide files that can go undetected unless close inspection.

Steganography: The art of hiding a data inside another data/medium is called steganography.

e.g.: hiding data within an image file

The secret message is called overt file and the covering file is called covert file.

Types of Steganography:

- Image Steganography
- Document Steganography
- Folder Steganography
- Video Steganography
- Audio Steganography
- White Space Steganography

5.Clearing Tracks: Once the hacker gains access, they cover their tracks to escape the security personnel. They do this by clearing the cache and cookies, tampering the log files and closing all the open ports. This step is important because it clears the system information making hacking a great deal harder to track.

Once an attacker finishes his work, he wants to erase all tracks leading the investigators tracing back to him. This can be done using

- Disable auditing&Clearing logs.
- Modifying logs, registry files.
- Removing all files, folders created.

# Tools used for Hacking:

NMAP: Nmap stands for Network Mapper. It is an open source tool that is used widely for network discovery and security auditing. Nmap was originally designed to scan large networks, but it can work

equally well for single hosts. Network administrators also find it useful for tasks such as network inventory, managing service upgrade schedules, and monitoring host or service uptime.

Nmap uses raw IP packets to determine –

- what hosts are available on the network,
- what services those hosts are offering,
- what operating systems they are running on,
- what type of firewalls are in use, and other such characteristics.

Nmap runs on all major computer operating systems such as Windows, Mac OS X, and Linux.

Metasploit: Metasploit is one of the most powerful exploit tools. It's a product of Rapid7 and most of its resources can be found at: www.metasploit.com. It comes in two versions –

commercial and free edition. Metasploit can be used with command prompt or with Web UI.

With Metasploit, you can perform the following operations –

- Conduct basic penetration tests on small networks
- Run spot checks on the exploitability of vulnerabilities
- Discover the network or import scan data
- Browse exploit modules and run individual exploits on hosts

Burp Suit: Burp Suite is a popular platform that is widely used for performing security testing of web applications. It has various tools that work in collaboration to support the entire testing process, from initial mapping and analysis of an application's attack surface, through to finding and exploiting security vulnerabilities.

Burp is easy to use and provides the administrators full control to combine advanced manual techniques with automation for efficient testing. Burp can be easily configured and it contains features to assist even the most experienced testers with their work.

Angry IP Scanner: Angry IP scanner is a lightweight, cross-platform IP address and port scanner. It can scan IP addresses in any range. It can be freely copied and used anywhere. In order to increase the scanning speed, it uses multithreaded approach, wherein a separate scanning thread is created for each scanned IP address.

Angry IP Scanner simply pings each IP address to check if it's alive, and then, it resolves its hostname, determines the MAC address, scans ports, etc. The amount of gathered data about each host can be saved to TXT, XML, CSV, or IP-Port list files. With help of plugins, Angry IP Scanner can gather any information about scanned IPs.

Cain & Abel: Cain & Abel is a password recovery tool for Microsoft Operating Systems. It helps in easy recovery of various kinds of passwords by employing any of the following methods:

- sniffing the network,

- cracking encrypted passwords using Dictionary, Brute-Force and Cryptanalysis attacks,
- recording VoIP conversations,
- decoding scrambled passwords,
- recovering wireless network keys,
- revealing password boxes,
- uncovering cached passwords and analyzing routing protocols

Cain & Abel is a useful tool for security consultants, professional penetration testers and everyone else who plans to use it for ethical reasons.

Ettercap: Ettercap stands for Ethernet Capture. It is a network security tool for Man-in-the-Middle attacks. It features sniffing of live connections, content filtering on the fly and many other interesting tricks. Ettercap has inbuilt features for network and host analysis. It supports active and passive dissection of many protocols.

You can run Ettercap on all the popular operating systems such as Windows, Linux, and Mac OS

EtherPeek: EtherPeek is a wonderful tool that simplifies network analysis in a multiprotocol heterogeneous network environment. EtherPeek is a small tool (less than 2 MB) that can be easily installed in a matter of few minutes.

EtherPeek proactively sniffs traffic packets on a network. By default, EtherPeek supports protocols such as AppleTalk, IP, IP Address Resolution Protocol (ARP), NetWare, TCP, UDP, NetBEUI, and NBT packets.

SuperScan: SuperScan is a powerful tool for network administrators to scan TCP ports and resolve hostnames. It has a user friendly interface that you can use to −

- Perform ping scans and port scans using any IP range.
- Scan any port range from a built-in list or any given range.
- View responses from connected hosts.
- Modify the port list and port descriptions using the built in editor.
- Merge port lists to build new ones.
- Connect to any discovered open port.
- Assign a custom helper application to any port.

QualysGuard: QualysGuard is an integrated suite of tools that can be utilized to simplify security operations and lower the cost of compliance. It delivers critical security intelligence on demand and automates the full spectrum of auditing, compliance and protection for IT systems and web applications. QualysGuard includes a set of tools that can monitor, detect, and protect your global network.
WebInspect: WebInspect is a web application security assessment tool that helps identify known and unknown vulnerabilities within the Web application layer.
It can also help check that a Web server is configured properly, and attempts common web attacks such as parameter injection, cross-site scripting, directory traversal, and more.

LC4: LC4 was formerly known as L0phtCrack. It is a password auditing and recovery application. It is used to test password strength and sometimes to recover lost Microsoft Windows passwords, by using dictionary, brute-force, and hybrid attacks.
LC4 recovers Windows user account passwords to streamline migration of users to another authentication system or to access accounts whose passwords are lost.
LANguard Network Security Scanner monitors a network by scanning connected machines and providing information about each node. You can obtain information about each individual operating system.
It can also detect registry issues and have a report set up in HTML format. For each computer, you can list the netbios name table, current logged-on user, and Mac address.
Network Stumbler: is a WiFi scanner and monitoring tool for Windows. It allows network professionals to detect WLANs. It is widely used by networking enthusiasts and hackers because it helps you find non-broadcasting wireless networks.
Network Stumbler can be used to verify if a network is well configured, its signal strength or coverage, and detect interference between one or more wireless networks. It can also be used to non-authorized connections.
ToneLoc stands for Tone Locator. It was a popular war dialling computer program written for MS-DOS in the early 90's. War dialling is a technique of using a modem to automatically scan a list of telephone numbers, usually dialling every number in a local area code.
Malicious hackers use the resulting lists in breaching computer security - for guessing user accounts, or locating modems that might provide an entry-point into computer or other electronic systems.
It can be used by security personnel to detect unauthorized devices on a company's telephone network.

# Conclusion:

In the present situation of modernized and electronically dependent works, it becomes more and more important to be protected from hacking and other cyber attacks. This information is mainly gathered from internet and the courses we've done on ethical hacking. Ethical Hacking doesn't teach you to be a black hat hacker but try to become white hat hacker and serve the society. This article presents clearly presents the brief overview, respective rules to be followed, various types, different phases and tools for ethical hacking. A perspective view on ethical hacking is presented in this article.

*References:*

[1] Engebretson, Patrick. The basics of hacking and penetration testing: ethical hacking and penetration testing made easy. Elsevier, 2013.
[2] https://www.geeksforgeeks.org/
[3] https://www.simplilearn.com/
[4] https://www.javatpoint.com/
[5] https://www.eccouncil.org/ethical-hacking/
[6] https://www.greycampus.com/opencampus/ethical-hacking
[7] https://www.computerhope.com/jargon/e/ethihack.htm
[8]https://studyabroad4lessblog.wordpress.com/2017/11/08/ethical-hacking-advantages-and-disadvantages

Compiled by:

Mr. Azeem Shaik (CSI : 01475757) is a student of BE III year (CSE) at Kallam Haranadhareddy Institute of Technology, Guntur, Andhra Pradesh. He can be reached at azeem.shaik.315213@gmail.com.

Mr.Parimi Vijay Kumar (CSI : 01475741 ) is a student of BE III Year (CSE) at Kallam Haranadhareddy Institute of Technology, Guntur, Andhra Pradesh. He can be reached at vijaykumarchowdary05@gmail.com

Mr. Aftab Shaik (CSI : 01475756) is a student of BE III Year (CSE) at Kallam Haranadhareddy Institute of Technology, Guntur,AndhraPradesh. He can be reached at shaikaftab02@gmail.com

Under the Guidance of:

Prof.N Md Jubair Basha is a CSI Life Member: 01085752 working in the CSE department & CSI Student Branch Coordinator at Kallam Haranadhareddy Institute of Technology, Guntur, Andhra Pradesh, India. He has published 20 papers in various National/ International Conferences and Journals. He is an active member of CSI and ACM and can be reached him at nawabjubair@gmail.com

# PRO-CRATE

Aditya Veer, Shardul Kulkarni, Pranav Shreeram

ABSTRACT

Procrate is an attractive, easy, convenient and thoughtfully created project sharing platform, build for students, faculty or enthusiast, where they can upload and share the projects and ideas conveniently. Procrate is sharing projects and ideas, asking questions/doubts, searching for projects, reading and writing blogs, following other users and much more. Procrate is equipped with HASHTAG feature which optimizes the Recommendation and Search functions to meet users' requirements as per his/her profile or preference.

User's rank is decided on XP i.e. Reward Points on access of multiple features of the Procrate the User will keep on earning the Reward Point, providing User a motivation to explore the Procrate frequently. An algorithm like PageRank algorithm ensures the quality Project/Idea gets proper exposure. It ranks the Project based on number of upvotes it gets considering its quality and amount of upvotes. i.e. a user with higher 'User Rank' upvotes any other users project, that upvoted projects 'Project Rank' will be scaled up according to Upvoters 'User Rank'. This functionality helps to get optimum search results and Recommendations.

## Introduction

Procrate, is an online project sharing website. Students will be able to Upload, Download and Review the projects on the website. The students as well as faculty can login in the system. Firstly, the requirements were identified which can be classified as user requirements and functional requirements.

User Requirements:

Students:

Common interdisciplinary project sharing platform at institute level.
Personal space to upload and work on project.
Customizable search option.
Comments Section for each project.
Upvote/ downvote option for projects.
Request download of project reports.
"Ask for help" forum.
Private chat box.
Pull request to provide any suggestion.
Accessibility to read and write blogs, relating to a project that is uploaded, also about any idea.
Accessibility to follow any user registered on Procrate.

Teachers:

Plagiarism checker.
Report the project.

Functional Requirements:

- Login.
- Signup.
- Add personal details.
- Create project.
- Add branch.
- Comment.
- Search.
- Upvote/Downvote.
- Check Plagiarism.

# Motivation

While working on mini projects in first year of our engineering, we encountered some problems. Some of these problems were faced by almost all the students. We found out that the main problems were that there were not enough references, lack of technical knowledge and experience, no enough technical help and the most important problem was that we observed repetition in the topic selection of the projects. And in some of the projects, plagiarism was seen. In most of the cases, teachers cannot check whether the project is copied or not.

Hence, we thought of this concept which will enable students to see the existing projects that can not only help them to find a new one, but the incomplete projects or the projects having some future scope could be useful for them to work on. This will assist in reducing the plagiarism as well.

# Objectives and Scope

The primary objective is to provide a common institute level interdisciplinary platform for sharing of project/code with other students or faculty of the institute.
To provide the facility to register as a new or old user, share or upload your project and write a blog about your project.
To give the accessibility to follow any user, like and comment on a blog or a project or any idea, read blogs, download any project on the uploader's approval, contact the other user, etc.
To facilitate the users with a Help forum so they can ask and answer the queries.

# Problem Statement

To provide a common institute level interdisciplinary platform for sharing of project/code with other students or faculty of the institute.

# System Design

Hardware Requirements:

The project is software based wherein we are creating a website. Hence the hardware requirements aren't much except for a computer with internet connectivity, server, etc.

Software Requirements:

The system is divided into two parts: server side and the client side.

The server side includes the web framework, the backend language, and the database. The client side includes the frontend designing and the website.

Server Side:

Web Framework: Django 2 is frameworks MVC (Model-View-Controller) architecture helps to build proper well-structured web application.

Backend language: Python 3.x has many libraries expands the functional scope of web application.

Database: MySQL is one of the supported open-source databases of Django.

Client Side:

Frontend Language: HTML 5 is latest version available and supported by Django for content display. CSS 3 is used to enhance and beautify the user interface JavaScript brings dynamicity to our page.

The dependencies and the best features of the project are as follows: Dependencies:

Project Data uploaded by the USER.

Hashtags: Functionality such as Search optimization, Project Recommendation, Questioning are solely based on Hashtags.

User preferences regarding the topic he/she is interested in.

Showstoppers:

Adding irrelevant or vague data (hashtags / project contents / questions)

Poor internet connectivity

Entity Relationship Diagram:

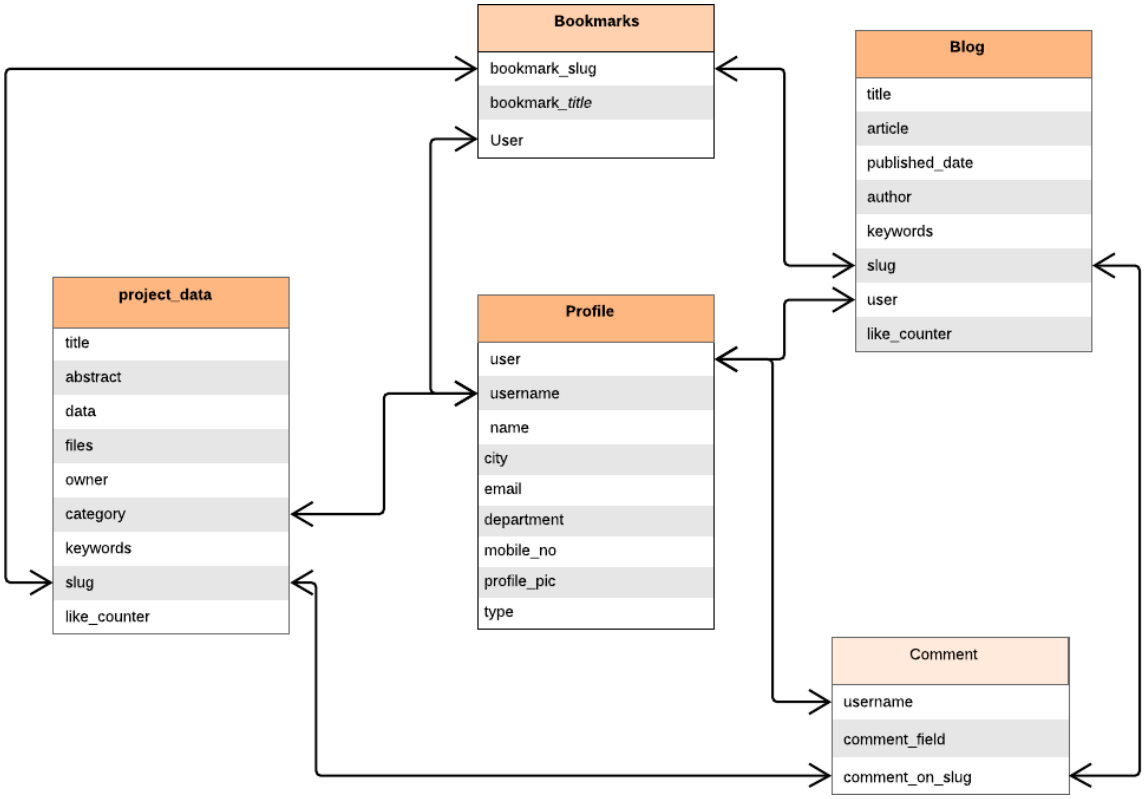



Fig. 3.1. E-R Diagram

Use Case Diagram:

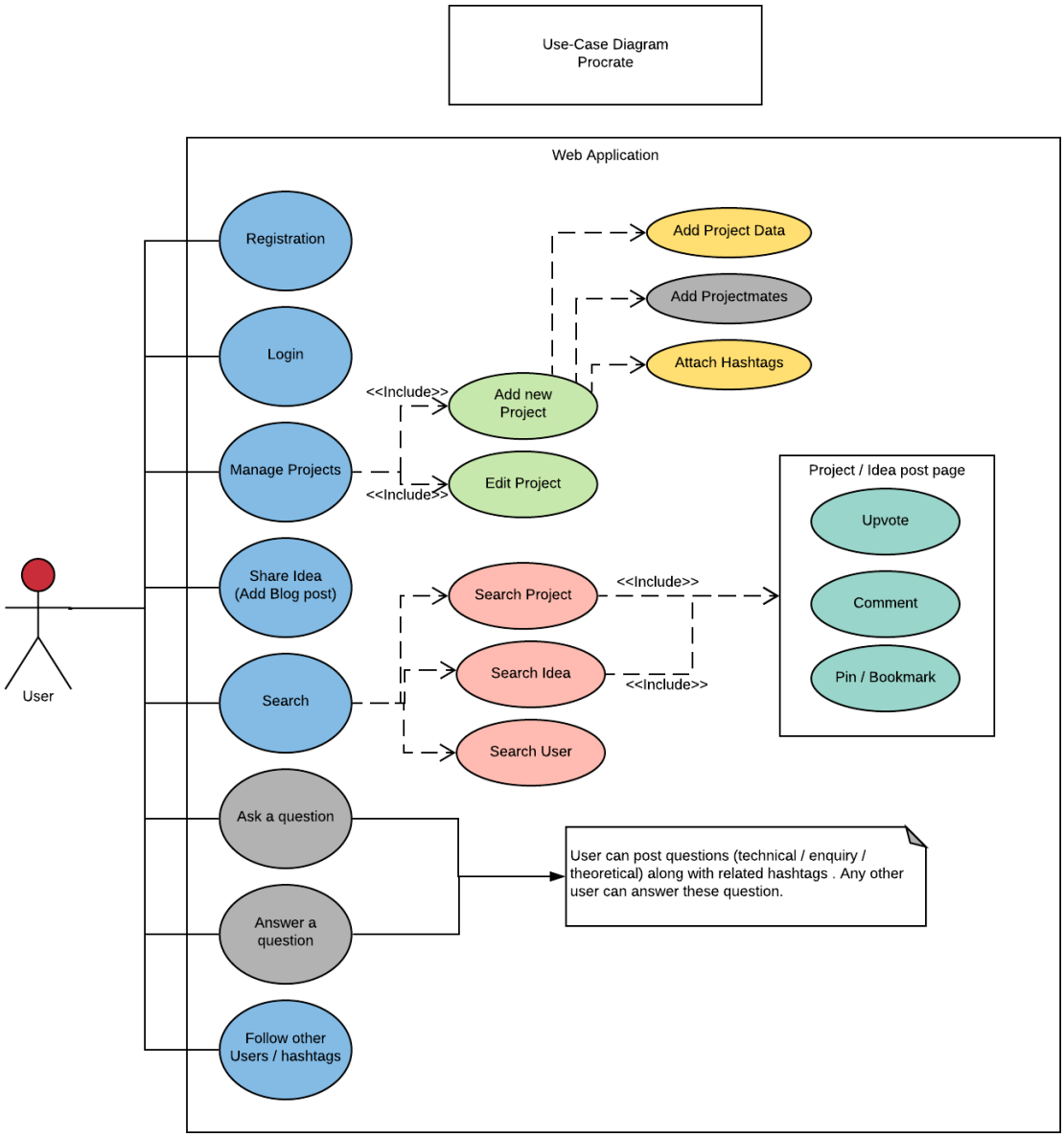



Fig. 3.2. Use Case Diagram

Sequence Diagram:

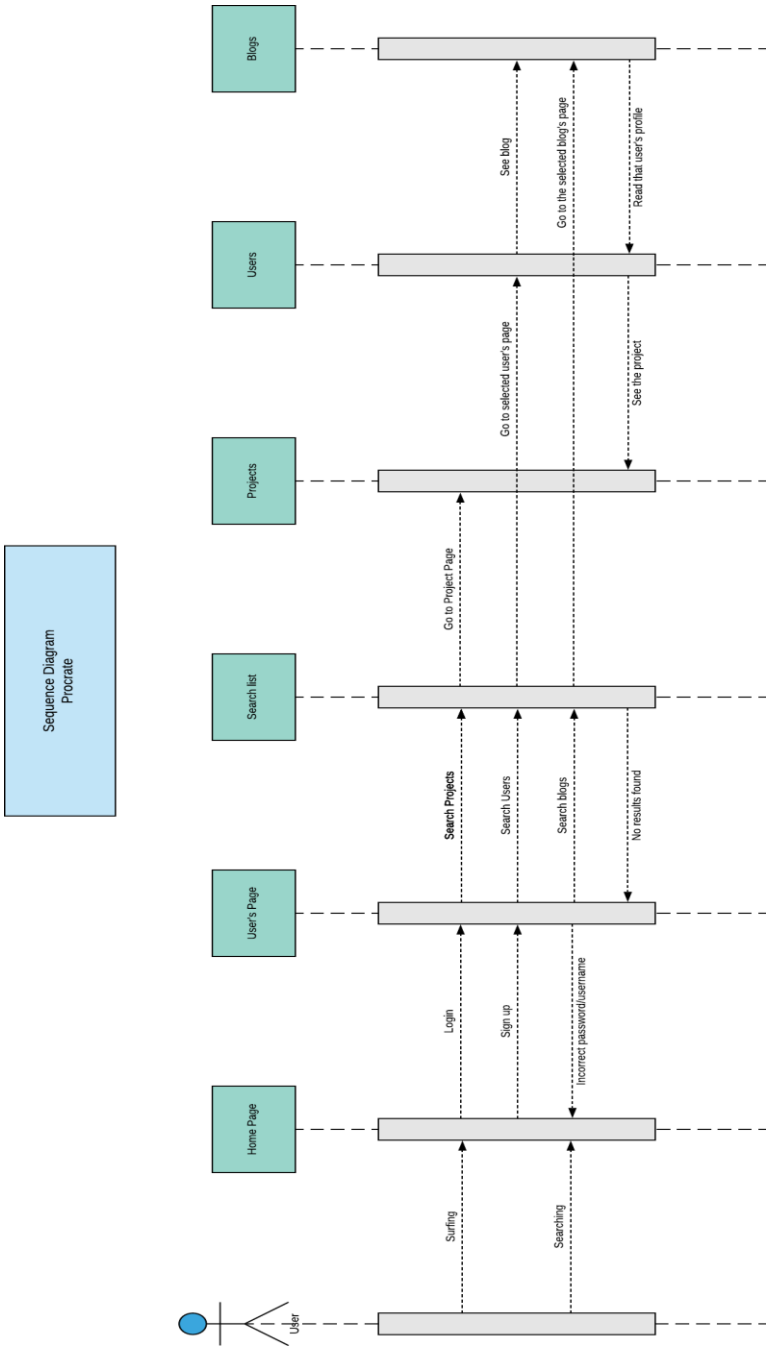



Fig. 3.3. Sequence Diagram

Activity Diagram:

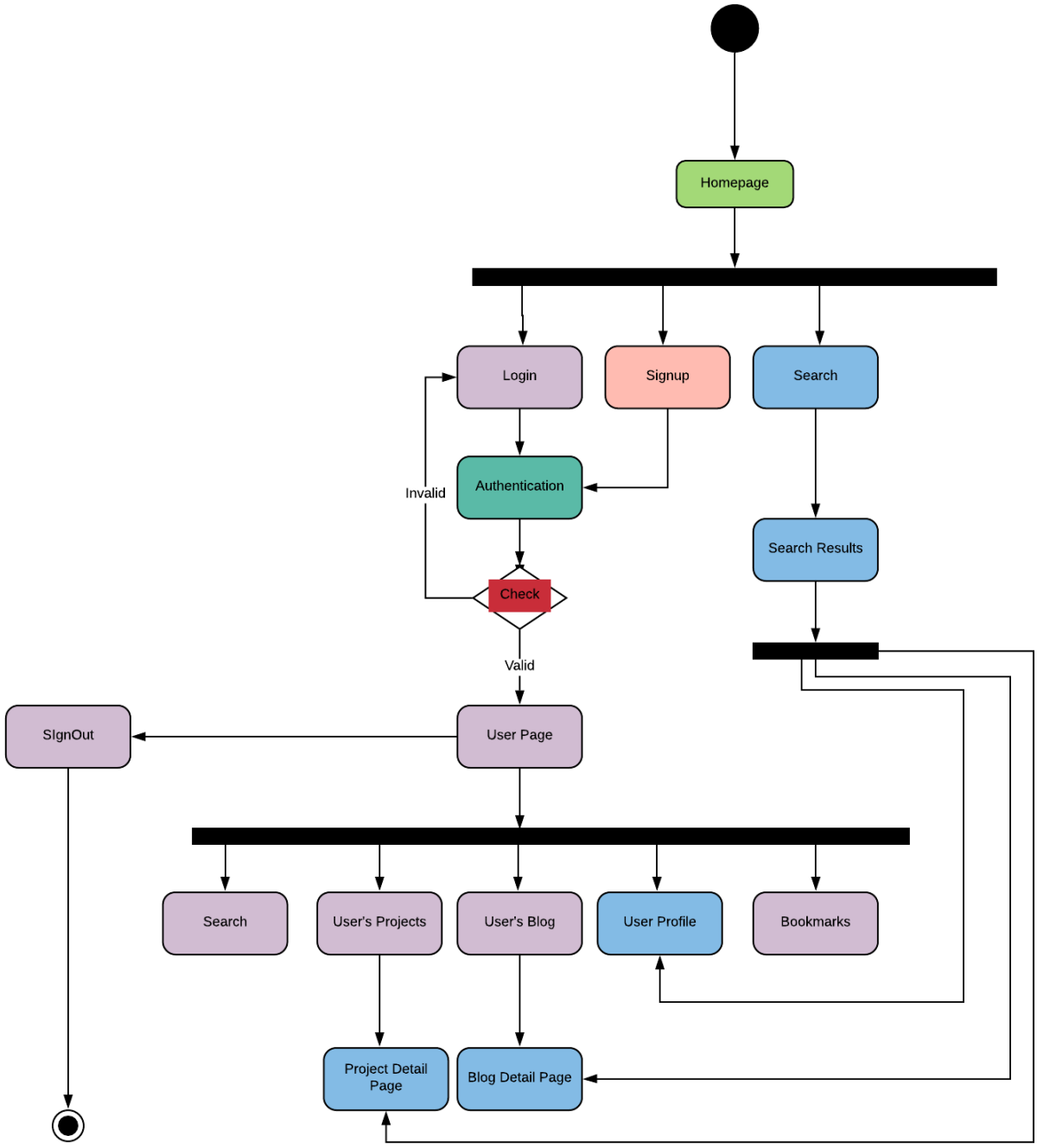



Fig. 3.4. Activity Diagram

# Implementation Details

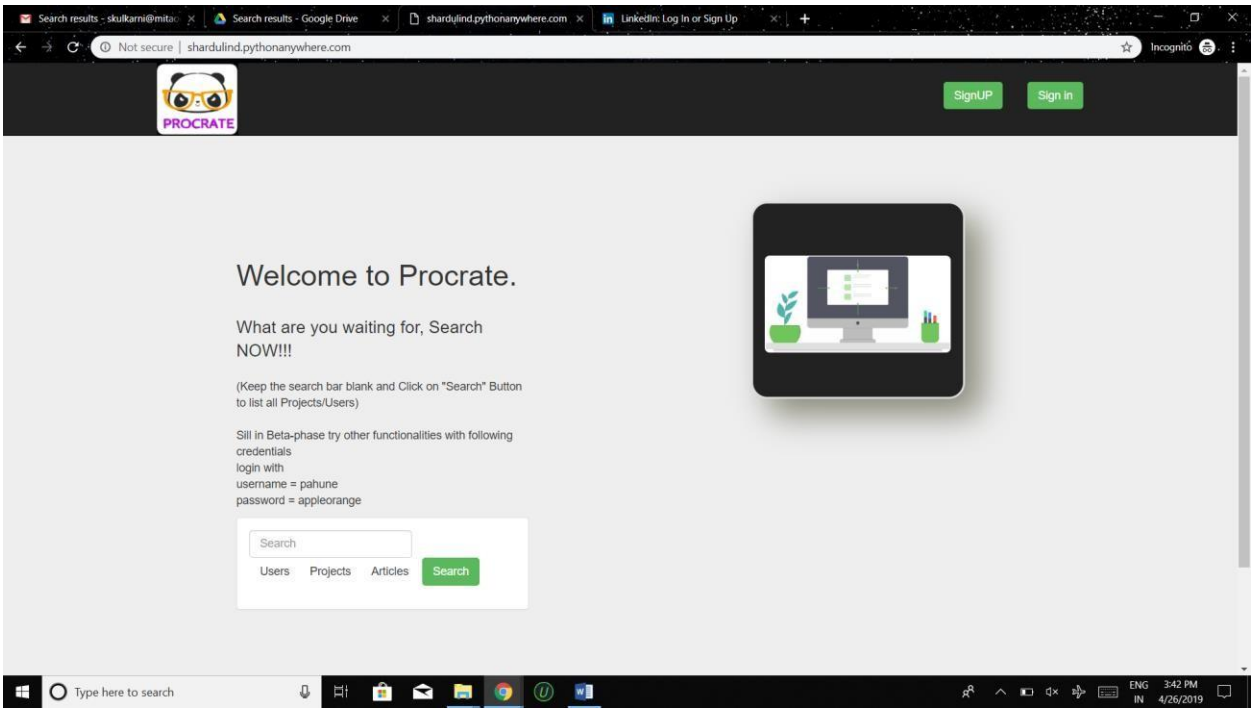
Procrate web application is completely based upon Django Web Framework which is powered by Python. Django works with MVC architecture i.e. MODEL-VIEW-CONTROLLER. Model part deals with data types and databases, the VIEW deals with the templates of the webpages and Controller connects the backend with the frontend as well as the User/Client Requests.

In Django a web-application is made of multiple apps. Each app is meant for functionality. In case of Procrate, we have created following apps they are: Accounts, Projects, utility, blogs. The frontend part is built using HTML5 and CSS3 with the help of Bootstrap.

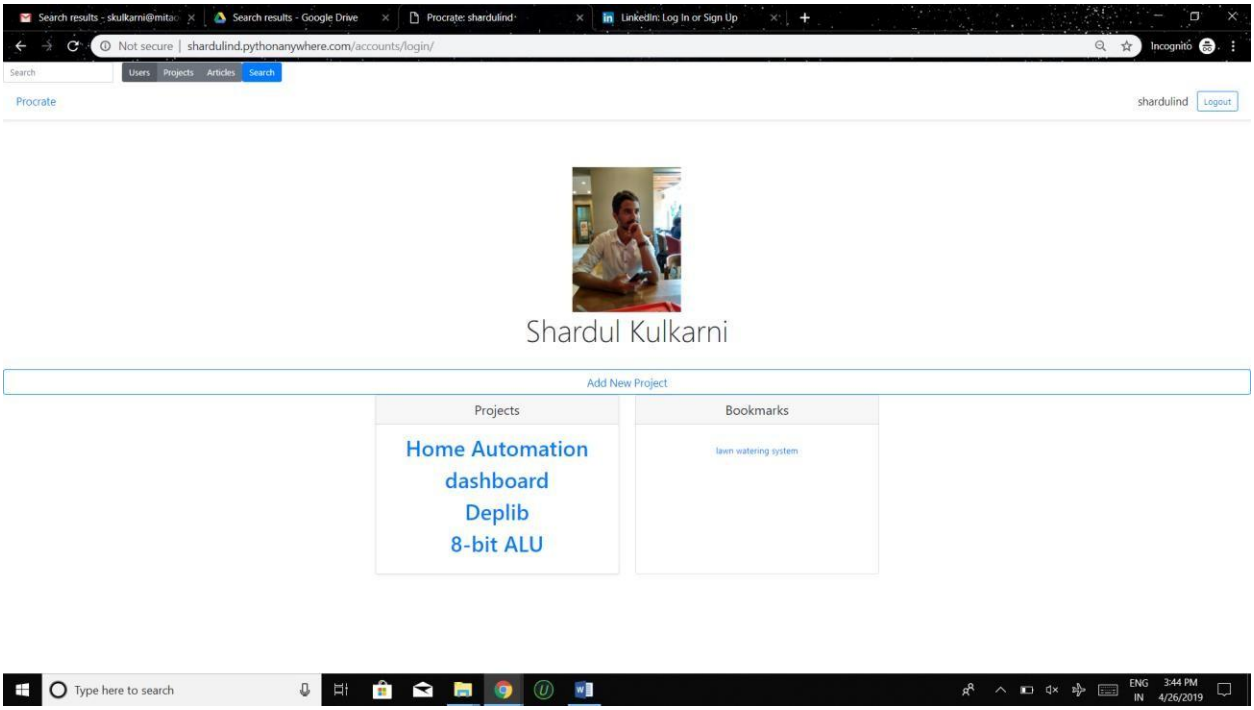
In the server side of the project we used Django 2 as the framework which is MVC (Model- View-Controller) architecture that helps to build well-structured web application. Python 3 has many libraries that expand the functional scope of the web application as the backend language. MySQL is one of the supported open-source databases of Django 2.

# Results

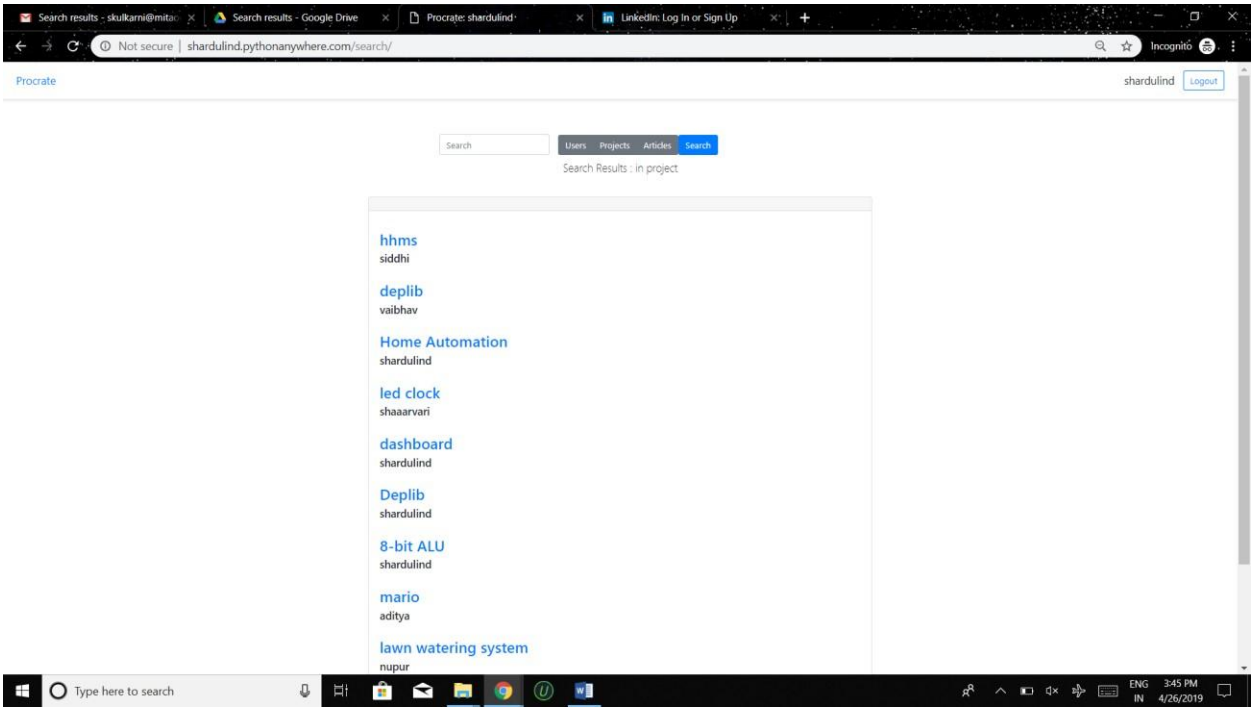
SCREENSHOT: Homepage:




User Page


Search results - skulkarni@mitac
Search results - Google Drive
Procrate: shardulind
LinkedIn: Log In or Sign Up
Not secure | shardulind.pythonanywhere.com/accounts/login/
Incognito
Search
Users
Projects
Articles
Search
Procrate
shardulind
Logout
Shardul Kulkarni
Add New Project
Projects
Home Automation
dashboard
Deplib
8-bit ALU
Bookmarks
lawn watering system
Type here to search
ENG
IN
3:44 PM
4/26/2019

Search


Search results - skulkarni@mitac
Search results - Google Drive
Procrate: shardulind
LinkedIn: Log In or Sign Up
Not secure | shardulind.pythonanywhere.com/search/
Incognito
Procrate
shardulind
Logout
Search
Users
Projects
Articles
Search
Search Results : in project
hhms
siddhi
deplib
vaibhav
Home Automation
shardulind
led clock
shaaarvari
dashboard
shardulind
Deplib
shardulind
8-bit ALU
shardulind
mario
aditya
lawn watering system
nupur
Type here to search
ENG
IN
3:45 PM
4/26/2019

# Conclusion and Future Scope

The objective of the project to provide the project sharing platform having a userbase of a College is optimally achieved in Procrate. Procrate is capable to handle user accounts and creating user profile, upload projects, bookmark projects and Comment on the projects.

Many more dynamic functionalities are to be added on the Procrate such as adding teammates on project so as to give proper contributions to every member. Give more rights to user related to project such as hiding the files, providing citation permissions, providing a upvote/like button which will help to optimize the search results.

*REFERENCES*

[1] Django Unleashed by Andrew Pinkham
[2] "Django Official Documentation" https://docs.djangoproject.com/en/2.2/
[3] "Bootstrap Documentation" https://getbootstrap.com/docs/4.3/getting-started/introduction/

Complied by:

Mr. Aditya Veer (01514849) is a student of BTech ||| year (CSE) at MIT Academy of Engineering Alandi. He can be reached at adityaveer8524@gmail.com

Mr. Shardul Kulkarni (01514826) is a student of BTech ||| year (CSE) at MIT Academy of Engineering Alandi . He can be reached at skulkarni@mitaoe.ac

Under the guidance of :

Mr. Pranav Shreeram (F8001620) is a professor at MIT Academy of Engineering Alandi . He can be reached at prshriram@it.mitaoe.ac.in

# Simple but Impactful Tips to stay secure in Modern World

Sayed Sohail Jamadar, Gayatri Bajantri

This article focuses on some of the important tips for securing our online presence in every single day of our modern life. You might be getting several tips on internet regarding “how to stay secure online”, but most of those miss out to mention these essential points which are to be listed below:

## Password:

You may find everyone saying to use good and long passwords, but strong and good password is not enough. That is, the first tip is to make use of different passwords for different accounts/websites. The reason behind using different password is if unluckily any of your account gets compromised it would not affect your another account and when you use same password all over the internet there is a high risk of getting your all accounts compromised.

## Antivirus:

Antivirus software you all may have in your phone/computers, but have you ever tested your antivirus? Many of the free antivirus software available on internet provide assurance of virus free device, but none of them are up to the worth. Try checking out list of some good and top antivirus software and see if your antivirus is on the list or not. Many of the paid antivirus software provide free trial which can help you to select best antivirus for your device. I recommend to use Bitdefender antivirus or Kaspersky antivirus, as they provide constant updates and latest antivirus definitions for helping you to stay secure.

## Avoid using Modded/Cracked Applications:

Instead of purchasing license/key of paid application, people try to install modded application. As the hackers implement harmful Trojan virus in modded applications which tries to gain access to your devices without your permission. Many of the recent reports have covered that these recent modded applications are built with ransomware virus, which is considered to one of the most harmful and dangerous virus. So avoid using such Modded applications and protect your data from getting compromised.

## Avoid suspicious web pages:

Try to avoid visiting suspicious and unwanted websites. Do not enter sensitive information like credit card details, OTP, password, mobile number on untrusted websites. Many of such websites are developed for phishing and such other illegal activities.

## Email Leak:

Have you been getting lots of spam mails on your email? The reason behind that could be your email been pwned/leaked. Make sure you follow the tips below to overcome this problem:

- Report such spam mails.
- Use strong passwords.

- Enable 2 factor authentication (like OTP).

In Conclusion, ones privacy on the internet is very important because of all the applications, services, scams and viruses on the internet that are waiting for any given chance to steal someone's personal material. Now it's your own duty to protect your privacy, maintain security and use right software.

## Compiled by :

Mr. Sayed Sohail Jamadar (2SA18CS404) is a student of BE III year (CSE) at Secab Institute of Engineering and Technology College, Vijayapura Karnataka. He can be reached at sohailjamadar2@gmail.com

## Under the guidance of :

Mrs. Gayatri Bajantri: She is a Professorin CSE, Secab Institute of Engineering and Technology College, Vijayapura Karnataka. She can be reached at gayatrib.cse@secab.org

# Basics of Semantic Web and Its Implications Towards Educational Systems

M. Rajakumar, P. Kiruthika,

## Introduction:

The concept of semantic web brought development in our present world wide web. The improvement made is that the web pages are structured and tagged. So that the data can be directly interpreted by the computer without humans. The semantic web technologies have its applications in almost all fields. Some of its most common applications are in oil and gas enterprises, publishing and media industries, life science and pharma industries and in insurance sectors. Thus, in this paper the applications of semantic web towards the educational sector has been explained in detail.

## Basics Of Semantic Web:

The semantic web can be viewed as an improvement in the present web. The semantic web provides meta data which is understandable to the interpreter. The data sent over the web is machine – interpretable because of the metadata. These metadata add better descriptions to the content of data which is already present. With this additional information it is easy for anyone to interpret. The computers itself can think and interpret like a human. This technology makes many things easier.

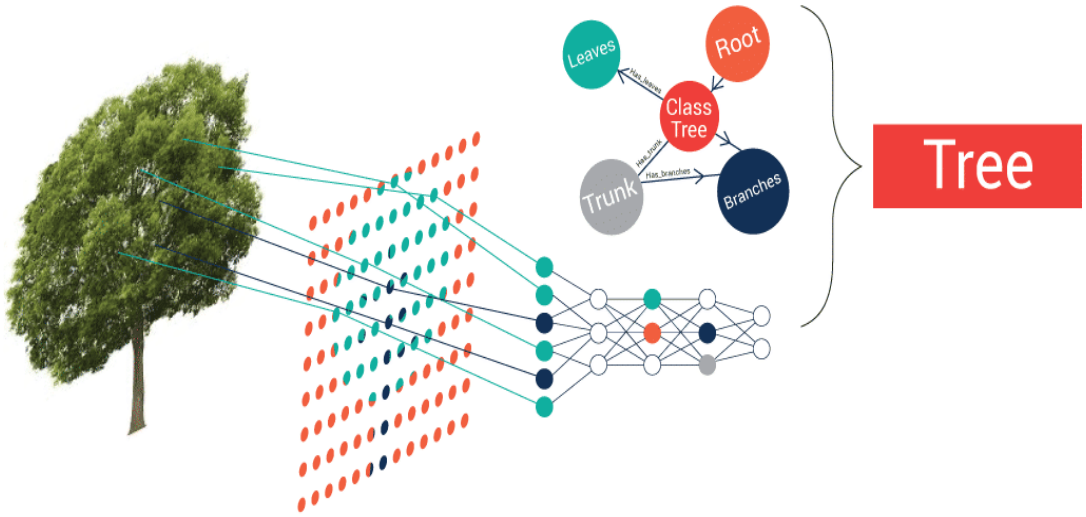



*Fig. 1 – Structure of Semantic Web*

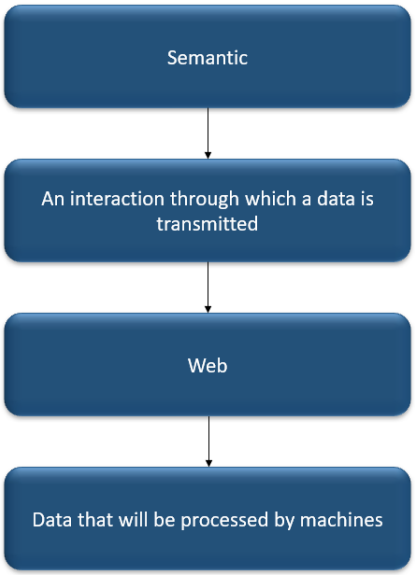



*Fig.2 – Definition of Semantic Web*

The founder of semantic web is Tim Berners-Lee. He explained semantic web as follows, in which semantic means, data that will be processed by machines and web means, an inter connection through which a data is transmitted from URI's to resources. According to Tim Berners-Lee. Semantic web comes under the combination of

- Automation of Information retrieval
- Internet of Things
- Personal Assistants

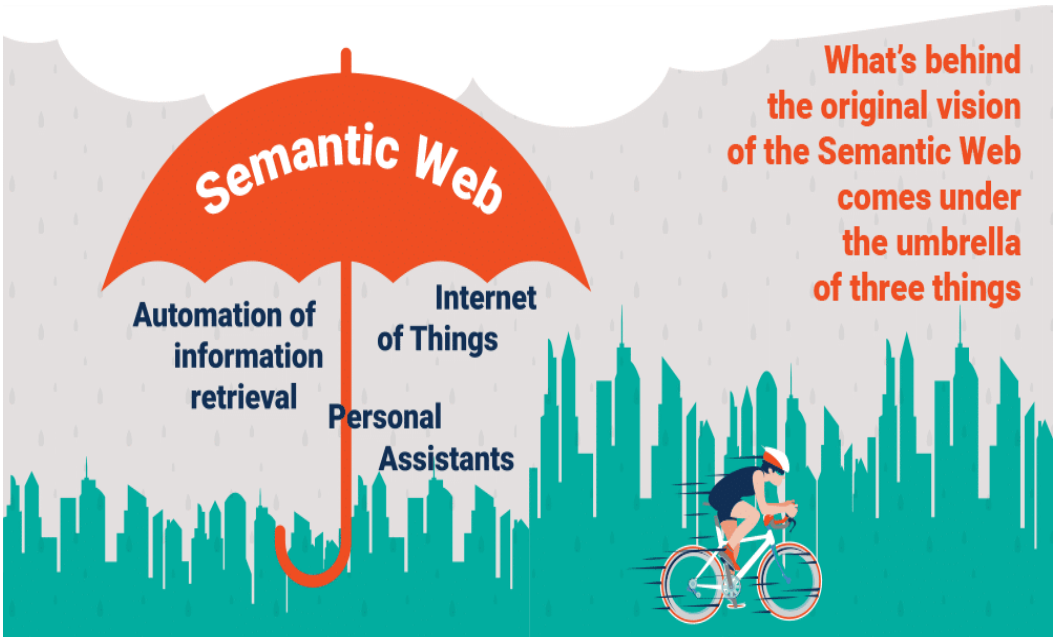



*Fig. 3 – Semantic Web umbrella*

The term semantic web evolves with the presence of two terms namely,

- Linked Open Data
- Semantic Metadata

## Linked Open Data

Linked data is a kind of structured data which is drawn in the form of a graph and links across various servers. The founder of semantic web Tim Berners-Lee officially framed four rules of Linked data in 2006.

They are:

- Use URI's as name for identifying things
- Use HTTP URI's to make those names visit
- Provide useful information in the URI using the standards (RDF * SPARQL)
- Link these URI to other URI's
- Linked open data is easy for both people and machines for handling of various servers and interpret through it.

Linked Open Data consists of

- Factual Data
- Ontologies (Schematic Definitions)
- Classes of objects
- Relationship among objects
- Attributes

Some examples for Linked open Data are encyclopaedia, scientific databases, travelling, etc, …


*Fig. 4 – Linked data*

## Semantic Metadata

Semantic Metadata is adding more semantic tags to regular web pages which makes it semantic web. The purpose of this semantic metadata is to make the webpage stronger. It helps the user to search easily based on semantic category. And this gives the user the exact page he wants. But almost 30% of web page is semantic web in today's world.

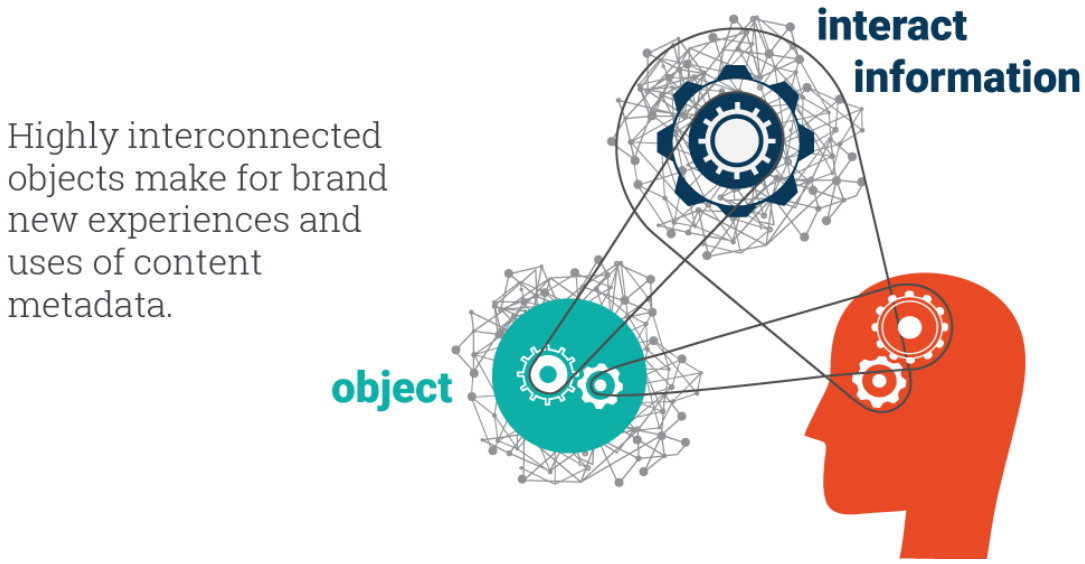



*Fig. 5 – Semantic meta data*

Standards

Fundamental Standards considered are:

- Resource Description Framework (RDF)
- CPARQL protocol and RDF Query Language (SPARQL)
- RDF

A language for showing relationship between defined objects in the form of a graph.

SPARQL

A query language for RDF data format

# Semantic Web In Education Systems

The concept of semantic web differs from the usual web in the way of semantic structure of web pages. The main purpose of semantic web is to make the content interpretable to both humans and machines. Thus, in semantic web software agents interpret the data in the web and complete the execution.

Focussing on education system, semantic web provides better education assistants than web – based systems. Some of the tasks performed by semantic web for education are, providing featured learning materials for the learners, ensure for the good interaction between the learner and webpages, check whether the learner meets all the required services etc, …

In 2004, Anderson and WhiteLock defined Educational Semantic web based on three features:

The first is the capacity to store and retrieve the data from the web

The second is automated retrieval of data from the web without the help of humans

The third is the support of the internet or broadband to access the web

Some technologies which contains the above features are:

*Ontologies:*

The data read by the humans from the web is in the natural language for readability. But for the machine's understanding the semantics of the data has to be interpreted. So, if we represent the data semantically it will be easy for the machine to store, retrieve and analyse data.

*Learning Standards:*

The standards are applied in resources for developing, exchanging, qualifying, etc, … The semantic web standards such as RDF (Resource Description Framework) can be applied for resources such as IEEE, IMS, etc, …

*Semantic Web Services:*

Semantic web performs two things:

It makes the data in the Worldwide Web (www) as

- Machine Processable
- Machine Interpretable

By doing this, the static data is WWW is transformed as distributed data.

*Intelligent Agents:*

Intelligent Agents are autonomous agents of the software type. These agents are used for convenient interaction with the web users.

REFERENCE MODELS FOR SEMANTIC WEB EDUCATIONAL SYSTEMS

In education system, generally the relationship exists only between the machine and the learner. The learner can be individual learner or group learners. The module involved are:

- Roles
- Resources
- Interface environment
- SWBES
- ROLES

In Education systems, there are several roles such as follows:

Teacher's Role: The role of teachers is to monitor the learners and check whether the classes are going on properly

Learner's Role: The main backbone of semantic web - based learning is the interest of the listeners. The reason for the listener's interest is their longing towards knowledge gain

Author's Role: Author's represent the content of the data in a structured way to the learners. They perform activities such as

- Preparing educational content

- Provide detailed instructions
- Adaption and Personalization activities

Group's Role: Nowadays Group learning are more concentrated. Because in Group learning interaction and sharing of thoughts will take place, which makes the learning more effective.

Developer's Role: Developers role is to develop an application for interaction and to add additional features to it.

Auditing's Role: The role of the auditor is to audit the proper working of the system. Every role of each player is audited and improvisation steps for the system is suggested.

Administrator's Role: They are a kind of artificial agents known as controller agents. They control the overall system from the beginning to the end.

Educational Resources

The successful hit of an educational system application relies on its resources. The simpler educational resources, the easier will be the understanding. The way for making the educational resources easier is giving more simpler examples, activities and others. Thus, the learners will make use of easy educational resources and the system will become popular.

Interface Environment

Designing an interface environment is very essential. Because, the interface environment should support every role mentioned earlier. And also, the interface environment should act as an interface between the semantic web-based system and the player (of any roles). Apart from choosing of interface environment the type of resource and context is also essential. Thus, the interface environment should be specially designed with different kinds of tools and frameworks. Example: Mobile phones, PDAs, Browser, etc, …

SWBES : SWBES is acronym for Semantic Web Based Educational System. SWBES provides a path to achieve the longing goal of learning. And it is implemented with the names of Learning Management system, Virtual Learning Environments, Pervasive Educational system and so on.

The components of SWBES are:

Ontologies

The data read by the humans from the web is in the natural language for readability. But for the machine's understanding the semantics of the data has to be interpreted. So, if we represent the data semantically it will be easy for the machine to store, retrieve and analyse data.

Educational Ontologies : Educational Ontologies are responsible foe verifying the knowledge about the particular thing. This includes Domain module Ontologies, Pedagogical ontologies, Task Ontologies, etc, …

Interaction Ontologies : The components responsible for interaction such as artificial agents, semantic web services and tools used are called interaction ontologies.

Context Ontologies : Context Ontologies deals with ensuring the communication and publication of those resources on the web. As a result, the interaction takes place smoothly.

Tools

For educational environment there may be many tools. And the kind of total differs according to the choice of user such as browser, PDAs etc, …

Educational Tools : For the proper functioning of the SWBES, the educational tools provide more interaction of the player with the system. Some of the tools are authoring tools, teacher tools, semantic wiki tools and others.

Tools to SW support

This literally means the semi-automatic features that are built in. These are used to support the artificial agents. Some tools are ontology mapping, social semantic web, ontology visualization and others.

Administrative tools :To configure the settings according to the roles such as learner, teacher and others.

Semantic web services :Semantic web performs two things:

It makes the data in the Worldwide Web (www) as

- Machine Processable
- Machine Interpretable

By doing this, the static data is WWW is transformed as distributed data.

Educational SWS :Educational SWS are used to achieve the system's goal. Activities such as integration of content, preparing the educational resources, evaluating and comparing among the other educational applications and so on.

Support SWS :Activities other than education while taking place in a system should also be supported such as searching, mining and others.

Intelligent Agents :Intelligent Agents are autonomous agents of the software type. These agents are used for convenient interaction with the web users.

Tutoring Agents :Useful for checking of proper interaction system with the user. This includes, personalization of content to various users and others.

Support Agents :To support other agents for completion of its work

Controller Agents :They are also known as administrative agents. Their task is to provide access, find untruthful agents and others.

# Conclusion

In this current scenario of the world, even though there are computers, humans are still lazy. So semantic web makes the work of humans easier than world wide web. The structured tagged metadata makes the machine itself to interpret the data. And also, the applications of semantic web

towards educational sector has also been explained in detail in this paper. Thus, semantic web is considered as a most useful invention for human life betterment.

*REFERENCES*

[1] https://www.google.com/search?q=semantic+meta+data&rlz=1C1CHBF_enIN781IN781&sxsrf=ALeKk03iUop-vtb3StKHpMlwCaZA13X8A:1586681144455&source=lnms&tbm=isch&sa=X&ved=2ahUKEwi9tfLvv-LoAhXyyDgGHZp3AyUQ_AUoAnoECA4QBA&biw=1242&bih=597#imgrc=nAGSC3l0fu4IkM

[2] https://www.ontotext.com/knowledgehub/fundamentals/what-is-the-semantic-web/

[3] http://www.cs.cmu.edu/~sisotani/artigos/Towards%20a%20Reference%20Model%20to%20Semantic%20Web-based.pdf

[4] https://expertsystem.com/semantic-web-technologies-applications-enterprises/

[5] https://www.explainablestartup.com/2016/08/the-history-of-semantic-web-is-the-future-of-intelligent-assistants.html

Compiled by:

MR M. RAJAKUMAR (01491535) is a UG student of SRM Valliammai Engineering College in Computer Science and Engineering department. His areas of interests include IoT and Cyber security.

MS P. KIRUTHIKA [01491613] is a UG student of SRM Valliammai Engineering College Pursuing degree in B. TECH INFORMATION TECHNOLOGY. Her area of interests includes Cyber security and Blockchain.

Under the Guidance of :

Dr. M. Senthil Kumar (LM- I1504760) is currently working as an Associate Professor in Computer Science and Engineering department at SRM Valliammai Engineering College of Tamil Nadu. He is a CSI-Student Branch Counsellor of the College. His research interests are in IOT, Big Data, Software Engineering and development of new tools for effort estimation.

# Ad Hoc Bluetooth Communication System

Naishadh Vora, Aayush Shah,
Umang Sakhare, Amit D. Joshi

Abstract:
Messaging applications that leverage the internet and cell tower infrastructure for communication have proliferated the market in recent years. However, communication based on Bluetooth remains a niche sector. In this article, we have attempted to delineate the process of implementing an ad hoc Bluetooth based messaging application.

## Introduction:

An ad-hoc network refers to a network where other devices can temporarily plug in for the duration of the communication. Bluetooth is a short-range wireless network for ad-hoc communication, where the connections are established when the devices are in close proximity. Ad-hoc networks have seen an unprecedented growth in the current times where an increasing range of devices are getting equipped with near-field communication networks like Bluetooth and WiFi. This evolution is driving a new approach of communication that is replacing the need for cables, and allowing communication despite the absence of the internet.

This leads us to the challenges faced in ad-hoc communication: the communication is possible only when the devices are in close proximity, and the connections are not persistent. This article aims to demonstrate how multiple devices can communicate with each other in the absence of a direct, persistent connection using a broadcasting protocol that delivers a message utilizing the indirect connection established between devices. Instead of using a routing table that assumes pre-established, persistent connections, we have to rely on the fact that Bluetooth connections are temporary and the path to a device might change while the message is on the way.

## Implementation Details:

This application uses the users mobile number as a unique identifier for each device. It scans for nearby devices that might have Bluetooth enabled, and attempts to pair with them to establish a connection. The list of discovered devices is saved, and accessed every time the user wants to send a text message. The message is broadcasted to all devices in the proximity in a round robin fashion by establishing a connection and delivering the message individually to each device. Since the application is geared towards maximizing efficiency in case of dynamic networks, it does not generate routing tables - since the variable nature of the network would require frequent updates, leading to needless discovery messages.

The network uses JSON Objects for communication, and they consist of the following fields -

- Timestamp
- Recipient Number

- Sender Number
- Message
- Current Hops

Every time a message is to be sent by the user, a JSON Object is created with the current timestamp, and broadcasted to all the devices in the list. Every time a device receives the message, it checks the Timestamp. If a message with the same timestamp was received earlier, the message is dropped. Next, the current hops are reached. If the current hops exceeds the max hops allowed, the message is dropped. Otherwise, the number of hops is incremented by one. Then, assuming the device is not the intended recipient, the message is broadcasted to all devices in the current device's list, except the one it received the message from. This process continues until the message reaches the intended recipient, or gets dropped if the device does not exist in the network (either due to no new device to transmit to, or through reaching the max hop count).

There are further nuances to the process to increase the reliability and security of transmission. For example, the JSON objects are encoded before being sent, to prevent snooping during transmission. Furthermore, the process of sending messages may fail without the sending device having any knowledge of the same. To remedy this, acknowledgement messages are used with a retry system to aid reliability of communication.

# Applications:

This method of ad-hoc communication is used in our application for exchanging text messages, but can be used in countless other ways. Some other applications include:
Disaster management - discovery of devices constantly emitting distress signals can help first responders to find people stuck in case of a crisis
In case of a pandemic - People can be tracked if they have come in contact, or in vicinity of other infected humans by the devices maintaining records of devices that have come in close proximity.
In case of a complete power outage - Communication can still continue with a modified routing protocol in densely populated regions by smartly integrating all ad- hoc mediums- Bluetooth, WiFi, sound, and mobile networks with limited connectivity.

# Future Work:

This application can be built upon by utilizing WiFi and hotspots in a similar fashion, where multiple devices can connect to an open, pre-existing WiFi network or a hotspot created by one of the devices that is in proximity of the most number of devices. Further work could be done in combining the two in case of failure of any one technique, in order to prevent the transmission being blocked before it can reach its destination.

# Conclusion:

Ad hoc networks are the ideal way to aid communication in disaster struck areas, and in this article, we have outlined one possible implementation for the same. We have addressed the issue of the lack of persistent connections between devices by introducing a rudimentary protocol for connecting

to multiple devices at once, and defining a data transmission format suitable for broadcast without flooding the network pathways.

*References:*


[1] Gergely V. Záruba, Stefano Basagni and Imrich Chlamtac: Bluetrees – Scatternet Formation to Enable Bluetooth-Based Ad Hoc Networks, IEEE, 2001, ISBN 0-7803-7097-1

[2] Ching Law, Amar K. Mehta and Kai-Yeung Siu Performance of a New Bluetooth Scatternet Formation Protocol, Massachusetts Institute of Technology, 2001

[3] Alok Aggarwal, Manika Kapoor, Lakshmi Ramachandran, and Abhinanda Sarkar: Clustering algorithms for wireless ad hoc networks. In Proceedings of the 4th International Workshop on Discrete Algorithms and Methods for Mobile, 2000

[4] Godfrey Tan, Allen Miu, John Guttag and Hari Balakrishnan: Forming Scatternets from Bluetooth Personal Area Networks, MIT Laboratory for Computer Science, MIT-LCS-TR-826, 2001

[5] Charles E. Perkins: Ad Hoc Networking. AddisonWesley, 2001, ISBN 0-201-30976-9

[6] V. K. Quy, N. T. Ban, and N. D. Han, "An advanced energy efficient and high performance routing protocol for MANET in 5G," Journal of Communications, vol. 13, no.12, pp. 743-749, 2018

[7] L. Hanzo and R. Tafazolli, "A survey of QoS routing solutions for mobile ad-hoc network," IEEE Comm. Surveys & Tutorials, vol. 9, no. 2, pp. 50-70, 2007



Mr. Naishadh A. Vora (111708060) is a student of TY B.Tech (Information Technology Engineering) at College of Engineering Pune, Maharashtra.
Email: vorana17.it@coep.ac.in
Membership Number: 01475202



Mr. Aayush B. Shah (111703003) is a student of TY B.Tech (Computer Engineering) at College of Engineering Pune, Maharashtra.
Email: aayushbs17.comp@coep.ac.in
Membership Number: 01475224



Mr. Umang S. Sakhare (111708071) is a student of TY B.Tech (Information Technology Engineering) at College of Engineering Pune, Maharashtra.
Email: umangs17.it@coep.ac.in
Membership Number: 01475211



Mr. Amit D. Joshi is an Asst. Professor and coordinator of CSI COEP Student Chapter at College of Engineering Pune, Maharashtra.
Email: adj.comp@coep.ac.in
Membership Number: 01180301

# Texture Analysis Techniques in Image Processing

Anamika Gupta, Aayush Khattar,
Anshuman Gupta, Vaishnavi Verma

Abstract

Texture is an important quality of an image. The article presents an overview of several approaches to image texture analysis. Four major areas of texture analysis i.e. Feature Extraction, Texture Segmentation, Texture Classification and Reconstruction of 3D Surface Geometry are discussed. Gray Level Co-occurrence Matrix (GLCM) techniques of feature extraction, Segmentation using Artificial Neural Networks and Clustering, Texture classification using various classifiers are also discussed. Further, the Active and Passive methods for Reconstruction of 3D surface geometry are explained in brief.

## Introduction:

In this era of smartphones, DSLR cameras and numerous imagery techniques in fields like medical science, remote sensing, archaeology, etc., there is an abundance of images in our life. But to harness the information that is present in images, various manipulation and processing techniques are required. The field of applying numerous manipulation techniques and operations to enhance images or extract meaningful information from images which may or may not be evident from naked eyes is called Image Processing [4]. When this processing is done by making use of computers then it is called Digital Image Processing [4]. Some applications of digital image processing include image restoration and sharpening, colour processing, pattern recognition and many more.

One can see a digital image as a 2-dimensional array or a grid of pixels with each pixel having different intensities. An RGB image is an image in which each pixel value has 3 different intensity values - red intensity value, green intensity value and blue intensity value, which can be used to represent colored pixels. Whereas, a grayscale image has only one intensity value which represents gray level intensity to display black and white images. Pixel intensity is represented using an 8-bit value ranging from 0 to 255 [10].

Texture analysis in image processing refers to the detection and identification of different textures or patterns like smoothness, bumpiness, etc associated with an image so as to classify different areas or regions of an image with an aim to further analyse different aspects of the image [1]. This technique evaluates different textures by using various spatial features of an image like energy, entropy, correlation, etc which describe how various pixels are related to other pixels in an image or a particular region of an image [1][5].

There are four major areas in texture analysis. These are-

Feature extraction: Extracting spatial features (like energy, entropy, contrast), transform features, edge and boundary features, etc of various regions of an image using techniques like GLCM, pattern spectrum, etc. [1]. This technique is used for facial recognition [19], fingerprint recognition [18], etc..

Texture segmentation: Differentiating different textures in an image on the basis of the extracted features [1]. This technique is used in detecting various regions of interest in biomedical images, radiological analysis, etc. [21].

Texture classification: Classification of different regions of an image based on the different textures [1]. Texture classification is used in a wide range of applications such as industrial inspection, image retrieval, medical imaging, remote sensing, etc. [20].

Reconstruction of 3D surface geometry: Images are 2-dimensional representations of real world objects. After different spatial features have been detected, reconstruction of how the texture would have looked like in 3-dimensional space can give necessary insights for further study [1]. For example, 3D rendering of objects made of wood or marble [17].

## 1. Feature Extraction

One of the most common techniques to extract spatial features is the Gray Level Co-occurrence Matrix method (GLCM).

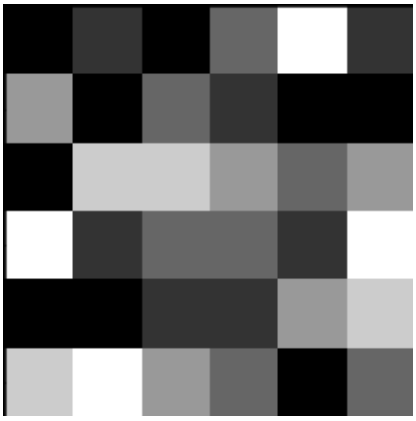
Gray Level Co-Occurrence Matrix (GLCM):

The GLCM method is a statistical method that is used to examine spatial relationships between different pixel intensities in the image which can be used to identify and analyse different textures within the image [1][3].

The construction of GLCM is fairly straightforward. The first step is to convert the image into a grey-scale image so as to deal with grey level intensities. Then construct a matrix GD of size LxL, where L represents the maximum levels of pixel intensities in the image. For example, if the highest pixel intensity level is 4, then we would have 5 levels namely 0,1,2,3 and 4.The second step is to specify d and Θ to form a distance vector D(d,Θ) where:

d is the distance or number of pixels between the 2 pixels under consideration, i.e., if d = 1, then we consider the immediate next pixel in constructing the GLCM.

Θ is the angle between the original pixel and the pixel we are comparing. For e.g., the pixel on the right of the original pixel has Θ = 0°, the pixel above the original pixel has a theta angle of Θ = 90° and so on [2].



| | | | | | |
|---|---|---|---|---|---|
| 0 | 1 | 0 | 2 | 5 | 1 |
| 3 | 0 | 2 | 1 | 0 | 0 |
| 0 | 4 | 4 | 3 | 2 | 3 |
| 5 | 1 | 2 | 2 | 1 | 5 |
| 0 | 0 | 1 | 1 | 3 | 4 |
| 4 | 5 | 3 | 2 | 0 | 2 |

*Figure 1: Image(6x6 pixels)* *Fig. 2: Pixel intensity levels of the image*.

Now initialize each entry of GD with zeroes. Then for each pixel (say P1) in the image, consider every pixel(say P2) that is at distance d and angle Θ from P1. If grey level intensities of P1 and P2 are i and j respectively, then increment GD(i,j) by 1. After the whole matrix is formed, the GD is transposed to give GDT. Then, the two matrices are added to form a symmetric GLCM. The symmetric GLCM is then normalized to further compute the various spatial features of the image [2].

| | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 0 | 4 | 4 | 4 | 1 | 1 | 0 |
| 1 | 4 | 2 | 3 | 1 | 0 | 3 |
| 2 | 4 | 3 | 2 | 3 | 0 | 1 |
| 3 | 1 | 1 | 3 | 0 | 2 | 1 |
| 4 | 1 | 0 | 0 | 2 | 2 | 1 |
| 5 | 0 | 3 | 1 | 1 | 1 | 0 |

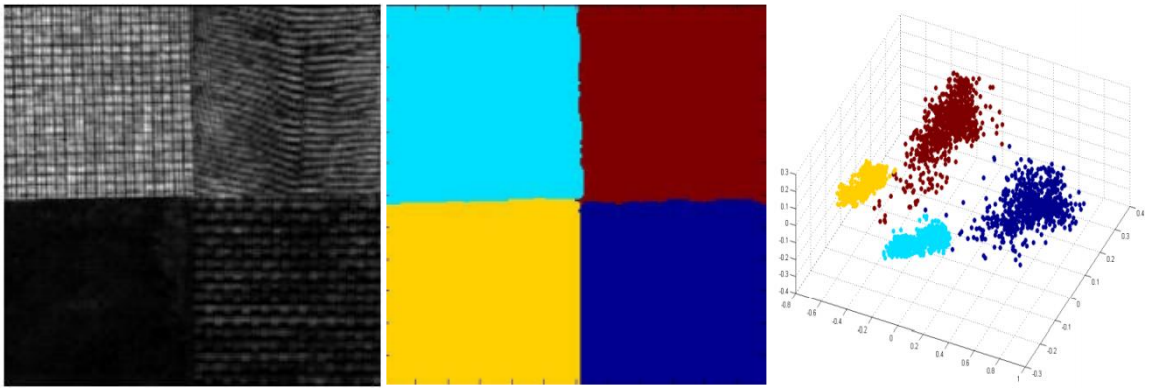
| | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 0 | 0.067 | 0.067 | 0.067 | 0.017 | 0.017 | 0.000 |
| 1 | 0.067 | 0.033 | 0.050 | 0.017 | 0.000 | 0.050 |
| 2 | 0.067 | 0.050 | 0.033 | 0.050 | 0.000 | 0.017 |
| 3 | 0.017 | 0.017 | 0.050 | 0.000 | 0.033 | 0.017 |
| 4 | 0.017 | 0.000 | 0.000 | 0.033 | 0.033 | 0.017 |
| 5 | 0.000 | 0.050 | 0.017 | 0.017 | 0.017 | 0.000 |

*Figure 3: (a) GLCM matrix and (b) normalized GLCM matrix of the image given in Fig. 1*

Then the various textural features are computed using the GLCM such as the Angular Second Moment, Contrast, Correlation, Variance, Inverse Difference Moment, Sum average, Sum variance, Sum entropy, Entropy, Difference Variance, Difference Entropy, the 2 Information Measure of Correlation and Maximal Correlation Coefficient. These textural features are computed for all the GLCMs (by varying d and Θ) produced for a given image. Then the mean and range of each of these measures averaging over all GLCMs are fed into feature segmentation algorithms and classifiers for further processing [4]. Since, some of these features might be strongly correlated, feature selection methods like Principle Component Analysis can be used to reduce the dimensionality of the feature set [5][7].

## 2. Texture Segmentation

The purpose of texture segmentation is to distinguish image curves that separate different textures. To segment textured images, one should initially be able to discriminate textures and isolate them based on textural properties [6].



(a) (b) (c)

*Figure 4: (a) Raw image having four regions of different textures (b) Regions of different textures identified using image segmentation based on textural features and (c) Clustering of image data points in a three-dimensional feature space to differentiate different clusters of regions [7].*

There are various segmentation strategies which are based on region growing, estimation theory – maximum likelihood, split-and-merge, Bayesian classification, probabilistic relaxation – an iterative approach for using context information to reduce local ambiguities, Clustering, artificial neural networks [1]. Two approaches out of the above mentioned are discussed in this article [1].

### Clustering-based Segmentation

The clustering-based techniques focus on finding homogenous groups or clusters of pixels or regions which have similar textural features. These clustering algorithms focus on finding clusters such that intra-class distance is minimized and inter-class distance is maximized between the clusters in the feature space. Hence, the regions with similar textural characteristics would lie in the same cluster and therefore, be marked as belonging to a single texture. One of the most common clustering techniques is K-means clustering which focuses on creating k centers around which data points are clustered [1][9].

### Artificial Neural Networks(ANN) based Segmentation

The use of ANN, especially Deep Neural Networks based models for Computer Vision have been on a rise and can also help in textural segmentation. Convolutional Neural Networks (CNNs) can be used for effectively extracting different textural regions of the image, even without the help of handcrafted features extracted from the images. This is because CNNs consist of filter-like layers which when trained effectively can extract meaningful features on their own and thus perform textural segmentation discriminating various textures in the image [8][15][16].

# 3. Texture Classification

It is the classification of different regions of an image based on the different textures. Texture classification process includes two stages: the learning stage and the recognition stage. In the learning stage, the objective is to build a model for the texture content of each texture class present in the training data, which generally comprises images with known class labels. The texture content of the training images is captured with a suitable texture analysis method, which yields a set of textural features such as textural properties of the images, such as spatial structure, contrast, roughness, orientation, etc., for each image. The extracted features are then fed into a supervised machine learning algorithm to train it for the recognition stage. In the recognition stage, firstly, the texture content of the given unknown sample is described with the same texture analysis method. Then the textural features of the given sample are compared to those of the training images by using a supervised classification algorithm and the sample is then assigned to the category with the best match.

The following supervised machine learning algorithms can be used for the classification of different textures:

KNN Classifier: In KNN, the classification of an object relies on the majority vote of its neighbors, with the object being assigned to the category most common amongst its k nearest neighbors, where k is a positive integer and typically small [14].

SVM Classifier: The objective of the SVM algorithm is to identify and differentiate complex patterns or categories of objects in a particular class. SVM is used over other methods because SVM can be numerically determined and simpler to analyze theoretically [13].

ANN Classifier :ANN Classifier focuses on extracting various components of an image using a multi-layered neural network. The image is passed on to the neural network which keeps on extracting different abstract features of the image layer-by-layer and then assigns the category to the image based on these smaller features [12].

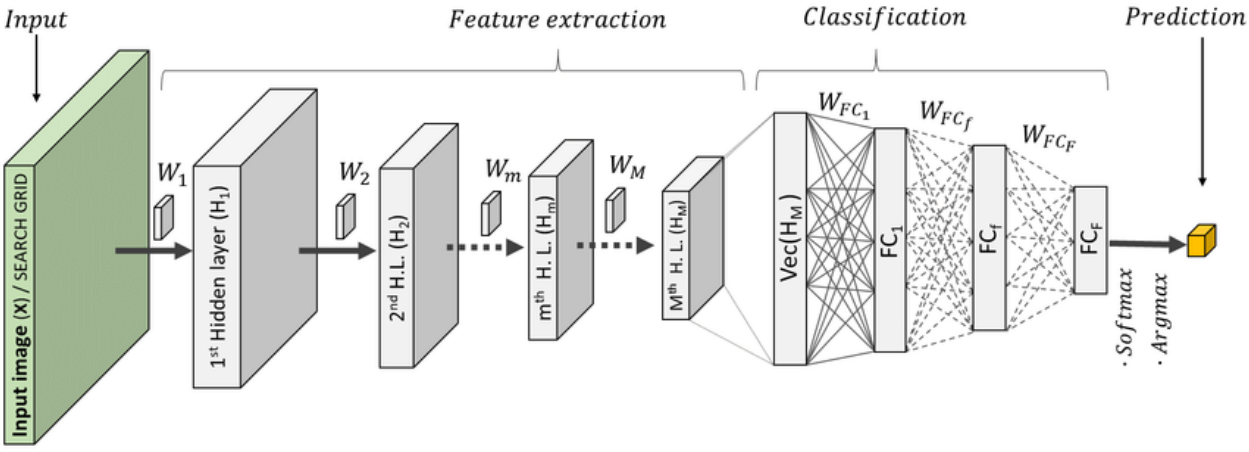



*Figure 5: Layer by layer feature extraction and classification in Artificial Neural Networks [16]*

Naive-Bayes Classifier: Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. The crux of the classifier is based on the Bayes theorem [14].

# 4. Reconstruction of 3D Surface Geometry

3D Reconstruction is the procedure that includes catching the appearance of the textures. It may be done in two different ways either in dynamic/active mode or latent/passive mode:

### Active Methods

This technique called go information strategies gives a 3D reconstruction of the item, furnished with a depth map (an image or image channel that contains data identifying the separation of the surfaces of scene objects from a viewpoint) by numerical guess approach by building the article in a situation which is dependent on the model. These strategies effectively meddle with the reproduced object, either precisely or radiometrically utilizing rangefinders to gain a depth map [11].

### Passive Methods

Passive techniques don't interfere with reconstructed objects; they just utilize a sensor to quantify the brilliance reflected or transmitted by the item's surface to induce its 3D structure through picture understanding. Monocular signs techniques and Binocular sound system vision are two of the most well-known strategies utilized in Passive strategy for 3D reconstruct [11].

### Monocular cues Method

This method uses several images from a viewpoint for 3D construction, it uses 2D features to measure 3D shape, 3D reconstruction through monocular cues methods is simple and quick as only one camera is needed. Monocular cues method can be used in three forms Shape-from-shading, Shape-from-shading, Shape-from-texture [11].

### Binocular stereo vision

At the point when binocular sound system vision is utilized, different pictures are gathered to get the 3-dimensional geometric data about an item. Two cameras are utilized all the while to gather the pictures from various points, or one camera is utilized to take numerous photos from alternate points of view [11].

# Conclusion

In this paper we discussed different texture analysis techniques for image processing, Four major areas on image texture analysis are discussed. Various applications of this area are mentioned along with their techniques used. In future, we plan to use these techniques on a real-life dataset.

*References*

[1] Materka, A., & Strzelecki, M.. Texture analysis methods–a review. Technical university of lodz, institute of electronics, COST B11 report, Brussels, 10(1.97), 4968. (1998).

[2] Vadakke Veettil, Bino & Unnikrishnan, Avittathur & Balakrishnan, Kannan. Gray Level Co-Occurrence Matrices: Generalisation and Some New Features. International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2(2012).

[3] Khan, A., Poonacha, K.N., Britt, A., McSweeney, B., Santos, N., Donovan, G., Yurk, V.M., & Scully, T.A.. Statistical Analysis of GLCM Texture Features and Microstructures in SEM Images of Crassostrea virginica Exposed to Atrazine. Proceedings of 11th International Conference on Bioinformatics and Computational Biology, vol 60, pages 170-180 (2019).

[4] Eduardo A.B. da Silva, Gelson V. Mendonça, The Electrical Engineering Handbook, (2005).

[5] Haralick, R. M., Shanmugam, K., & Dinstein, I. H.. Textural features for image classification. IEEE Transactions on systems, man, and cybernetics, (6), 610-621 (1973).

[6] Turner, M.R. Texture discrimination by Gabor functions. Biol. Cybern. 55, 71–82 (1986).

[7] Rampun, Andrik & Strange, Harry & Zwiggelaar, Reyer.Texture Segmentation Using Different Orientations of GLCM Features. ACM International Conference Proceeding Series. 10.1145/2466715.2466720. (2013).

[8] Andrearczyk, V., & Whelan, P. F. Texture segmentation with fully convolutional networks. arXiv preprint arXiv:1703.05230 (2017).

[9] Kaur, Dilpreet, and Yadwinder Kaur. Various image segmentation techniques: a review. International Journal of Computer Science and Mobile Computing 3.5 : 809-814 (2014).

[10] Tyagi, Vipin. Understanding digital image processing. CRC Press (2018).

[11] Moons, Theo, Luc Van Gool, and Maarten Vergauwen. 3D Reconstruction from Multiple Images: Principles. Now Publishers Inc. (2009).

[12] Ashour, Mohammed W., Khaled M. Mahar, and Mahmoud F. Hussin. "Artificial Neural Network for Texture Classification Using Several Features: A Comparative Study." The 2007 International Arab Conference on Information Technology (ACIT'2007) (2007).

[13] Kim, Kwang In, et al. "Support vector machines for texture classification." IEEE Transactions on Pattern Analysis and Machine Intelligence 24.11: 1542-1550 (2002).

[14]Armi, Laleh, and Shervan Fekri-Ershad. "Texture image analysis and texture classification methods-A review." arXiv preprint arXiv:1904.06554 (2019).

[15] P. G. H. H. Gunasekara, J. V. Wijayakulasooriya and H. A. C. Dharmagunawardhana. Image texture analysis using deep neural networks. IEEE International Conference on Industrial and Information Systems (ICIIS), Peradeniya (2017).

[16] Avalos, Sebastian & Ortiz, Julián. Geological modeling using a recursive convolutional neural networks approach. arXiv preprint arXiv:1904.12190 (2019).

[17] Tom McReynolds, David Blythe, Chapter 5 - Texture Mapping, In The Morgan Kaufmann Series in Computer Graphics, Advanced Graphics Programming Using OpenGL, Morgan Kaufmann, Pages 73-102 (2005).

[18] Chih-Jen Lee and Sheng-De Wang, "Fingerprint feature extraction using Gabor filters," in Electronics Letters, vol. 35, no. 4, pp. 288-290,(1999).

[19] Barbu, Tudor. "Gabor filter-based face recognition technique." Proceedings of the Romanian Academy 11.3 277-283 (2010).

[20] Dongxiao Zhou. Texture Analysis and Synthesis Using a Generic Markov-Gibbs Image Model. University of Auckland (2006).

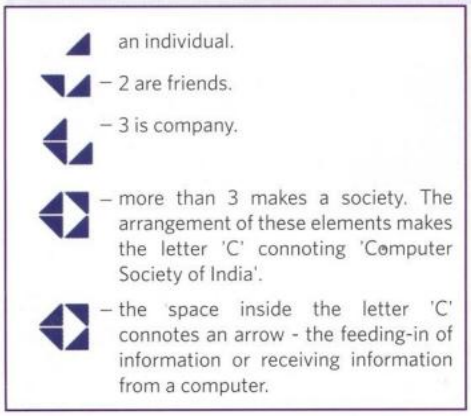
[21] Madasu, Vamsi Krishna, and Prasad Yarlagadda. "An in depth comparison of four texture segmentation methods." 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007). IEEE (2007).

## Compiled by


- Mr. Aayush Khattar is a student of B.Sc. (H) Computer Science, II year at Shaheed Sukhdev College of Business Studies, University of Delhi. He can be reached at aayushkhattar@acm.org.
- Mr. Anshuman Gupta is a student of B.Sc. (H) Computer Science, II year at Shaheed Sukhdev College of Business Studies, University of Delhi. He can be reached at anshumangupta1@acm.org.
- Ms. Vaishnavi Verma is a student of B.Sc. (H) Computer Science, II year at Shaheed Sukhdev College of Business Studies, University of Delhi. She can be reached at vaishnaviverma@acm.org.


## Under the guidance of

Dr. Anamika Gupta, Associate life member of CSI (CSI: 00118185), Sr. Assistant Professor, S.S. College of Business Studies, University of Delhi. She has more than two decades of teaching and research experience in Machine learning, image processing, data science with several research publications in reputed conferences and journals. She can be reached at anamikargupta@sscbsdu.ac.in.




**CSI Headquarter :**
Samruddhi Venture Park, Unit No. 3, 4th Floor, MIDC, Andheri (E), Mumbai-400093
Maharashtra, India
Phone : 91-22-29261700
Fax : 91-22-28302133
Email : hq@csi-india.org

**CSI Education Directorate :**
CIT Campus, 4th Cross Road, Taramani,
Chennai-600 113, Tamilnadu, India
Phone : 91-44-22541102
Fax : 91-44-22541103 : 91-44-22542874
Email : director.edu@csi-india.org

**CSI Registered Office :**
302, Archana Arcade, 10-3-190,
St. Johns Road, Secunderabad-500025,
Telengana, India
Phone : 040-27821998

# CSI ADHYAYAN

Published by
Prof. A. K. Nayak
Director
Indian Institute of Business Management, Patna
For Computer Society of India

Chair -Publication Committee
Dr. D. D. Sarma
Chief Scientist (R), CSIR-NGRI, Hyderabad

Chief Editor
Prof. Vipin Tyagi
Jaypee University of Engineering and Technology
Raghogarh, Guna (MP)

Editor
Dr. Vishal Jain
BVICAM
Delhi