

## Undergraduate Thesis Topic Allocation

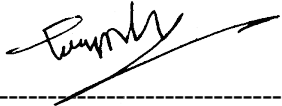
All the prospective students of L4/T1 of January 2022 term are advised to look at the thesis topics published by the different thesis supervisors (See below). The students are encouraged to make a list of several topics & supervisors that interest them. The allocation of topics will be done according to the merit order of the students. Students who have completed 108Cr. By July 2022 term will be eligible for thesis allocation.

An online meeting to allocate the thesis topics will be held on May 16<sup>th</sup>, 2022 (Monday) at 7:00 PM. The Zoom meeting link is:

<https://bdren.zoom.us/j/63702265241?pwd=S0k3ampaNU5kZjVuNlhVMFhJVVhZEZz09>

Meeting ID: 637 0226 5241

Password: 729605



---

Nafis Irtiza Tripto  
Assistant Professor & BUGS Secretary  
Department of CSE, BUET

Topic	Description	Students	Teacher ID
Knowledge Acquisition and Interrogation using Natural Language	<p>We will want to build an ontological infrastructure or knowledge graph for knowledge representation as well as the associated processes to automatically acquire knowledge from texts and reason over it.</p> <p>We will start with <a href="#">this base document</a>.</p>	1-2	Dr. Muhammad Masroor Ali
Towards Semantic Datums for Clinical Information	<p>We will try to find a way to represent clinical information using semantic datums or knowledge graphs.</p> <p>We will start with <a href="#">this base document</a> (Page 49).</p>	1-2	Dr. Muhammad Masroor Ali
Towards Better Annotation of Newspaper Articles using Knowledge Graphs	<p>We will try to find a way to annotate the newspaper articles using knowledge graphs.</p> <p>We will start with <a href="#">this base document</a> (Page 67).</p>	1-2	Dr. Muhammad Masroor Ali

<p>Algorithms and Computation (Exact topics will be fixed by discussing with the students)</p>	<p>Along with the penetration of ICT in society and the advance and spread of sensors, measurement instruments and observation equipment for gathering information in the real world, the amount of data obtained from various fields has grown exponentially and continues to become more diverse and more frequent in occurrence. This data is termed as “big data.” How to handle big data is a very important issue in computer science. In the theoretical area, developing efficient algorithms for handling big data is an urgent task. In the areas like web graph and social networks analysis, large scale protein interaction analysis, evacuation planning etc., where algorithmic techniques are needed to apply on big data, traditional efficient polynomial time algorithms are not useful. For example, taking input of a graph with 1 peta vertices and 2 peta edges takes 3.5 days with a 10 BIPS computer, and finding a maximum matching will take .2 million years. Thus design of efficient fundamental algorithms for property testing, matching, visualizing, substructure enumeration etc. in the context of big data is necessary for dealing with future information technology.</p> <p>Most real-world data sets are relational, which can be modeled as graphs, consisting of vertices and edges. Algorithms for automated drawings of graphs are the fundamental algorithms behind all sorts of visualization of data sets. In the last few decades graph drawing algorithms were primarily used for VLSI design automation where the underlying graphs were planar graphs, and hence structural properties and fundamental algorithms for planar graphs have been studied extensively [1,2]. However, most of the real-world graphs are non-planar and dynamic. In particular, many scale-free networks, which can be used to model web-graphs, social networks and biological networks consist of sparse non-planar graphs. To analyze and visualize such real-world networks we need to solve fundamental mathematical and algorithmic research questions on non-planar graphs and dynamic graphs. In recent years researchers have concentrated their attention in this area and several</p>	<p>7</p>	<p>Dr. Md. Saidur Rahman</p>
----------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------	------------------------------

week-long workshops and symposiums have been organized to define and formulate these research problems [3,4,5].

Generating networks with a desired degree distribution is an important topic of research in network science. Often network researchers need to perform empirical experiments on various types of networks and gain valuable observations. Synthetic network generators can help a lot here by reducing loads on real world data. Among various degree distributions, the power law degree distribution found in scale-free networks is the most common. Albert and Laszlo Barabási have mentioned a number of models for generation of scale-free networks [6]. However, these models don't impose any restriction on generated networks to satisfy the property of planarity, connectivity etc. Thus there is a need for developing algorithms for generating scale-free networks satisfying desired graph theoretic properties.

In analyzing big data, it is often required to clarify small and middle size structures in big data so that complicated big data will become small understandable data [7]. Data analysis such as machine learning becomes easy to design, visualization becomes clear, and the solutions will have connections to human understanding. Particlization discloses latent structures in customer data, economy data, and social networks, and supports innovations in new businesses on information technologies through useful data analysis. To deal with relational data like graphs, graph neural networks may give better performance over other machine learning approaches.

Dealing with big data is a real *challenge* in the upcoming 4th industrial revolution. Many existing algorithmic techniques will not perform well in handling big data. Finding new algorithmic paradigms to cope with these challenges is a great need. This research project aims to solve fundamental mathematical and algorithmic research problems that will evolve

while dealing with big data. These problems include characterizing the classes of non-planar graphs, development of efficient algorithms for automatic drawing of large non-planar graphs, for generating scale-free networks etc. Developing and improving the techniques for analyzing useful latent structure inside big data using machine learning approaches such as graph neural networks is also an important focus of this research project.

#### References

1. T. Nishizeki and M. S. Rahman, Planar Graph Drawing, World Scientific, 2004.
2. M. S. Rahman and M. R. Karim, Drawing planar graphs, Proc. of WALCOM 2020, Lecture Notes in Computer Science, 12049, pp. 3-14, 2020.
3. S. Hong, M. Kaufmann, S. G. Kobourov, and J. Pach, Beyond-Planar Graphs: Algorithmics and Combinatorics, Report from Dagstuhl Seminar 16452 (2016), Dagstuhl Reports, Vol. 6, Issue 11, pp. 35–62, 2017  
([https://drops.dagstuhl.de/opus/volltexte/2017/7038/pdf/dagrep\\_v006\\_i011\\_p035\\_s16452.pdf](https://drops.dagstuhl.de/opus/volltexte/2017/7038/pdf/dagrep_v006_i011_p035_s16452.pdf))
4. S. Hong and T. Tokuyama (Eds.), Beyond Planar Graphs, Communications of NII Shonan Meetings, Springer, 2020.
5. A. Meidiana , S. Hong and Peter Eades. New Quality Metrics for Dynamic Graph Drawing. Proc. of Graph Drawing 2020, LNCS, 12590, pp. 450-465, 2020.
6. A. Barabási, Network Science, Cambridge University Press, 2016.
7. P. Mutzel, Algorithmic Data Science, 36th International Symposium on Theoretical Aspects of Computer Science (STACS 2019), LIPICS, 2019.

Artificial Neural Networks	Applying neural network based systems for medical diagnosis including (e.g. COVID, Dengue and water borne diseases)	2	Md. Monirul Islam (mdmonirulislam)
NLP	Devising machine learning based framework for NLP	2	Md. Monirul Islam (mdmonirulislam)
Nature Inspired Algorithms	Focus of this work will be applying (and modifying in some cases) nature inspired algorithms for solving practical problems such as fraud detection, traffic optimization, etc.	2	Md. Monirul Islam (mdmonirulislam)
A classification framework using machine learning for Titanium FPGA family processors	Titanium FPGA family has the parallel processing capability for edge computing machines. It reduces the bandwidth requirements as the classification is possible in the edge devices in the IoT applications. The main purpose of this research is to develop a framework for this classification problem so that any practical problem can be easily mapped to this generalized framework. For further information regarding Titanium FPGA family please read the following link: <a href="https://www.eetimes.com/fpga-comes-back-into-its-own-as-edge-computing-and-ai-catch-fire/">https://www.eetimes.com/fpga-comes-back-into-its-own-as-edge-computing-and-ai-catch-fire/#</a>	2	Dr. Md. Mostofa Akbar
Summarizing the video description for the visually impaired persons in the context of	A visually impaired person might get help while walking on the roads of a busy city like Dhaka. A camera can capture the video in front of a person and detect the images automatically. But the main problem is reporting the important objects. This will be notified through voice using text to speech converter. The main problem is to establish a strategy for identifying important object for the visually impaired person. Reporting too many objects	1 or 2	Dr. Md. Mostofa Akbar

<p>busy road of Dhaka</p>	<p>through sound makes a person difficult to follow. It is also great to find some pther techniques for notification such as sending signal to hand or leg.</p>		
<p>Automated License Plate Recognition System for Toll Plaza</p>	<p>Detail Design of an Automated License Plate Recognition System using feed from high resolution traffic IP Camera with the following features</p> <ul style="list-style-type: none"> <li>a. Support for Scalable Multiple concurrent feeds</li> <li>b. Able to locate, detect and crop license plates</li> <li>c. Support for Bangladesh License Plate Recognition</li> <li>d. Support for Multiple Vehicles in a single frame</li> <li>e. Support for classification of vehicle in broad categories</li> </ul> <p>This project is for a software development company and the interested students will be able get some finacial incentive if they can cotribute to any of the above mentioned problem in a regular basis.</p>	<p>2</p>	<p>Dr. Md. Mostofa Akbar</p>
<p>Blockchain Based Clinical Decision Support System</p>	<p>Patient-records from multiple sources will be linked using the blockchain technology and decision support system will be designed and implemented using the linked records</p>	<p>2</p>	<p>Dr. Abu Sayed Md. Latiful Hoque</p>

<p>Efficient Design and implementation of Blockchain database for privacy-preserved medical records</p>	<p>Blockchain technology has been designed basically for crypto currency. The performance is the main issue of blockchain in transaction processing.</p>	<p>2</p>	<p>Dr. Abu Sayed Md. Latiful Hoque</p>
<p>Big Data Analytics of Health Data</p>	<p>Health data analytics can be used to improve health services. The challenges are data integration, standardization, and noise reduction in data.</p>	<p>2</p>	<p>Dr. Abu Sayed Md. Latiful Hoque</p>
<p>Privacy and security for Healthcare Blockchains</p>	<p>Healthcare blockchains provide an innovative way to store healthcare information, execute healthcare transactions, electronic healthcare records and build trust for healthcare data sharing and data integration in a decentralized open healthcare network environment. Although the healthcare blockchain technology has attracted broad interests and attention in industry, government and academia, the security and privacy concerns remain the focus of debate when deploying blockchains for information sharing in the healthcare sector from business operation to research collaboration.</p>	<p>2</p>	<p>Dr. Mohammad Mahfuzul Islam</p>
<p>Steganography</p>	<p>Steganography is the way of hiding information inside data or image so that any intruder cannot understand whether information is being transmitted or not. In cryptography concept, information is encrypted and kept it secret to the intruders so that the intruder cannot understand what data is being transmitted. However, the intruder still know information is being transmitted and the intruder still can damage the information even if he/she cannot</p>	<p>2</p>	<p>Dr. Mohammad Mahfuzul Islam</p>



	<p>understand what is the actual information. Steganography is the methodology that kept information secret and the intruder can know nothing about its transmission. This is an emerging area of research in protecting privacy and security.</p>		
<p>Vehicular ad hoc networks</p>	<p>Intelligent transportation system realized through vehicular ad hoc network (VANET) where vehicles exchange information and content with each other and with the infrastructure. Up to date, excellent research results have been yielded in the VANET domain aiming at safe, reliable, and infotainment-rich driving experience. However, due to the dynamic topologies, host-centric model, and ephemeral nature of vehicular communication, various challenges are faced by VANET that hinder the realization of successful vehicular networks and adversely affect the data dissemination, content delivery, and user experiences.</p>	2	<p>Dr. Mohammad Mahfuzul Islam</p>
<p>5G Wireless Technology</p>	<p>The most challenging part of implementing 4IR is the shortage of wireless bandwidth. 5G technology gives 70 times more bandwidth compared with 4G. This is not only the bandwidth, with the implementation of 5G, a mobile phone will be as much powerful as super-computer because mobile will serve as an interfacing clients and all the computations will be carried out in the cloud. Fom effective implementation of 5G and getting benefits from it, new protocols need to be designed and requires address lots of reseach challenges.</p>	2	<p>Dr. Mohammad Mahfuzul Islam</p>
<p>Quantum Computing</p>	<p>We want to study quantum computing under this project and advance the current state of the art. This will be a learning experience for both the supervisor and supervisee.</p>	1 or 2	<p>Dr. M. Sohel Rahman</p>

<p>Bioinformatics and Computational Biology</p>	<p>We want to explore various problems in BI and CompBio. Topics include but are not limited to: Single cell transcriptomics, Single cell and bulk methylation, Epigenetic clocks, Spatial transcriptomics, Drug efficacy prediction. You may check these topics by searching google to get an idea. But don't get discouraged if you feel lost with the concepts/terminology (I also belong to your group:)). We will use various computational techniques; ML/DL is expected to play a strong role here.</p> <p>I am providing some basic description of the topics mentioned above:</p> <p>Single-cell (sc) transcriptomics: An sc dataset captures the activity level of every gene in the individual cells of a tissue. The datasets are “big” (~100K cells, ~20K genes) and sparse (i.e., the captured activity levels of most genes in a cell are zero). We aim to build new, improved models for sc data integration, imputation, clustering, and modeling temporally resolved sc data.</p> <p>Spatial transcriptomics (ST): ST is a new technology that combines single-cell gene activity measurements with each cell’s spatial location. We aim to develop new methods for ST data segmentation, integration, and modeling temporally resolved ST data.</p> <p>Sequence-to-methylation modeling: DNA Methylation refers to the modification of DNA nucleotides when a methyl-molecule attaches to them. DNA methylation at “wrong” places of the genome leads to diseases like cancers and Alzheimer’s. We will review and improve the current models for predicting how DNA mutations change DNA methylation status.</p> <p>Methylation clocks: DNA methylation of the genome is predictive of a person’s biological age (in contrast to their “calendar” age). A 25-year-old person with heart disease, for example, may have a heart that is functionally similar to the heart of an 80-year-old person. In this case, his biological age is 80 years, although the calendar age is 25. This project will review</p>	<p>Any</p>	<p>Dr. M. Sohel Rahman</p>
-------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------	--------------------------------

	<p>and improve the current methylation clock models and apply them to single-cell methylation data.</p> <p>Drug efficacy prediction: We aim to develop new models to predict the efficacy and toxicity of drug molecules.</p>		
Metaheuristics Optimization	We want to tackle various optimization problems in different branches of science and engineering using metaheuristics techniques. No prior knowledge is necessary.	Any	Dr. M. Sohel Rahman
On Devising Genetic Algorithm for Efficient Solution of Discrete Optimization Problems	<p>Many real-world discrete optimization problems are NP-hard and difficult to solve using traditional linear and convex optimization approaches. Furthermore, the time complexity of these optimization problems grows exponentially with the number of decision variables. Another difficulty that might arise is that the search space is intrinsically multimodal and non-convex. In such a case, an efficient optimization method that can deal with these problem characteristics is necessary. The genetic algorithm (GA) is a popular tool for solving DOPs. The original GA and its variants have been used to a variety of classic discrete optimization problems. Literature shows that the static mutation probability is commonly used for the GA and its variants which cause the imbalance between exploration and exploitation, limiting the performance of GA. To overcome this problem, in this research we shall explore time-varying mutation operator for GA. A balance between exploration and exploitation of the devised GA will be verified using the benchmark instances of a well-known combinatorial optimization problem i.e., the 0-1 knapsack problem.</p>	2-3	Dr. A.K.M. Ashikur Rahman

<p>On DoS and DDoS attacks in routing protocols of wireless multi hop networks.</p>	<p>Mobile ad-hoc networks (MANET) are the most widely used ad-hoc wireless communication networks. MANETs are decentralized wireless networks that communicate without pre-existing infrastructure. This is an art of networking without any network. MANET can function as standalone or can be connected to external networks. The security of the ad-hoc network is most challenging issue. Denial-of-service attacks and distributed denial-of-service attacks are most common attacks in almost every layer of the networks. In this thesis we will be focusing on routing layer and explore the possibilities and counter measures of DoS and DDoS attacks in such networks.</p>	<p>2-3</p>	<p>Dr. A.K.M. Ashikur Rahman</p>
<p>Smile and beyond: impact of smile on face recognition accuracy</p>	<p>Have you ever wondered what made Shahrukh Khan (SRK) very special and the king of Bollywood? SRK has the best smiling face of the world that could even move mountains. Anyone can easily recognize SRK simply by looking at his smile. In fact, using smile rather than neutral images might bring significant improvements in face matching accuracy. In this research at first, we shall review the state-of-the-art techniques to identify smile in facial images. Then we will be building a complete end-to-end solution that can detect smiles in a video stream in real-time using deep learning along with traditional computer vision techniques. Finally, we shall find the impact of smile on face recognition accuracy in depth.</p>	<p>2-3</p>	<p>Dr. A.K.M. Ashikur Rahman</p>
<p>Applied Machine Learning -1</p>	<p>Study of different GAN models and application.</p>	<p>Maximum 2</p>	<p>Dr. Mahmuda Naznin</p>
<p>Applied Machine Learning-2</p>	<p>Applied machine learning in the domain of cognitive science.</p>	<p>Maximum 2</p>	<p>Dr. Mahmuda Naznin</p>

Data Collection and Analysis	Design and development of an android based tool to facilitate capturing geo diversified multilingual medical data.	1 or 2	Dr. Mahmuda Naznin
------------------------------	--------------------------------------------------------------------------------------------------------------------	--------	--------------------

Different Topics on Image processing and computer vision

Image processing and computer vision are both vast areas of research. Below pls find some of their major branches. Some of the areas also use machine learning algorithms. Actual topics will be selected after consultation with me based on your interest and capability.

**Image & video search and retrieval:** this includes feature extraction, image representation and retrieval algorithm design.

**Security and surveillance:** there are many biometric and non-biometric applications like face, iris, finger print, palm print recognition, crime scene analysis

**Medical imaging:** automatic segmentation, detection and classification of anomalies, like tumors and other disease patterns in X-ray, CT, MRI, PET, angiogram, ECG, echo, endoscopy images

**Document Analysis/Classification/Preservation:** Character recognition, document binarization Object and shape recognition, historical document enhancement, land map digitization, etc

**Intelligent transport system:** license plate detection/recognition, road sign detection/recognition, etc

**Environmental monitoring and remote sensing/satellite imaging:** land usage determination, changes in land usage, satellite image enhancement/de-noising

1 to 6

Dr. Md. Monirul Islam (mmislam)

**Image enhancement:** improve the visual quality of images using different methods

**Image de-noising:** removing noises from images

**Image segmentation:** dividing images into different semantically meaningful regions, e.g., object, etc

**Image Security:** this is not human's security; rather the security of image itself. Some areas are: watermarking and registration, image tampering/forgery detection, forgery detection of bank notes, etc

**Image compression:** represent images compactly for transmission and storage.

**Image classification:** too many applications, e.g., disease classification, OCR, leaf classification, crop disease classification

**Some of the specific titles (but not restricted to):**

**Deep learning model for feature extraction for image /video retrieval and classification, 2-3 students**

**Land map digitization 2-3 students**

	<p><b>Brain/lunge tumor detection and classification from MRI/CT images 2-3 students</b></p> <p><b>Mathematical model for detection of image tampering/forgery 2-3 students</b></p> <p><b>Plant/crop disease classification from leaf images 2-3 students</b></p> <p><b>Plant identification from leaf databases 2-3 students</b></p> <p><b>Image processing solution for authenticity detection of paper currencies 2-3 students</b></p>		
--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	--



<p>Intrusion detection in containerized environments</p>	<p>With the growing popularity of cloud technologies and particularly, containerized environments, it has been significantly important to protect them from various cyberattacks especially because of their heavy usage in critical infrastructures including 5G networks. Additionally, many unique natures of these new technologies, such as large-scale, dynamicity, non-persistence, very often make the traditional security solutions infeasible for them.</p> <p>This project will aim at addressing unique challenges in containerized environments and building a robust and efficient intrusion detection in containerized environments. Furthermore, the project will target to enhance the proposed method to include the intrusion tolerance capability which will allow to continue the business and mitigate the impact of cyberattacks in a containerized environment. Particularly, the project will conduct the following research tasks:</p> <ul style="list-style-type: none"> <li>● Study on security and other practical challenges related to containerized environments.</li> <li>● Deploy a testbed of a real-world container environment (e.g., Kubernetes/Docker).</li> <li>● Review the literature on intrusion detection in containers and identify the existing research gaps.</li> <li>● Design and implement an enhanced solution by considering a richer context and feature set.</li> <li>● Study container image-level diversification approaches and propose an automated diversification solution to keep the attack factor constant and prevent from parallel attacks.</li> <li>● Study whether/how container image diversification would improve the detectability of the proposed solution</li> </ul>	<p>2</p>	<p>Dr. Md. Shohrab Hossain</p>
----------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------	--------------------------------

	<p>This is a collaborative research work with Universitt of Ottawa, Concordia University, Canada and KAUST.</p>		
<p>Automated detection of web measurement tools like Selenium or OpenWPM</p>	<p>The goal is to do a differential analysis of normal web browsing behaviors against automated tools. Such analysis is useful to detect bots and frauds. The research activity would involve creating or reusing existing web automation tools and comparing them against real browser sessions and identifying ways to automatically detect if a given browser session is being generated by a real browser versus an automated bot.</p> <p>This is a collaborative research work with Department of Computer Science, North Carolina State University, USA.</p>	2	Dr. Md. Shohrab Hossain
<p>Detecting user activity from encrypted IoT traffic</p>	<p>The goal here is to develop automated techniques to determine user activities on IoT devices from encrypted network traffic. For this, we can utilize some open-sourced IoT data (some with labels). This is a collaborative research work with Department of Computer Science, North Carolina State University, USA.</p>	2	Dr. Md. Shohrab Hossain
<p>Assessing Impacts of Heat through Computing Technologies</p>	<p>Heat exhibits different impacts on human beings – covering both physiological and psychological cases. This is of particular importance in the case of Dhaka city owing to excessive heat exposures to its dwellers. This becomes even more important for outdoor workers such rickshaw pullers, daily labors, etc.</p> <p>In this thesis, our goal would be to investigate the impacts of heat on the most-affected communities such as rickshaw pullers. Computing technologies such as sensors, wearable,</p>	1	Dr. A. B. M. Alim Al Islam

	<p>etc., will be used in the process of the investigation. Real experimentation, as well as real data collection, is expected to be carried out here. Expected team size for this thesis will be 1-2.</p>		
<p>Embedded Fault Tolerance</p>	<p>Fault tolerance is now being in widespread usage, however, mostly at the cost of high resource utilization. There remain many applications (such as embedded system based applications) that cannot afford such high resource availability. Thus, research on enabling fault tolerance sustaining resource constraints has recently emerged as a prominent research area.</p> <p>In this thesis, our goal would be to come up with effective mechanism(s) that can enable fault tolerance in embedded devices having resource constraints. Attaining this goal would require development of new light-weight mechanism(s) and their device-level experimentations. Expected team size for this thesis will be 1-2.</p>	<p>1</p>	<p>Dr. A. B. M. Alim Al Islam</p>
<p>Quantum Intelligence</p>	<p>Thinking classically is different from thinking quantumly, and so does in the case of intelligence. In the realm of emerging quantum computing era, AI needs to be redesigned in many cases leveraging the core strengths of quantum computing sustaining its limitations. Thus, quantum intelligence would be substantially different from classical AI.</p> <p>In this thesis, our goal would be to come up with new quantum intelligence algorithms that can enable artificial intelligence in real quantum computers having special circuits and connections. Attaining this goal would require development of new quantum intelligence</p>	<p>1</p>	<p>Dr. A. B. M. Alim Al Islam</p>

	<p>algorithm(s) and their experimentations in real quantum computers. Expected team size for this thesis will be 1-2.</p>		
<p>Simulating Road Traffic of Dhaka for Effective Policy Making</p>	<p>On-road experience in Dhaka city is perhaps one of the worst parts in lives of its dwellers. However, it is believed that policy making and their enforcement could substantially ameliorate the on-road experience. Attempting to do so directly in real settings may end up with backfiring consequences in case the enforced policy does not realize both micro-level and macro-level impacts, which are often very difficult to foresee. Here comes the role of simulation of to-be-attempted policies to understand their possible impacts even before their enforcements in real settings.</p> <p>In this thesis, our goal would be to come up with new effective on-road policies (such as making a road link unidirectional, removing slow vehicles from a link, etc.) that can engender better on-road experience. Attaining this goal would require development of new policies and their experimentations using a simulator named Dhakasim (developed in our research group). Expected team size for this thesis will be 1-2.</p>	<p>1</p>	<p>Dr. A. B. M. Alim Al Islam</p>

<p>Computing Solutions to Serve the Under-served</p>	<p>Digital divide has already segregated people between technology-literate and technology-illiterate sects. Such divide exhibits mourning consequences when a large part of whole population remains technology-illiterate as well as under-served from the perspectives of attaining the basic needs. This is even more prominent in the context of Bangladesh, where there remain many marginalized under-served communities (such as beggars, street children, blind people, etc.) that are mostly technology-illiterate. Thus, it is extremely challenging to devise technological solutions to serve these under-served communities.</p> <p>In this thesis, our goal would be to devise new technology-enabled solutions to facilitate serving some basic needs (food, shelter, education, etc.) of the marginalized communities. Attaining this goal would require development of new solutions following HCI-based studies such as VSD, participatory design, etc. Expected team size for this thesis will be 1-2.</p>	<p>1</p>	<p>Dr. A. B. M. Alim Al Islam</p>
<p>Future Networking Technology: 5G and beyond 5G</p>	<p>Different G's (2G, 3G, and 4G) have attained considerable successes paving the path of yet another "G", i.e., 5G, in near future, or it is already in. However, achieving success of the 5G, or beyond 5G, in specialized cases is still under research. An example of such cases is IoT, which demands a special treatment to 5G to get it rolling in its full potent in the system architecture of IoT. 5G security, 5G integration with SDN, etc., are some other related research areas.</p> <p>In this thesis, our goal would be to devise new protocol(s) or networking architecture to enable enhanced and secured performance of 5G or beyond 5G, perhaps in specific applications. Attaining this goal would require development of new protocol(s) or networking</p>	<p>1</p>	<p>Dr. A. B. M. Alim Al Islam</p>

	<p>architecture and experimentation through simulation and/or testbed implementations. Expected team size for this thesis will be 1-2.</p>		
<p>Virtualizing Land Aisles in Bangladesh</p>	<p>Disputes in lands are arguably the most frequent causes of legal cases and tensions, which span over a society, a neighborhood, or even a family. One of the main reasons behind these scenarios is the completely manual processing of land areas – in sketches and in real measurements. Here, a first step to overcome the scenarios could be virtualizing (or digitizing, or automating) the land areas.</p> <p>In this thesis, our goal would be to devise new computing solutions to virtualize the land areas in Bangladesh. This can start from simple Mouja maps. Here, developing solutions would require exploring cross-section of different domains such as image processing, machine learning, etc. Expected team size for this thesis will be 1-2.</p>	<p>1</p>	<p>Dr. A. B. M. Alim Al Islam</p>

<p>Code Entity Recognition from Natural Language and Code Amalgam using Code-NL Pretrained Models</p>	<p>Often in software development environments (e.g. code review, StackOverflow, Github) developers use a mixture of natural language and programming language to communicate with each other about source code issues. This creates ambiguity in text understanding and can confuse an NLP model.</p> <p>Automatically identifying these code tokens from a completely unstructured text could help automatic error localization and automatic bug repair with AI. Furthermore, it can also improve the readability of code-NL amalgam text to developers.</p> <p>Research Question:</p> <p>Can we automatically identify code tokens from code-NL amalgam using pretrained language models like PLBART or Code-T5?</p> <p>Data:</p> <p>Most StackOverflow question and answers allows user to provide a special tags to indicate code tokens inside a text. We can download a StackOverflow data dump, and use these special tokens for prediction. This gives us a huge corpus of annotated data.</p>	<p>2</p>	<p>Dr. Anindya Iqbal</p>
-------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------	--------------------------

<p>Detecting Android malware using machine learning techniques</p>	<p>Malware is one of the major issues regarding the Android operating system. We have seen Signature-based malware detection techniques in the literature. However, the techniques were not able to detect unknown malware. Despite numerous detection and analysis techniques are there, the detection accuracy of new malware is still a crucial issue. In this project, we will use different machine learning techniques to detect unknown malware.</p> <p>There will be collaboration with Dr. Shahrear Iqbal, Research Officer, Cyber Security at National Research Council, Canada in this project.</p>	<p>2</p>	<p>Dr. Anindya Iqbal</p>
<p>Advanced Persistent Threat (APT) detection using system provenance graphs</p>	<p>The term APT is used for a wide range of cyber-attacks executed over a long period of time to gain access to confidential information or to compromise high-integrity resources while remaining undetected. Recent works suggest that system provenance graphs are a more effective data source for APT detection. A system provenance graph is a directed acyclic graph (DAG) that represents causal relationships between running processes and objects (e.g., files, network flow, threads) in a system. It can connect events that are temporally distant but causally related. In this project, we will use the system provenance graph to detect advanced persistent threats. We will investigate how to convert this data to a numerical form (graph embedding) so that we can leverage machine learning algorithms.</p> <p>There will be collaboration with Dr. Shahrear Iqbal, Research Officer, Cyber Security at National Research Council, Canada in this project.</p>	<p>2</p>	<p>Dr. Anindya Iqbal</p>



<p>Commonsense Reasoning in Bangla</p>	<p>In recent years, commonsense knowledge and reasoning have received renewed attention from the NLP community. A collection of papers and resources can be found here (<a href="https://github.com/yuchenlin/awesome-commonsense">https://github.com/yuchenlin/awesome-commonsense</a>). The use of ConceptNet (<a href="https://conceptnet.io/">https://conceptnet.io/</a>) is also ubiquitous for commonsense reasoning. We will focus primarily on Bangla, where it is not yet explored.</p>	<p>1-2</p>	<p>Dr. Rifat Shahriyar</p>
<p>Study of Bias in Bangla Language Models</p>	<p>In recent years, detecting and mitigating bias have received significant interest from the NLP community. A collection of papers and resources can be found here (<a href="https://github.com/uclanlp/awesome-fairness-papers">https://github.com/uclanlp/awesome-fairness-papers</a>). We will investigate if Bangla language models learn bias during self-supervised training. We will study BanglaBERT, BanglaGPT, and BanglaT5 style models. We will focus on bias detection first and gradually explore bias mitigation approaches. Potential NLP applications include machine translation and question answering.</p>	<p>1-2</p>	<p>Dr. Rifat Shahriyar</p>
<p>Sublinear time algorithms for maximum inner product search (MIPS)</p>	<p>Maximum Inner Product Search (MIPS) is a standard computational task where we are given a set of vectors <math>V</math> and a query vector <math>q</math>, and the goal is to find the vector in <math>V</math> that has the maximum inner product with the query vector <math>q</math>. MIPS is extensively used in real-world systems like web search, information retrieval, and recommendation systems. Although linear-time algorithms are de facto for MIPS, they often fail to scale on very large datasets. In this work, we will explore sublinear time data structures and algorithms with provable guarantees for MIPS. Possible directions may include approximation, parallelization, sampling, etc. Strong competitive programming background is recommended.</p>	<p>1-2</p>	<p>Dr. Rifat Shahriyar</p>

<p>Reusable Garbage Collection Implementation with Rust</p>	<p>MMTk (<a href="https://www.mmtk.io/">https://www.mmtk.io/</a>) is a framework for the design and implementation of memory managers. MMTK was originally implemented using Java for Jikes RVM (<a href="https://github.com/JikesRVM/JikesRVM">https://github.com/JikesRVM/JikesRVM</a>). MMTK core (<a href="https://github.com/mmtk/mmtk-core">https://github.com/mmtk/mmtk-core</a>) is the Rust port of MMTK. MMTK core maintains three VM bindings for MMTk (OpenJDK, JikesRVM, and V8). All the bindings can use garbage collection (GC) algorithms implemented in the MMTk core. This work will focus on implementing any existing high-performance GC algorithm in MMTk Core with Rust and then, if possible, will target novel algorithm design.</p>	<p>1-2</p>	<p>Dr. Rifat Shahriyar</p>
<p>Data Analytics on Serverless Platform</p>	<p>Serverless computing has recently attracted a lot of attention from research and industry due to its promise of ultimate elasticity and operational simplicity. However, there is no consensus yet on whether or not the approach is suitable for data processing. In this thesis, we'll explore possible ways to perform data analytics on serverless computing by designing a serverless distributed data processing framework and tweaking serverless architecture. We'll try to answer several important technical questions that need to be solved to support data analytics on serverless platform which offers a cost and performance advantage over existing solutions.</p>	<p>Any</p>	<p>Dr. Muhammad Abdullah Adnan</p>

<p>Geo-Distributed Data Analytics and Machine Learning</p>	<p>The recent explosion of data volumes has reignited the focus on scale-out data analytics, and has fostered the world of Big Data systems. While these paradigms suffice for a single data center, we have reached a new inflection point where the combination of big and geographically distributed data requires new approaches for geo-distributed analytics processing and machine learning to minimize wide-area bandwidth costs. Centralized approaches together with heuristics such as data reduction or ad-hoc distributed querying may suffice in the short term. However, they are not sustainable as data volumes grow relative to transoceanic bandwidth and regulatory concerns become paramount.</p> <p>In this research project, we intend to develop an efficient training scheme for a machine learning algorithm that will work on Geo-distributed Big Data. Here, data are distributed among geographically separated data centers. Our goal is to train a distributed regression/support vector machine/neural network/etc. model that will use optimum communication (i.e. bandwidth) with the data centers as well as it will preserve the data privacy and converge quickly with a satisfactory accuracy.</p>	<p>Any</p>	<p>Dr. Muhammad Abdullah Adnan</p>
<p>Speech to Bangla Sign Language Generation</p>	<p>The goal of this thesis is to research and develop techniques to generate gesture in real time from Bangla speech. We will formulate a deep learning based model on Bangla Sign Language and develop a tool that will translate verbal Bangla speeches into the corresponding gestures of Bangla Sign Language in real time. We'll also show the effectiveness of generative models for gesture generation in avatars for out of vocabulary speeches. The outcome will be a portable application that will provide a bridge of communication between the hearing and hearing-impaired people.</p>	<p>Any</p>	<p>Dr. Muhammad Abdullah Adnan</p>

Computationally Efficient Natural Language Processing	This research is focused on making NLP models more efficient both at training and testing time. This work will enable applications of such models at scale and potentially make them directly runnable on mobile devices without losing the quality of models run on servers. Model compression techniques such as quantization, knowledge distillation, model pruning, etc. and more efficient, modular, sparse architectures may need to be explored.	Any	Dr. Muhammad Abdullah Adnan
Security in Machine Learning Systems	With machine learning being integrated into real-life Systems, new avenues of attacks have opened up for attackers. Attackers have been working hard to figure out all the potential attacks an ML system could fall victim to. Hence, in order to secure a machine learning system, Security must be built into the model from the very beginning of the machine learning lifecycle. From the start of development through the production stages the entire time the model is in use, the system must be actively secured. In this research, we'll explore and analyze security aspects of ML systems such as Data confidentiality, System manipulation, Adversarial examples, Transfer learning attack, Data poisoning, etc. and innovate new techniques for securing machine learning systems.	Any	Dr. Muhammad Abdullah Adnan

<p>Alignment-free estimation of phylogeny</p>	<p>Phylogeny refers to the evolutionary relationships among a set of entities. phylogeny represents the evolution of a particular gene within a group of species. The study of gene phylogeny not only helps identify the historical relationships among a group of organisms, but also aids in other biological research such as drug and vaccine design, protein structure prediction and so on. In sequence-based methods of phylogeny reconstruction, the input is a set of homologous sequences from different species. After obtaining an alignment of these sequences, different methods are applied to extract the phylogenetic information. These include distance-based methods, heuristics for either Maximum-Likelihood (ML) or Maximum-Parsimony (MP), and Bayesian Markov Chain Monte Carlo (MCMC). Examples of distance-based methods include Neighbor Joining (NJ), BIONJ, RapidNJ, FastME, QuickTree, Clearcut etc. RAxML and FastTree, the most popular tools of tree estimation from alignment to date, use heuristics for ML. Since all these methods rely on sequence alignment, the process therefore is time consuming. Also, any error in the alignment significantly affects the downstream processes, resulting in poor estimation of the gene tree.</p> <p>While the most popular distance-based methods compute the distance matrix from sequence alignment, it is technically possible to use these approaches without the alignment step, so long as the distance measure is able to reflect the number of substitutions per site, which underlies classical alignment-based phylogeny reconstruction. In fact, several such alignment-free phylogeny estimation methods are found in literature. In this work, we will review the existing alignment-free methods and objectively compare their performance. We will then propose new and improved alignment-free methods. We will compare our methods with state-of-the-art alignment free methods. We will also compare the performance of our methods against RAxML and FastTree.</p>	<p>1-2</p>	<p>Dr. Mohammad Saifur Rahman</p>
-----------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------	-----------------------------------

<p>Absent words to solve several problems in bioinformatics</p>	<p>An absent word with respect to a sequence is a word that does not occur in the sequence as a factor. A minimal absent word (MAW) is a word that is absent in a sequence but all its proper factors occur in that sequence. On the other hand, a relative absent word (RAW) is a word that occurs in a target sequence but is absent in a reference sequence. A RAW is minimal if none of its proper factors are RAW for the same pair of target and reference sequences. Perhaps the most significant use of absent words is in the field of bioinformatics. Recently, MAW has been used to compute the distance between two species. Similar effort has also been made to investigate the variation in number and content of MAWs within a species using four human genome assemblies. Silva et al. utilized RAW for differential identification of sequences that are derived from a pathogen genome but absent from its host. They applied this concept in analyzing Ebola virus genome from the 2014 outbreak and discovered the presence of short DNA sequences in the Ebola virus genome that appear nowhere in the human genome. The pathogen specific signatures identified from such analysis can be useful for quick and precise action against the infectious agents. MAW and RAW have also been used in gene phylogeny reconstruction.</p> <p>In this work, we will explore using absent words to solve several other problems in bioinformatics. Firstly, in protein attribute prediction problems, distance measures stemming from absent words could potentially be used during dataset preparation to avoid homology bias. CD-HIT is currently the most popular software in this regard. We will design experiments with different datasets from different protein attribute prediction problems to examine the efficacy of our proposed distance measures in reducing homology bias in any benchmark dataset.</p> <p>Secondly, for a collection of protein sequences, the pairwise distance matrix, based on absent words, can be computed and used as a custom kernel with SVM to produce prediction</p>	<p>1-2</p>	<p>Dr. Mohammad Saifur Rahman</p>
-----------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------	-----------------------------------

models for different protein attribute prediction problems. For the distance matrix to qualify as a valid kernel, Mercer's condition must be satisfied. Even if our custom kernels do not satisfy Mercer's condition, a given training dataset can possibly result in a positive semidefinite Hessian, in which case the training will converge perfectly well, despite lacking the theoretical basis for the maximum margin classifier. Therefore, we propose experimentally evaluating the performance of these custom kernels in several representative protein attribute prediction problems.

Thirdly, the absent word analysis can be useful in overlap detection in long reads of DNA sequences. Identifying overlaps between error-prone long reads is essential for error correction and de novo assembly. We will review the current state-of-the-art read-to-read overlap tools for error-prone long reads, including BLASR, DALIGNER, MHAP, GraphMap and Minimap. Several of these tools are built around the core idea of short k-mer matches for overlap detection. We, on the other hand, will explore how MAW and RAW can be used in this regard.

<p>Parkinson's Analysis with Remote Tasks</p>	<p>Park (Parkinson's Analysis with Remote Kinetic-tasks) is an online (<a href="https://parktest.net/index.html">https://parktest.net/index.html</a>) system built by University of Rochester that enables the measurement of Parkinson's disease for anyone, anywhere - via webcam. PARK instructs and guides users through six motor tasks and one audio task, and records their performance via webcam. Dr. Ehsan's team has published a sequence of papers using machine learning and deep learning, analyzing the dynamics of facial expressions and the motion of hand gestures to automatically predict PD vs. non-PD. However, their model lacks data from the Indian subcontinent.</p> <p>Tithi and Noshin et al. (CSE'16) worked on this project last year. They first created a Bangla language version of the framework (<a href="https://parktestbangla.web.app/">https://parktestbangla.web.app/</a>). Then they worked with collaborating Neurologists to collect data from patients. We now have a Bangladeshi dataset containing data from approximately 170 subjects, around 32 of which are Parkinson's patients. While they have conducted some analysis on the dataset, more remains to be done. In this project, we will analyze the Bangladeshi dataset to create models based on different motor-neuron tasks to distinguish between PD vs. Non-PD.</p> <p>Reference Material</p> <ol style="list-style-type: none"> <li>1. <a href="https://parktestbangla.web.app/">https://parktestbangla.web.app/</a></li> <li>2. <a href="https://www.facebook.com/toKnowParkinsons">https://www.facebook.com/toKnowParkinsons</a></li> <li>3. <a href="https://roc-hci.com/current-projects/park-parkinsons-analysis-with-remote-kinetic-tasks">https://roc-hci.com/current-projects/park-parkinsons-analysis-with-remote-kinetic-tasks</a> / [See the reference papers]</li> </ol>	<p>1-2</p>	<p>Dr. Mohammad Saifur Rahman</p>
-----------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------	-----------------------------------



<p>Single-cell and Spatial Transcriptomics (ST) data analysis</p>	<p><i>Research in these topics will be conducted in collaboration with Dr. Md. Abul Hassan Samee, Assistant Professor, Molecular Physiology and Biophys, Baylor College of Medicine</i></p> <p>The recent boom in microfluidics and combinatorial indexing strategies, combined with low sequencing costs, has empowered single-cell sequencing technology. Thousands—or even millions—of cells analyzed in a single experiment amount to a data revolution in single-cell biology and pose unique data science problems. Single-cell RNA sequencing (scRNA-seq) enables transcriptome-wide gene expression measurement at single-cell resolution, allowing for cell type clusters to be distinguished, the arrangement of populations of cells according to novel hierarchies, and the identification of cells transitioning between states. In the paper linked below, the authors have outlined eleven challenges that will be central to bringing this emerging field of single-cell data science forward. For each challenge, the authors highlight motivating research questions, review prior work, and formulate open problems. We can work on one or more of these areas in collaboration with Dr. Samee.</p> <p>Spatial transcriptomics (ST), on the other hand, captures gene expression from intact tissues at single-cell resolution. ST data reveal principles of tissue architecture, such as cellular colocalization patterns, and variations in gene expression across the tissue. Studying ST data from diseased samples could identify and characterize tissue regions for targeted therapy design. Some questions we could ask: Can we identify modules of genes that show systematic covariation of expression across the tissue? Can we segment an ST data into regions of high gene expression similarity?</p> <p>Nuwaisir and Turza et al. (CSE'16) have worked on proposing novel spatial clustering methods to identify the different segments (healthy vs. diseased) in a tissue sample, given the ST data frame. Their work outperformed state-of-the-art methods such as BayesSpace</p>	<p>2-3</p>	<p>Dr. Mohammad Saifur Rahman</p>
-------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------	-----------------------------------

and SC-MEB. Further work could be done to improve on their work and/or introduce new research questions.

Tousif, Enan and Romiz et al. (CSE'16) have worked on developing methods for cell localization in a tissue sample. However, their method does not work for multiple cell co-localization. This is something we would like to investigate further.

#### Reference material

1. <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-1926-6>
2. <https://www.youtube.com/watch?v=hWWkpe4Ewgo>
3. [https://satijalab.org/seurat/archive/v3.2/spatial\\_vignette.html](https://satijalab.org/seurat/archive/v3.2/spatial_vignette.html)

<p>Efficient algorithms for reconstructing phylogenetic trees (evolutionary trees)</p>	<p><b>(Two projects on phylogenetic tree estimation will be assigned)</b></p> <p>The overarching goal of these two research projects is answering impactful biological questions, especially those related to the study of evolution, by developing algorithms that can accurately analyze very large genome-scale datasets. The ongoing big data revolution in genomics can vastly increase our understanding of biology only if our computational toolkit can keep up with the pace of ever increasing abundance of molecular data.</p> <p>In these projects, we will be developing efficient algorithms for inferring phylogenetic trees (evolutionary trees) from genome-scale data. Phylogenetic trees provide insights into basic biology, including how life evolved, the mechanisms of evolution and how it modifies function and structure, disease evolution, criminal investigation etc. A species tree represents the evolutionary history of a group of organisms, while a gene tree shows the evolutionary pathways of a particular gene within a group of organisms. Interestingly, different genes evolve in different ways, meaning that they do not necessarily have identical evolutionary histories. This is known as gene tree discordance.</p> <p><b>Project 1:</b> In this particular project, our research will contribute to the problem of fast and accurate species tree estimation from genes sampled throughout the whole genome, considering the presence of gene tree discordance.</p> <p><b>Project 2:</b> Developing novel methods to address the problem of gene tree estimation error while estimating species trees.</p>	<p>1-4</p>	<p>Dr. Md. Shamsuzzoha Bayzid</p>
----------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------	-----------------------------------

	<p><b>Prior knowledge of Biology is not required.</b></p>		
<p>Application of machine learning in predicting protein attributes</p>	<p>Proteins are considered to be the building blocks of life. To understand the molecular foundation of life, it is critical to study individual proteins, and their biological attributes such as protein structure, folding, protein-protein-interaction, etc. With the success of the human genome project and advancements in sequencing technologies, there has been a rapid growth in the number of sequence-known proteins. However, advancement is much slower in determining their biological attributes. To mitigate this gap between sequence-known proteins and attribute-known proteins, we have to face the challenge of developing fast and highly accurate methods to predict protein attributes by analyzing protein sequences. In this project, we aim to apply machine learning techniques to predict protein attributes (in particular, protein structures, protein foldings, protein-protein interaction) from protein sequences.</p> <p><b>Please see the following articles:</b></p> <ol style="list-style-type: none"> <li>1. <a href="https://doi.org/10.1093/bioinformatics/btaa531">https://doi.org/10.1093/bioinformatics/btaa531</a></li> <li>2. <a href="https://doi.org/10.1093/bib/bbab578">https://doi.org/10.1093/bib/bbab578</a></li> </ol> <p><b>Prior knowledge of Biology is not required.</b></p>	<p>1-2</p>	<p>Dr. Md. Shamsuzzoha Bayzid</p>

<p>Racially Diverse Cellular Models of Alzheimer's Disease for Better Diagnosis and Cure</p>	<p>The increase in the number of patients with Alzheimer's disease (AD) is a serious problem to achieve healthy longevity. The current system for diagnosis and management of AD is far from sufficient, partly due to the lack of racial/ethnic diversity in research data, which has been mainly focused only on Caucasians. Racial/ethnic differences in biological risk factors may help to explain disparities in the incidence and prevalence of AD. Apolipoprotein E (APOE) polymorphism has been identified as the major genetic determinant of sporadic AD (sAD), representing over 95% of AD cases. There are three major isoforms: APOE ε2, APOE ε3, and APOE ε4. While the ε4 allele of the APOE gene is the strongest genetic risk factor for sAD, the biological mechanisms by which the APOE ε4 increases the risk of developing sAD in different racial groups are still unknown. Induced pluripotent stem cells (iPSCs) hold great promise to model AD as such cells can be differentiated in vitro to the required cell type. Using human induced iPSC driven cells, our study aims to investigate the differential effects of APOE on brain cell types in two distinct ethnic/racial groups: Japanese and Caucasians. <b><i>We will leverage CRISPR-Cas9 genome editing in this regard.</i></b></p> <p><b>Hypothesis</b></p> <p>Previous studies tested the associations of APOE ε alleles with sAD risk among Caucasians, African-Americans, Hispanics, and East Asians. Because these analyses revealed substantial heterogeneity by race/ethnicity, we hypothesize that there will be race/ethnic-specific effects of APOE ε4/ε4 genotype on gene expression and function in brain cell types. Specifically, a comparison of Japanese and Caucasian cell lines may indicate different APOE-mediated pathways in sAD pathogenesis. Thus, our proposed study aims to investigate the impact of genetic differences on AD risk by utilizing genome-edited iPSC-based modeling of AD from different ethnic/racial groups.</p>	<p>1-2</p>	<p>Dr. Md. Shamsuzzoha Bayzid</p>
----------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------	-----------------------------------

	<b>This project will be done in collaboration with:</b>		
--	---------------------------------------------------------	--	--

i) Icahn School of Medicine at Mount Sinai, New York, USA

ii) Tokyo Metropolitan Institute of Medical Science, Japan.

<p>Machine Learning Model for Prediction of Cognitive Decline in Older Adults</p>	<p>Progression to dementia due to Alzheimer’s Disease (AD) involves multiple pathways that affect cognition. Cognitive impairment refers to the loss of thinking abilities, such as memory, attention, executive function, and language, which negatively affects individuals’ daily life. In the stage of mild cognitive impairment (MCI), affected individuals experience cognitive loss without significant decline in their day-to-day activities. In the stage of severe cognitive impairment, which is referred to as dementia, individuals begin to lose their abilities to execute daily activities. Individuals who develop dementia follow a trajectory from a stage of normal cognition to MCI and dementia. Predicting the early onset of dementia due to AD has major implications for timely clinical management and outcomes. However, diagnosis at early stages of AD is impeded by heterogeneity in patient populations and comorbidities that may lead to a diagnosis of MCI without progression to AD. Early prediction of AD conversion has major implications for treatment and prognosis.</p> <p>Recent advances in machine learning (ML) allow us to develop predictive models of AD by mining multimodal datasets that include both genetic and non-genetic data from large patient cohorts. There have been many previous studies on identifying genetic risk factors of AD and identified numerous significant genetic loci that are associated with the disease. With respect to the investigation on non-genetic factors, education, age, gender, blood pressure, cholesterol, and body mass index, were identified as predictors of AD. To the best of our knowledge, there are few studies that analyze the combined effects from both genetic and non-genetic factors on AD. Polygenic risk scores (PRS) are used to aggregate the effects of thousands of trait-associated DNA variants discovered in genome-wide association studies (GWAS). PRS have been used to estimate individual-specific genetic susceptibility and predict disease. However, prior studies focused on uni-variate PRS models and used samples of European ancestry with limited diversity. Here, we aim to leverage state-of-the-art ML techniques to develop a multi-PRS model that captures both genetic and non-genetic risk</p>	<p>1-2</p>	<p>Dr. Md. Shamsuzzoha Bayzid</p>
-----------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------	-----------------------------------

	<p>factors to predict AD in older adults across ancestral groups, and thereby understanding the heterogeneity of AD.</p> <p><b>This project will be done in collaboration with:</b></p> <p>i) Icahn School of Medicine at Mount Sinai, New York, USA</p> <p>ii) Tokyo Metropolitan Institute of Medical Science, Japan.</p>		
<p>Algorithms and Computation (Exact topics will be fixed by discussing with the students. Students can work individually or in groups.)</p>	<p>Research Area:</p> <p>Algorithms (exact, approximation, online, randomized, distributed, heuristics and metaheuristics, ML &amp; GNN), Graph Drawing, Graph Algorithms, Network Science, Bioinformatics, Cyber Security, Graph Neural Networks, Data Clustering, Big Data Analytics and Visualization</p> <p>You may call me if you are interested.</p> <p>A brief description of our RISE research project:</p> <p>Development of Efficient Graph Algorithms for Big Data Analytics and Visualization</p>	<p>2-3</p>	<p>Dr. Sadia Sharmin</p>



Along with the penetration of ICT in society and the advance and spread of sensors, measurement instruments and observation equipment for gathering information in the real world, the amount of data obtained from various fields has grown exponentially and continues to become more diverse and more frequent in occurrence. This data is termed as “big data.” How to handle big data is a very important issue in computer science. In the theoretical area, developing efficient algorithms for handling big data is an urgent task. In the areas like web graph and social networks analysis, large scale protein interaction analysis, evacuation planning etc., where algorithmic techniques are needed to apply on big data, traditional efficient polynomial time algorithms are not useful. For example, taking input of a graph with 1 peta vertices and 2 peta edges takes 3.5 days with a 10 BIPS computer, and finding a maximum matching will take .2 million years. Thus design of efficient fundamental algorithms for property testing, matching, visualizing, substructure enumeration etc. in the context of big data is necessary for dealing with future information technology.

Most real-world data sets are relational, which can be modeled as graphs, consisting of vertices and edges. Algorithms for automated drawings of graphs are the fundamental algorithms behind all sorts of visualization of data sets. In the last few decades graph drawing algorithms were primarily used for VLSI design automation where the underlying graphs were planar graphs, and hence structural properties and fundamental algorithms for planar graphs have been studied extensively [1,2]. However, most of the real-world graphs are non-planar and dynamic. In particular, many scale-free networks, which can be used to model web-graphs, social networks and biological networks consist of sparse non-planar graphs. To analyze and visualize such real-world networks we need to solve fundamental mathematical and algorithmic research questions on non-planar graphs and dynamic graphs. In recent years researchers have concentrated their attention in this area and several

week-long workshops and symposiums have been organized to define and formulate these research problems [3,4,5].

Generating networks with a desired degree distribution is an important topic of research in network science. Often network researchers need to perform empirical experiments on various types of networks and gain valuable observations. Synthetic network generators can help a lot here by reducing loads on real world data. Among various degree distributions, the power law degree distribution found in scale-free networks is the most common. Albert and Laszlo Barabási have mentioned a number of models for generation of scale-free networks [6]. However, these models don't impose any restriction on generated networks to satisfy the property of planarity, connectivity etc. Thus there is a need for developing algorithms for generating scale-free networks satisfying desired graph theoretic properties.

In analyzing big data, it is often required to clarify small and middle size structures in big data so that complicated big data will become small understandable data [7]. Data analysis such as machine learning becomes easy to design, visualization becomes clear, and the solutions will have connections to human understanding. Particlization discloses latent structures in customer data, economy data, and social networks, and supports innovations in new businesses on information technologies through useful data analysis. To deal with relational data like graphs, graph neural networks may give better performance over other machine learning approaches.

Dealing with big data is a real *challenge* in the upcoming 4th industrial revolution. Many existing algorithmic techniques will not perform well in handling big data. Finding new algorithmic paradigms to cope with these challenges is a great need. This research project aims to solve fundamental mathematical and algorithmic research problems that will evolve

while dealing with big data. These problems include characterizing the classes of non-planar graphs, development of efficient algorithms for automatic drawing of large non-planar graphs, for generating scale-free networks etc. Developing and improving the techniques for analyzing useful latent structure inside big data using machine learning approaches such as graph neural networks is also an important focus of this research project.

#### References

1. T. Nishizeki and M. S. Rahman, Planar Graph Drawing, World Scientific, 2004.
2. M. S. Rahman and M. R. Karim, Drawing planar graphs, Proc. of WALCOM 2020, Lecture Notes in Computer Science, 12049, pp. 3-14, 2020.
3. S. Hong, M. Kaufmann, S. G. Kobourov, and J. Pach, Beyond-Planar Graphs: Algorithmics and Combinatorics, Report from Dagstuhl Seminar 16452 (2016), Dagstuhl Reports, Vol. 6, Issue 11, pp. 35–62, 2017  
([https://drops.dagstuhl.de/opus/volltexte/2017/7038/pdf/dagrep\\_v006\\_i011\\_p035\\_s16452.pdf](https://drops.dagstuhl.de/opus/volltexte/2017/7038/pdf/dagrep_v006_i011_p035_s16452.pdf))
4. S. Hong and T. Tokuyama (Eds.), Beyond Planar Graphs, Communications of NII Shonan Meetings, Springer, 2020.
5. A. Meidiana , S. Hong and Peter Eades. New Quality Metrics for Dynamic Graph Drawing. Proc. of Graph Drawing 2020, LNCS, 12590, pp. 450-465, 2020.
6. A. Barabási, Network Science, Cambridge University Press, 2016.
7. P. Mutzel, Algorithmic Data Science, 36th International Symposium on Theoretical Aspects of Computer Science (STACS 2019), LIPICS, 2019.

<p>Understanding Users, Cognitive and Affective Factors</p>	<p>Cognition and affection are important for interaction design. Cognition and affection has been applied to interaction design and cognitive research has led to the design of more effective interactive products, mental models, conceptual frameworks that are useful for interaction design.</p> <p>Research area: HCI, AI.</p> <p>Actual topic will be decided based on discussion.</p>	<p>2-3</p>	<p>Dr. Sadia Sharmin</p>
<p>Attention Engineering for local applications</p>	<p>Facebook engineers your attention. Facebook algorithms choose and schedule content and ads optimally. Based on your media use, cookies, browser history, and myriad other factors, they present you with the most engaging content at the best possible time. The addictive cycles that Facebook, YouTube, Snapchat, Reddit, Netflix, and others proliferate alter your perceptions and change your mental effectiveness. They cause people to easily consume hyperbolic journalism, biased information, misinformation, and disinformation. Individuals receive content that confirms their beliefs and ignores or chastises conflicting beliefs. This confirmation bias can lead to lapses in judgement and vapid, linear thought. Attention engineering strategies, ironically, contribute to scattered attention, less productive work, less critical thinking, and less self-control. We would like to investigate what is the case with new/our local applications.</p>	<p>1-2</p>	<p>Dr. Sadia Sharmin</p>

<p>Deep learning for misinformation detection( Fake news detection in social media/ Citizen journalism)</p>	<p>Recently, the use of social networks such as Facebook, Twitter, and Sina Weibo has become an inseparable part of our daily lives. It is considered as a convenient platform for users to share personal messages, pictures, and videos. However, while people enjoy social networks, many deceptive activities such as fake news or rumors can mislead users into believing misinformation. Besides, spreading the massive amount of misinformation in social networks has become a global risk. Therefore, misinformation detection (MID) in social networks has gained a great deal of attention and is considered an emerging area of research interest. Several studies related to MID have been studied to new research problems and techniques. While important, however, the automated detection of misinformation is difficult to accomplish as it requires the advanced model to understand how related or unrelated the reported information is when compared to real information. The existing studies have mainly focused on three broad categories of misinformation: false information, fake news, and rumor detection. Therefore, related to the previous issues, we present a comprehensive survey of automated misinformation detection on (i) false information, (ii) rumors, (iii) spam, (iv) fake news, and (v) disinformation.</p> <p>We want to investigate using deep learning approach.</p>	<p>2-3</p>	<p>Dr. Sadia Sharmin</p>
<p>Study and design of methods to analyze scRNA-seq and spatial transcriptomics data</p>	<p>All cells in an organism share almost the same DNA sequence and the same set of genes. However, different cells look and behave differently as the sets of genes expressed vary across cell types. Single cell RNA-seq (scRNA-seq) data is used to study this. Spatial transcriptomics assays can provide the spatial context of the cells. Here, we will study various challenges in this area and develop methods to address those challenges. Possible topics include Imputation of missing data using RNA velocity RNA velocity estimation in isoform resolution RNA abundance estimation in spatially resolved scRNA-seq data.</p>	<p>2</p>	<p>Dr. Atif Hasan Rahman</p>

	<p>This will be a collaboration with Md. Abul Hassan Samee, Ph.D., Assistant Professor, Baylor College of Medicine No prior knowledge of biology is necessary but will be helpful</p>		
<p>Reference-free methods in computational biology</p>	<p>The reference genomes of most organisms including agricultural crops remain incomplete. This makes methods for various tasks in computational biology that rely on the reference inaccurate. Here, we will study and develop reference free methods for tasks such as association mapping. Association mapping is the process of finding which variants in the genome are associated with diseases and traits. We will develop a method for finding relations between traits and variations in genomes in a reference free manner using k-mers. No prior knowledge of biology is necessary but will be helpful.</p>	2	<p>Dr. Atif Hasan Rahman</p>
<p>Structural variant calling in genomes using deep learning</p>	<p>While the human reference genome can now be termed complete through the efforts of the telomere-to-telomere (T2T) consortium, genomes of different individuals vary from each other. These variations are small sequence variations as well as structural variations such as reversals, translocations, deletions, duplications, etc. Many of them are related diseases. In this thesis, we will develop a deep learning based method to detect locations of structural variations (breakpoints). No prior knowledge of biology is necessary but will be helpful.</p>	2	<p>Dr. Atif Hasan Rahman</p>

<p>Addressing challenges of designing efficient fault tolerant and elastic VNF state management system under various consistency requirements</p>	<p>Network Functions can be stateless or stateful. Network function virtualization (NFV) has enabled network functions to be implemented on virtual machines. One of the main goals of designing virtual network functions (VNF) is to achieve elasticity, i.e., the number of virtual instances can grow and shrink based on traffic load.</p> <p>There are some challenges of designing an elastic state management system. One of the core challenges in achieving elasticity is to determine how NF states are stored and shared across multiple NF instances. State management can be even more complicated when NF instances can fail. Other challenges include limited buffer size, service level agreements, consistency requirements. In this project, our goal is to address the aforementioned challenges. Moreover, we aim to find the consistency requirements for different NF types and find tradeoffs between performance and consistency for different NF types.</p>	<p>2</p>	<p>Dr. Rezwana Reaz</p>
<p>Leveraging fog-cloud computing for implementing IoT firewalls</p>	<p>IoT devices are vulnerable to many attacks. So it is important to implement IoT firewalls to filter out unwanted traffic before they can reach IoT devices. One way to implement an IoT firewall is to store and run a firewall in each IoT device customized for that device. However, most IoT devices come with very limited memory and computational capability. So storing and running a firewall within an IoT device is not a practical solution. Other options include implementing the firewall in the network hub, in a gateway router, or in a distant node like fog/cloud.</p> <p>In this project, we will explore the fog-cloud based architecture to implement firewalls for IoT devices in a smart-healthcare scenario. Such an approach involves traffic redirection to a cloud or fog nodes and execution of the firewall in the cloud/fog nodes. While such an approach is beneficial in terms of scalability, economy, memory utilization, and computational</p>	<p>2~3</p>	<p>Dr. Rezwana Reaz</p>

	capacity, it poses several challenges in terms of latency, stateful inspection of firewalls, and privacy. In this project, we will address these challenges.		
Study of application of Blockchain in network function chaining	Software Defined Networking (SDN) can be used with NFV to steer traffic through a sequence of virtual network functions creating a Service Function Chain (SFC). SFC imposes security challenges when implemented in a multi-tenant and multi-domain environment. A recent trend is to use blockchain technology in securing service function chains that span over multiple domains. In this project, we will investigate the challenges involved regarding availability and scalability issues in such designs and outline possible solutions.	1	Dr. Rezwana Reaz