Thesis Topics

SL	Topic	Description	Students	Teacher
1	Research on	We will start with the tutorial, Knowledge-Augmented Methods for Natural Language Processing, available at this location. Since this tutorial opens doors to many research topics, after the study of about four weeks, the students will branch out to the narrow topics in groups of 2. The students will be expected to work in the fields (as and when applicable) of machine learning, natural language processing, and knowledge graph.	3 groups of 2 each, total 6 students	Dr. Muhammad Masroor Ali (mmasroorali)
2	Algorithms and Compuation (Exact topics will be fixed by discussing with the students. Students can work individually or in groups.)	Research Area: Algorithms (exact, approximation, online, randomized, distributed, heiristics and metaheuristics, ML & GNN), Graph Drawing, Graph Algorithms, Network Science, Bioinformatics, Cyber Security, Graph Neural Networks, Data Clustering, Big Data Alalytics and Visualization For detalis pls check https://cse.buet.ac.bd/research/group/gd/ http://saidurrahman.buet.ac.bd/ https://drive.google.com/file/d/1L9ZLJiRRfCdeWBztb38g1jGbZsPIvT7r/view You may call me if you are interested. (https://cse.buet.ac.bd/research/group/gd/) (http://saidurrahman.buet.ac.bd/) (https://drive.google.com/file/d/1L9ZLJiRRfCdeWBztb38g1jGbZsPIvT7r/view) Some current students are working on the following topics: Minimum Bend Orthogonal Convex Drawing of Planar Graphs	8	Dr. Md. Saidur Rahman (saidurrahman)

SL	Topic	Description	Students	Teacher
		On a new subclass and an application of pairwise compatibility graphs		
		Randomization in Double Coverage Algorithm on a Line for Online K-Server Problem		
		Impacts of Monetary Policy and Credit Ratings on Call Money Market		
		Twin-width of Planar Graphs		
		Practical Overlapping Community Detection in Weighted Graphs using Deep GCNs		
		Generating scalefree networks with desired graph theoretic properties		
		Generating Spanning Trees with Constraints		
		Approximation algorithms for graph drawing		
		A brief description of my RISE research project: Development of Efficient Graph Algorithms for Big Data Analytics and Visualization 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		

SL	Topic	Description	Students	Teacher
		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].		

SL	Topic	Description	Students	Teacher
		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.		
3	Automatic Design of Convolutional Neural Network	Convolutional neural networks (CNNs) have exhibited remarkable success on various image classification tasks. However, the performance of CNNs highly relies upon their architectures, which are often manually designed with expertise in both CNNs and the investigated problems. Therefore, it is difficult for users, who have no extended expertise in CNNs and the problems. The aim of this work is to devise automatic approaches for designing CNNs for various image processing tasks.	2-3	Dr. Md. Monirul Islam (mdmonirulislam)
4	Handwritten Bangla Character Recognition	Necessity of recognizing handwritten Bangla characters is increasing day by day because of its various applications. Although there have been many attempts in recognizing handwritten English characters, Chinese characters, few attempts have been made for handwritten Bangla characters. The objective of this research is to develop deep learning based models for handwritten Bangla characters.	2-3	Dr. Md. Monirul Islam (mdmonirulislam)
5	Design and Implementation of Display Controller	Generally rasberry pi, ATmega etc are used for display controller. Such system consumes power and slower as those are general purpose system. Our target is to implement an integrated circuit so that it works faster with less power consumption and space. We will use verilog code to design the circuit	1	Dr. Md. Mostofa Akbar (mostofa)

SL	Topic	Description	Students	Teacher
	using FPGA based ASIC	VHDL. The target design will be deployed in an FPGA based system. This will never be as efficient as VLSI circuit but it will be definitely better than off the shelf aurdino based system.		
6	Teaching and Learning Patterns for Pointers in C programming language	The main target is to investigate different way of introducing concept of pointers to tbe students and find the best possible way of learning pointers. This requires survey among the students to find the effectiveness of the methodology.	1	Dr. Md. Mostofa Akbar (mostofa)
7	Review of MMKP heuristics	There are several MMKP heuristics with different complexities. The solution quality and time reqiurement for solving MMKP by these heuristics are not the same. But there is no comprehensive comparison among these heuristics as those were tested in different environment. We would like to prepare an open source object with sample data sets as well as comparative results. The main target is to write review paper on this topic.	1	Dr. Md. Mostofa Akbar (mostofa)
8	Exploring Techniques of Automated User Interface Testing.	The reseacher will explore the following: Testability requirement of the application will be defined. Generation of commands from the test cases. The technique of qualtifying visual outputs to determine whether the test paases on not.	1-2	Dr. Md. Mostofa Akbar (mostofa)
9	Algorithms and Theory	We plan to work on some theoretical problems mostly focusing in graphs and/or stringology. The main goal would be to devise and analyse algorithms, prove necessary and/or sufficient conditions etc. Ocasionally this can be supplemented with experimental validation. Interest and confidence in Discrete Math, DS/Algorithms course would be preferable.	1-3	Dr. M. Sohel Rahman (msrahman)

SL	Topic	Description	Students	Teacher
10	Quantum Computing	We want to study quantum computing under this project and advance the current state of the art. This will be a learning experinece for both the supervisor and supervisee.	1-2	Dr. M. Sohel Rahman (msrahman)
11	Computational Biology and Bioinformatics	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. Also, we may need to do optimization (metaheuristic based or otherwise).	Any	Dr. M. Sohel Rahman (msrahman)
12	Computational Modeling, ABM	We want to leverage the power of computational modeling, in particular, Agent Based Moedling to model various aspects in nature, society etc. The scope includes but is not limited to, disease modeling (e.g., COVID 19), traffic modeling, social/ecological modeling etc.	1-3	Dr. M. Sohel Rahman (msrahman)
13	Various Aspects related to Machine Learning	We would love to study aspects like Introduced Bias, adversarial robustness and model fairness and interaction between and among these.	1-2	Dr. M. Sohel Rahman (msrahman)
14	Let's chat about chatGPT	Recently generative AI based predictive models like chatGPT, Google BARD etc have drawn lots of attention of the research community. The world has seen some astonishing capability of these generative tools in answering many questions, summarizing long texts, designing and drawing with human like skills, devising programming codes to complex problems etc. However, there is a mixed reaction. Some researchers are very much enthusiastic about its capabilities and some are very much negative about it. For example, some research papers are highlighting a plethora of its failure in different context and some are arguing that chatGPT is a "jack of all trades, master of none..." In	3-4	Dr. A.K.M. Ashikur Rahman (ashikurrahman)

SL	Topic	Description	Students	Teacher
		this research problem we will explore and evaluate chatGPT's capability in a wider context, for example in solving mathematical problems, in answering questions in bangla or solving GRE like logical reasoning problems. **DISCLAIMER**: This research problem has been generated using chatGPT ... ;-)		
15	Angry faces are better than neutral faces for face recognition	Face recognition is an important research problem with many applications. Accurate face recognition is always a challenging problem. Facial expressions may improve face recognition accuracy as many persons have some unique way to express themselves. For example, the unique smile of Shahrukh Khan (SRK) has touched many hearts and has the ability to move even the mountains. Researchers have successfully shown that using smile rather than neutral images might bring significant improvements in face matching accuracy. But how about other facial expressions such as anger or disgust? In this research, at first we shall review the state-of-the-art techniques to identify anger/disgust in facial images. Then we will be building a machine learning based models to find the impact of anger/disgust on face recognition accuracy in depth and compare the results with smile's impact on the accuracy.	3	Dr. A.K.M. Ashikur Rahman (ashikurrahman)
16	IoT based automotive battery management system for monitoring car battery health	The purpose of an automotive battery is to provide power for starting a car. It also provides power for short-term use of car lights, car stereo, or wipers when the engine is off. Over the lifetime of a battery, discharge-recharge reactions happen thousands of times. Each cycle wears out the plates a bit, and over time the lead deteriorates. Deep discharging, which happens when the battery is used to run the stereo, lights or other electrical systems in the car when the engine is off, is responsible for a good portion of battery failures. In order to ensure reliable and efficient use of automotive batteries, it is necessary to monitor the state of the battery condition continuously in real time. Therefore an IoT-based battery monitoring system can be used to track the health of the battery. In this research, we will design and implement an arduino based IoT system equipped with SIM cards that would be mounted on top of a car battery and will be using the cellular network to send SMS to	3	Dr. A.K.M. Ashikur Rahman (ashikurrahman)

SL	Topic	Description	Students	Teacher
		the owner of the car when the health condition of the battery deteriorates. This is a collaborative project with Rahim Afroze.		
17	On DoS and DDoS attacks in MAC protocols in full duplex channel of wireless multi hop networks	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 MAC layer in full duplex channel and explore the possibilities and counter measures of DoS and DDoS attacks in such networks. Some knowledge of ns-3 simulator will be an asset as this project will be a continuation project from a previous group who worked and published a paper.	3	Dr. A.K.M. Ashikur Rahman (ashikurrahman)
18	Efficient and Effective Charging of Electric Vehicles (EVs)	**Subject areas: Machine Learning, Algorithms (In collaborations with researchers from University of Melbourne)** **The global electric vehicle (EV) market is expected to reach $823.75 billion by 2030. The demand of EVs has skyrocketed in recent years due to their fuel-efficient, high-performance, and low-emission nature. One of the fundamental problems EVs are facing these days is efficient and effective charging.Given a source and a destination of an electric vehicle, a set of charging stations' location, the problem is to find a sequence of charging stations where the EV can charge if charging is needed to reach the destination. The objective can be finding the sequence for charging that provides the lowest time to reach to the destination (as electric charging takes some time even with the superchargers), and/or provides the lowest cost for charging. There can be queuing at the charging stations (waiting time), and the charging time at different stations can be different based on the charging rate of that station. The search space becomes larger if partial**	1-2	Dr. Mohammed Eunus Ali (eunus)

SL	Topic	Description	Students	Teacher
		charging is allowed. In this work, we will explore an important problem of EV charging, which requires efficient algorithmic/learning solutions. Initial Reading - Meryem Abid, Mohammed Tabaa, Asmae Chakir, Hanaa Hachimi, Routing and charging of electric vehicles: Literature review, Energy Reports,Volume 8, Supplement 9, 2022, Pages 556-578, https://doi.org/10.1016/j.egyr.2022.07.089		
19	Learning to Index 3D Point-Cloud for Efficient Retrieval	Subject areas: Deep Learning, Data Structures (in collaboration with researchers from university of Melbourne) A 3D point cloud, a database of points, is a digital or virtual representation of physical objects in 3D space. Various applications like medical imaging, manufacturing, city planning, 3D gaming, virtual reality make extensive use of these 3D point data point clouds. In most of these applications on 3D point clouds, 3D object retrieval is a fundamental problem. Specifically, given a set P of point clouds (each representing a 3D object), a query point cloud Q, the goal to find a set of point cloud objects from P such that for each point cloud p in that set, Q is similar to a portion of p. For example, given a seat as the query, it can return a car, as the query matches with the car seats. Such problem is also important for robot navigation or autonomous cars, as from a partial view from the sensors (that is, a query object), it may need to find the objects that match with it. Such as, from a partial view of a bicycle that's captured in its sensor (the other parts are obstructed from view), an autonomous vehicle needs to match it with existing objects' dataset to infer that it's a bicycle to make navigation decisions. As such matching for decision making needs to be done efficiently on the fly, a learned index for 3D point clouds that can provide fast answers is needed. There are many learned indexes for general 1D data, and some papers on 2D data. However, there's no learned index specifically for	1-2	Dr. Mohammed Eunus Ali (eunus)

SL	Topic	Description	Students	Teacher
		point cloud objects where each object consists of many 3D points. For this problem, the objective is also different - to do similarity checks instead of exact matching for retrieval. **Initial Readings** 1. **M. A. Uy and G. H. Lee. 2018. PointNetVLAD: Deep Point Cloud Based Re-trieval for Large-Scale Place Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).** 2. **J. -Y. Chen, C. -H. Lin, P. -C. Hsu and C. -H. Chen, "Point Cloud Encoding for 3D Building Model Retrieval," in IEEE Transactions on Multimedia, vol. 16, no. 2, pp. 337-345, Feb. 2014, doi: 10.1109/TMM.2013.2286580.**		
20	How the toxicity and biases change in Chain-of-Thought Prompting in LLMs	**Subject areas: NLP, HCI (In collaborations with researchers from Bosch Research)** **Large language models (LLMs) have demonstrated unprecedented capabilities in language generation, re-shaped natural language processing (NLP), and become integral AI assistants in various services including healthcare, biology/nutrition science, education, media and news, customer service, law enforcement, security, and surveillance. Since the users of AI systems come from diverse backgrounds with critical information, toxicity and biases in them can lead to catastrophic outcomes. Therefore, in addition to the capabilities, understanding of such limitations of LLMs is also essential. Now, on how to prompt the LLMs, a new prompting paradigm Chain-of-Thought (CoT) that generates both the answer as well the step-by-step reasoning behind it, achieves stronger performance on many tasks and has been adopted as a de facto prompt. While previous works show that Large Language models are found to demonstrate biases towards particular ethnicity, religion, culture, or group of people, the study was often limited to straightforward prompting and it is unclear how the toxicity and biases evolve with CoT. In this**	1-2	Dr. Mohammed Eunus Ali (eunus)

SL	Topic	Description	Students	Teacher
		work, therefore, we aim to conduct a principled study and analysis of how the toxicity and biases change in language generation using CoT. **Initial readings:** 1. https://arxiv.org/pdf/2305.04388.pdf 2. https://arxiv.org/pdf/2304.05335.pdf 3. https://aclanthology.org/2021.acl-long.329.pdf 4. https://arxiv.org/pdf/2101.05783.pdf		
21	Reasoning with Graph, Chart, and Tabular data using LLMs	**Subject areas: NLP, Deep Learning (In collaborations with researchers from Bosch Research)** The increasing availability and complexity of data in various forms, such as graphs, charts, and tables, present unique challenges for effective data analysis and interpretation. Traditional methods for extracting interesting insights from these data types often require domain-specific expertise and manual effort. However, recent advancements in large language models, such as GPT-3 and its successors, have shown promising capabilities in natural language understanding and generation. This work aims to explore the application of large language models for reasoning with graph, chart, and tabular data. We will study the different ways to prompt the LLMs to leverage them for these specific data types, analyze their bottlenecks, and ways to improve reasoning capabilities. **Initial Readings:** 1. https://arxiv.org/abs/2203.10244 2. https://arxiv.org/abs/2305.10037	1-2	Dr. Mohammed Eunus Ali (eunus)

SL	Topic	Description	Students	Teacher
22	Reinforcement Learning (RL)-Based Scheduling Algorithm for Cost-Optimized and SLA-Aware Deployment of Microservices in Multi-Cloud Kubernetes Environment	**Subject areas: Systems and Software Engineering, Reinforcement Learning (In collaborations with researchers from University of Melbourne)** **Microservices-based applications, such as e-commerce (or similar) platforms, require efficient scheduling algorithms to optimize resource utilization, maintain Service Level Agreements (SLA)s, and minimize costs. This research project aims to develop an RL-based scheduling algorithm for a multi-cloud (a few servers/Virtual Machines will be locally deployed, while the rest of the servers/VMs will be deployed on the Cloud) Kubernetes environment to deploy and manage microservices-oriented applications. The proposed algorithm will focus on autoscaling a locally deployed cluster using cloud resources while ensuring low latency for critical services and improving overall throughput, all while minimizing costs (as we have to pay for using the cloud resources). More specifically, we will develop an RL-based scheduling algorithm that can make intelligent decisions regarding pod placement, autoscaling, and resource allocation in a multi-cloud Kubernetes environment.** **Initial Readings:** 1. **Survey Paper (ACM Computing Surveys): Kubernetes Scheduling: Taxonomy, Ongoing Issues, and Challenges. Link: https://dl.acm.org/doi/10.1145/3539606** 2. **Multi-cloud Paper (IEEE Transactions on Computers): SLA-Based Scheduling of Spark Jobs in Hybrid Cloud Computing Environments.** **Link: https://ieeexplore.ieee.org/abstract/document/9416782** 3. **RL Paper (IEEE Transactions on Parallel and Distributed Systems): Performance and Cost-Efficient Spark Job Scheduling Based on Deep Reinforcement Learning in Cloud Computing Environments. Link: https://ieeexplore.ieee.org/abstract/document/9599497**	1-2	Dr. Mohammed Eunus Ali (eunus)

SL	Topic	Description	Students	Teacher
23	Earthquake Early Warning Prediction from Sparse Seismic Wave Data	Earthquake early warning (EEW) systems have become increasingly vital to provide timely alerts, which can help to avoid huge loss of human lives and properties. A key task of an EEWS is to accurately predict the seismic intensity in different regions of interest in real-time. Over the last few years, different deep learning based techniques such as CNN, GNN, etc. have been employed to improve the accuracy of the prediction. However, these models often fail to predict the seismic intensity of the earth quakes, where initial seismic waves data are recorded in very few stations over a large geographic region. In this project, we will employ recent deep learning techniques such as a few-shot learning/meta learning approach to solve this problem. This project is a part of our initiatives with JIDPUS-BUET to develop an EEWS for Bangladesh. Initial Readings: https://academic.oup.com/gji/article/229/1/704/6447541 https://pubs.geoscienceworld.org/ssa/srl/article-abstract/93/3/1636/612500/DeepShake-Shaking-Intensity-Prediction-Using-Deep https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2022JB024595	1-2	Dr. Mohammed Eunus Ali (eunus)
24	Applied Machine Learning	In the domain of image processing. It is a part of collaborative project.	2-3	Dr. Mahmuda Naznin (mahmudanaznin)
25	Attack Model Study	In the domain of recommendation system.	2	Dr. Mahmuda Naznin (mahmudanaznin)

SL	Topic	Description	Students	Teacher
26	Data and behaviorial analysis	Improving learning using some accessible tools for people with special needs. It is a part of collaborative project.	2	Dr. Mahmuda Naznin (mahmudanaznin)
27	Security in edge computing	This is about security and privacy challenges in edge computing.	2	Dr. Mahmuda Naznin (mahmudanaznin)
28	Identification and behaviorial analysis with biometric signals	This project will be blending of Digital Signal Processing and ML/DL.	2	Dr. Mahmuda Naznin (mahmudanaznin)
29	Different Topics on Image processing and computer vision	Topic: 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	1 to 6	Dr. Md. Monirul Islam (mmislam)

SL	Topic	Description	Students	Teacher
		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		
		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		

SL	Topic	Description	Students	Teacher
		Brain/lunge tumor detection and classification from MRI/CT images 2-3 students **Mathematical model for detection of image tampering/forgery 2-3 students** **Plant/crop disease classification from leaf images 2-3 students** **Plant identification from leaf databases 2-3 students** **Image processing solution for authenticity detection of paper currencies 2-3 students**		
30	Natural Hazard Prediction	Area: Deep Learning, Machine Learning, AI Lightning, a natural hazard with potentially severe consequences, poses threats to human lives, outdoor activities, aviation safety, and critical infrastructure. On an average four people are killed weekly by lightning in Bangladesh with 363 deaths in 2021 alone, which is getting worse every year due to climate change. Besides, lightning is near the top of the list of all types of weather-related deaths. Hence, it needs no further saying why accurate prediction of lightning is a must to enable timely warnings, allow individuals to seek shelter, organizers to make informed decisions about outdoor events, and aviation authorities to adjust flight paths and schedules. Existing lightning prediction models generally apply numerical modeling with different parameterization schemes to simulate atmospheric conditions and dynamics using mathematical equations and computational techniques by ingesting meteorological data from various sources, such as weather stations, radar, and satellite observations. However, the change in space-time domain and the manual tuning of the model's parameters introduce biases negatively affecting these models' performance. In this project, we want to predict lightning by using ML techniques by tapping into the spatio-temporal aspect of it while utilizing weather related datasets. References:	2	Dr. Tanzima Hashem (tanzimahashem)

SL	Topic	Description	Students	Teacher
		[1]https://news.mongabay.com/2022/09/for-lightning-prone-communities-in-bangladesh-new-warning-system-may-not-be-enough/ [2] Yangli-ao Geng, Qingyong Li, Tianyang Lin, Lei Jiang, Liangtao Xu, Dong Zheng, Wen Yao, Weitao Lyu, and Yijun Zhang. 2019. LightNet: A Dual Spatiotemporal Encoder Network Model for Lightning Prediction. In SIGKDD.		
31	Federated Machine Learning on Crowdsourced Data	Area: Privacy, Machine Learning, Crowdsourcing There are two sources of crime data for predicting crime: (i) public crime data (reported to police), (ii) crowdsourced crime data. Crowdsourced data contains more and recent information since many crimes go unreported. Moreover, not all cities have regularly updated public crime data (e.g. Dhaka). Therefore, we want to utilize crowd data along with public crime data and external features (e.g. traffic flow) to train an efficient crime-prediction model. However, the crowdsourced data is sensitive; hence, many data providers may be unwilling to share the raw crime data with a centralized server for training an ML model. Therefore, this research project proposes to use federated learning to train our crime-prediction model on crowd data. In a federated learning setting, a model is trained on local devices (e.g., smartphones) with local data. Sensitive raw data are not transferred from the data provider's device; only model updates and aggregated information are sent to the centralized server in order to update a global model. This approach can be applied to other spatio-temporal application areas. References: 1. https://ieeexplore.ieee.org/document/9458643 2. https://www.computer.org/csdl/journal/tk/5555/01/10023962/1K9soExCZSo	2	Dr. Tanzima Hashem (tanzimahashem)

SL	Topic	Description	Students	Teacher
32	Fairness in Temporal Machine Learning Models	Area: AI, Machine Learning Ensuring fairness in machine learning models is of crucial importance because it helps mitigate bias and discrimination, builds trust in AI systems, and contributes to building a more equitable and inclusive society. There has been a significant amount of research to develop fair machine-learning models. However, temporal data, such as time series data, poses unique challenges for ensuring fairness in machine learning models due to evolving patterns and biases present in data over time. This research proposal aims to investigate novel strategies (i.e., Generative AI, Dynamic Time Warping) to integrate fairness considerations into temporal machine learning models. We will focus on scenarios where gender or demographic biases are evident, as these biases have significant implications in decision-making processes. More potential applications include fair credit scoring, personalized healthcare interventions, and equitable resource allocation. By developing fair and unbiased temporal machine learning models, this research will contribute to building more transparent and accountable AI systems that can reduce discriminatory outcomes not only for minority groups but also promote equal opportunities for all individuals with equal merits. References 1. https://dl.acm.org/doi/abs/10.5555/3495724.34957862. https://dl.acm.org/doi/10.1145/3457607	1-2	Dr. Tanzima Hashem (tanzimahashem)
33	Green Route Planner	Area: Location-Based Services, Machine Learning, Ubiquitous Computing Location-based services, especially the journey planners like Google or Bing Maps, have become an integral part of our life for moving on roads with convenience. Existing services mainly consider distance and traffic while planning the routes for the travellers. However, the shortest or the fastest	1-2	Dr. Tanzima Hashem (tanzimahashem)

SL	Topic	Description	Students	Teacher
		route is not always the best choice. Nowadays, air pollution and excessive noise on roads have increased a lot worldwide. According to World Health Organization (WHO), approximately seven million people worldwide die every year from health related issues like chronic obstructive pulmonary disease, lung cancer and acute respiratory infections resulted from air pollution [2]. Thus, while travelling on roads, people may want to travel on less polluted and noisy routes, which are known as green routes. In this research project, we aim to develop variant map based services that return a green route within a distance constraint for traveling from a source location to a fixed destination or a point of interest (e.g., a restaurant). Reference 1. https://www.helsinki.fi/en/researchgroups/digital-geography-lab/green-paths 2. https://www.thedailystar.net/environment/pollution/air-pollution/news/dhaka-again-ranks-worlds-most-polluted-city-2976026		
34	Identification of cybersecurity-related entities and events for CTI mining	Cyber threat intelligence is a branch of cybersecurity that deals with the collection, analysis, and dissemination of information about current and potential cyber attacks that pose a threat to an organization's assets. The goal of cyber threat intelligence is to provide organizations with actionable information that can help them identify and mitigate potential threats before they cause harm. The identification of cybersecurity-related entities and events is very crucial for CTI mining. In cybersecurity entity and event extraction, named entities in the unstructured text are located and classified into predefined cybersecurity categories, such as impacted organizations, locations, vulnerabilities, etc. On the other hand, events are classified into predefined cyber attack categories, such as phishing, Distributed Denial-of-Service (DDoS) attacks, etc. Students will have to use NLP and deep learning techniques for this task.	2	Dr. Md. Shohrab Hossain (mshohrabhossain)

SL	Topic	Description	Students	Teacher
35	Cybersecurity attack tactics, techniques, and procedures	Cyber Intelligence is the knowledge that allows a company to prevent or mitigate cyber-attacks by studying the threat data and provide information on adversaries. In this thesis, the goal is to determine how cyber threat actors and hackers prepare and execute cyber attacks by analyzing their Tactics, Techniques, and Procedures (TTPs). This is analogous to pathology study in healthcare, which aims to understand the causes and effects of disease or injury. In addition, profiles of hackers will be generated which will help trace the origin of cyber attacks. The establishment of a hacker profile aims to uncover the sources and resources of a threat actor, including cyber threat attribution and hacker assets.	2	Dr. Md. Shohrab Hossain (mshohrabhossain)
36	Vulnerability exploits, malware implementation and Threat hunting for Cyber Threat Intelligence	he goal of cyber threat intelligence is to provide organizations with actionable information that can help them identify and mitigate potential threats before they cause harm. It is becoming increasingly common and dangerous to be exposed to cybersecurity risks and malware threats. There are a wide range of vulnerabilities that can lead to data leaks, and threat agents can exploit them to compromise secure networks. Despite much attention paid to vulnerability and malware detection using code semantics, mining CTI sources beyond code is limited in terms of discovering practical information about vulnerability exploits and malware implementation. In this work , students will analyze documentation to discover vulnerabilities under a particular product or service, predict exploits, and find information about malware implementation for predicting software characteristics. Threat hunting is a proactive approach to identify previously unknown, or ongoing nonremediated threats, within an organization's network. During threat hunting, the suspicious activity patterns that may deemed to be resolved but isn't or have been missed are inspected. Student will try to identify vulnerabilities that might be exploited and malware implementation through CTI mining. Multiple	2	Dr. Md. Shohrab Hossain (mshohrabhossain)

SL	Topic	Description	Students	Teacher
		data sources will be consulted in order to obtain knowledge about vulnerability exploits and malware implementation.		
37	Smart Infrastructure: Leveraging LLM/NN-based AI and IoT for Digital Asset Management and Structural Health Monitoring	This study is expected to offer an exceptional opportunity to work with cutting-edge technologies and contribute to the advancement of smart infrastructure. Large Language Models (LLMs)/Transformer-Neural-Network powered AI systems, specifically tuned for IoT applications, have the potential to revolutionize the field of smart infrastructure. This thesis is expected to focus on various critical infrastructures including bridges, industrial complexes, etc. Real-world structural health monitoring data are expected to be provided by prestigious institutions such as the Department of Transportation (DOT) in the United States. Moreover, collaboration closely with a research team led by a professor from USA is expected. Additionally, an expert from the North American industrial sector is also expected to provide valuable insights to ensure that the research aligns with industry needs and standards. Expected team size for this thesis will be 1-2.	2	Dr. A. B. M. Alim Al Islam (razi)
38	Future Networking Technology: 6G and beyond 6G	Different networking technologies (2G, 3G, 4G, etc.) have attained considerable successes paving the path of yet another networking technology, i.e., 6G, in near future, or it is already in. However, achieving success of the 6G, or beyond 6G, in specialized cases is still under research. An example of such cases is autonomous vehicles, which demand a special treatment to 6G to get it rolling in its full potent. 6G security, 6G integration with SDN, etc., are some other related research areas. In this thesis, our goal would be to devise new protocol(s) or networking architecture to enable enhanced and secured performance of 6G or beyond 6G, perhaps in specific applications. Attaining this goal would require development of new protocol(s) or networking architecture and	1	Dr. A. B. M. Alim Al Islam (razi)

SL	Topic	Description	Students	Teacher
		experimentation through simulation and/or testbed implementations. Expected team size for this thesis will be 1-2.		
39	Computing Solutions to Serve the Under-served	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. 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.	2	Dr. A. B. M. Alim Al Islam (razi)
40	Intelligent Traffic Signaling for Dhaka City	It is commonly believed that inefficient traffic signaling is one of the most important contributors to the poor on-road experience in Dhaka city. As per development in the relevant research studies, intelligent traffic signaling could substantially ameliorate the on-road experience. However, in the case of Dhaka city, direct adoption of the solutions existing in the first-world countries is not a reality considering the fact that Dhaka presents substantially different traffic covering heterogeneous and non-lane based traffic. Therefore, new intelligent traffic signaling system for Dhaka city needs to be developed. In this thesis, our goal would be to come up with a new intelligent traffic signaling system that can ensure better on-road experience. Attaining this goal would require development of new systems	2	Dr. A. B. M. Alim Al Islam (razi)

SL	Topic	Description	Students	Teacher
		and their experimentations (partially developed in our research group). Expected team size for this thesis will be 2-3.		
41	Simulating Road Traffic of Dhaka for Effective Policy Making	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. 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.	1	Dr. A. B. M. Alim Al Islam (razi)
42	Quantum Intelligence	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. 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 algorithm(s) and their experimentations in real quantum computers. Expected team size for this thesis will be 1-2.	1	Dr. A. B. M. Alim Al Islam (razi)

SL	Topic	Description	Students	Teacher
43	Embedded Intelligence and Fault Tolerance	Intelligence and fault tolerance are now being in widespread usage, however, mostly at the cost of high resource utilization. There remain many applications (such as microcontroller-based applications) that cannot afford such high resource availability. Thus, enabling intelligence and fault tolerance sustaining resource constraints has recently emerged as a prominent research area. In this thesis, our goal would be to come up with effective mechanism(s) that can enable intelligence and 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.	1	Dr. A. B. M. Alim Al Islam (razi)
44	Assessing Impacts of Heat through Computing Technologies	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. 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, 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.	1	Dr. A. B. M. Alim Al Islam (razi)
45	Detecting Tax evasion from Income Tax returns cross-referencing with Bank Statement	Tabular document layout parsing and OCR techniques will be used to retrieve relevant data from bank statement and then will be cross matched with corresponding fields of Income Tax return. The returns may be electronically submitted or scanned printed/handwritten documents. We shall start with electronically submitted returns and then try to extend for printed and handwritten scanned returns.	1/2	Dr. Anindya Iqbal (anindyaiqbal)

SL	Topic	Description	Students	Teacher
46	Interactive tutoring with AI	Large Language Models (LLMs) like ChatGPT is transforming the almost every knowledge industry. Using this technology, for the first time, there is a possibility of creating an AI tutor who will provide personalized high quality education to school student in Bangladesh who do not have access to personal private tutors. Platforms such as Khan Academy (https://youtu.be/hJP5GqnTrNo), Chegg, are already integrating these types of tools into their learning platforms. There are a few interesting research directions on this problem. Two of them are below, • The current English LLM without any translation can be used for tutoring Bangladeshi students in Bangladeshi curriculum in Math, Physics, Chemistry, etc. The main challenge is enabling the AI to help student think, and guide them to correct resources to learn those skills instead of revealing the answers. • If the time period permits, using previous work in English to Bangla and Bangla to English Neural Machine Translation in Department of CSE, BUET, this project can integrate Large Language Models for creating personalized in Bangla as well. A PhD student from University of Rochester with 3 years of experience in this area will closely collaborate on this problem.	2/3	Dr. Anindya Iqbal (anindyaiqbal)
47	Can Contrastive Learning Improve Classification or Generation Performance for Software Engineering	Recently a research from our group found significant improvement in API Review Classification. However, such improvements were not observed on some other problems. This motivates us to explore other similar problems such as (i)Toxicity detection in Code review, (ii) Anti-social behavior (in Bangla) detection in Social Media, (iii) fake news detection, (iv)Transformation of sentence from Sadhu to Cholito, etc.	2 or 4 (multiple groups)	Dr. Anindya Iqbal (anindyaiqbal)

SL	Topic	Description	Students	Teacher
	or Bangla NLP Problems?			
48	Aligning Large Language Models with Human Values	The rapid advancement of large language models, such as OpenAI's GPT-3, has raised concerns about their potential to generate biased or harmful outputs. Ensuring that these models align with human values is crucial for their responsible deployment in various domains. This research aims to investigate and develop techniques for aligning large language models with human values. The BUET CSE NLP Group is actively engaged in pretraining billion-scale GPT and T5 models in Bangla, which will serve as primary testbeds for our proposed research. The project will explore directions such as instruction fine-tuning and reinforcement learning with human feedback to enhance the model's alignment with specific human values in the Bangla language context. Additionally, the project will focus on developing interpretability frameworks to shed light on the decision-making processes of these models, enabling better understanding and control over their output. If time permits, we will also explore ideas from game theory, cognitive science, and behavioral economics to design objective/reward functions aligned with human rationales. References: Language Models are Few-Shot Learners Scaling Instruction-Finetuned Language Models Training language models to follow instructions with human feedback LLaMA: Open and Efficient Foundation Language Models Alpaca: A Strong, Replicable Instruction-Following Model	1-2	Dr. Rifat Shahriyar (rifat)

SL	Topic	Description	Students	Teacher
49	Retrieval-Augmented Large Language Models	This research will focus on exploring the potential of retrieval-augmented large language models, aiming to enhance their performance in natural language understanding and generation tasks. Retrieval-based methods have shown promising results in improving the quality and relevance of generated responses in conversational AI systems. This project proposes to investigate techniques that combine the strengths of large language models with effective retrieval mechanisms to generate more contextually relevant and coherent responses. The research will involve designing and training retrieval models that can efficiently retrieve relevant information from large knowledge bases or corpora to support the language model's generation process. Furthermore, the project will explore methods for fine-tuning large language models using retrieval-based objectives, enabling it to leverage retrieved information for more accurate and informed responses. Potential applications will include (but not be limited to) open-domain and/or cross-lingual question answering. References: Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks Dense Passage Retrieval for Open-Domain Question Answering One Question Answering Model for Many Languages with Cross-lingual Dense Passage Retrieval UDAPDR: Unsupervised Domain Adaptation via LLM Prompting and Distillation of Rerankers REALM: Retrieval-Augmented Language Model Pre-Training	1-2	Dr. Rifat Shahriyar (rifat)
50	Vision-Enhanced Large Language Models	This research will aim to investigate the integration of vision in large language models to enhance their capabilities in visual understanding and generation tasks. While large language models have achieved remarkable success in natural language processing, they often lack the ability to comprehend and generate content related to visual information. This project proposes to explore techniques that combine the power of large language models with computer vision methodologies,	1-2	Dr. Rifat Shahriyar (rifat)

SL	Topic	Description	Students	Teacher
		enabling the models to understand and generate text based on visual input. The research will involve designing and training models that can effectively process and analyze images, extracting meaningful visual features to enrich the language model's understanding. Furthermore, the project will explore methods for fine-tuning the language model using visual-based objectives, allowing it to generate text that is coherent and aligned with the visual content. The proposal will have potential applications in areas such as visual question answering and multimodal summarization. References: VisualBERT: A Simple and Performant Baseline for Vision and Language An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale MultiModalQA: Complex Question Answering over Text, Tables and Images MSMO: Multimodal Summarization with Multimodal Output		
51	ML/DL without Python!	Python is the de-facto language for implementing machine learning/deep learning. PyTorch and TensorFlow are all Python-based frameworks. Unfortunately, Python is an inherently slow programming language. Recently (Feb 2023), Yann LeCun twitted 'ML would have advanced faster if another front-end language had been available and widely adopted instead of Python. One that is interactive yet fast & compilable, multithreaded (no GIL), isn't bloated, doesn't care about white spaces.' This thesis will explore other languages and whether they can replace Python as the de-facto. Mojo (https://www.modular.com/mojo) is claimed to be such a language that can achieve significant speedup over Python. However, Mojo is still under development. This thesis requires exploring different programming languages' capabilities, so strong programming/development interest/experience is preferred.	1-2	Dr. Rifat Shahriyar (rifat)

SL	Topic	Description	Students	Teacher
52	Table Structure Recognition and Table Summarization	The table structure recognition problem focuses on recognizing and organizing the table structure from the table image. First, detect each cell from the table image, then use the combination of positional features and the textual features of that cell to detect the shared relationship of that specific cell and other cells. The textual feature can be derived from any OCR output of the cell image block. Next, the table summarization system analyzes the contents of each column and row, like what type of information a specific row conveys, like an entity recognition system. The summary can be mapped or generated in any desired format; for example, it can be a JSON object or plain text. For reference, see this image.	1-2	Dr. Rifat Shahriyar (rifat)
53	Analyzing Large Language Models in Low-Resource Settings	We are witnessing a revolution in Natural Language Processing (NLP) with the rise of deep learning models like ChatGPT, GPT-4, and Llama. These models are collectively called Large Language Models (LLMs) owing to their large model size and huge pre-training data. Due to their sheer scale, these models have shown remarkable capabilities in language tasks such as sentiment analysis, summarization, natural language inference, code generation and so on. They also show interesting emergent properties never before seen in NLP models such as: interacting and solving problems by simply communicating in natural language prompts (prompt engineering and instruction fine-tuning), gaining state-of-the-art performance on datasets after learning from a few training examples (few-shot learning and in-context learning), displaying human-level reasoning skills (Chain-of-thought) and acting as knowledge hubs (LLM as Knowledge base). However, even though LLMs have achieved huge gains across NLP benchmarks, these models still struggle in certain use cases. One such use case involves solving problems in low-resource settings like NLP tasks in Bangla. While ongoing research has been discovering and applying methods to improve LLM performance in general, little work has been done in these problem settings. In this thesis, we will target this problem domain by:	1 or 2	Dr. Muhammad Abdullah Adnan (adnan)

SL	Topic	Description	Students	Teacher
		1. Analyzing the performance of LLMs on tasks in low-resource languages. 2. Developing effective prompts for low-resource settings. 3. Combining and modifying established ideas like "chain-of-thought" to these tasks. 4. Experimenting with new and unique prompting methods to improve performance on these tasks. We will start with a comprehensive survey of the current literature and available LLMs. This will help you become familiar with models like GPT-3, GPT-4, Llama, Bloom, Bard etc. as well as terms like in-context learning, prompt engineering, chain-of-thought reasoning, etc. Once you are settled in, we will explore and experiment with problems and datasets in low-resource languages to come up with novel solutions.		
54	Security and Privacy in Federated Learning	Machine learning needs massive data for training, and the cloud has made it more accessible. Distributed machine learning using the cloud has become very convenient. As a result, many organizations want to collaborate with data for training, which raises privacy concerns. Federated learning is a collaborative distributed machine learning technique in which multiple edge servers jointly learn a machine learning model: each edge server maintains a local model for its local training dataset, while a central cloud server maintains a global model via aggregating the local models from the edge servers. As a result, Federated learning opens doors for many applications. However, this also opens new attack surfaces for attackers. The security and privacy issue for federated learning and distributed machine learning is a relatively new but essential topic. As many edge servers collaborate, some of them can be adversarial and malicious. Adversarial edge servers can perform many types of attacks or breach privacy of the data. Moreover, the adversarial edge servers may try to hamper the global model's accuracy by sending poisonous local model updates.	1 or 2	Dr. Muhammad Abdullah Adnan (adnan)

SL	Topic	Description	Students	Teacher
		In this project, we shall develop a secure and privacy preserving federated learning framework for mitigation against different kinds of attacks. We shall investigate different types of attacks on federated and distributed machine learning and develop techniques for protection against them.		
55	Speech to Sign Language: A Deep Learning Approach for Bangla Sign Language Production	Language impacts the daily lives of members of any race, creed, and region of the world. Language helps express our feelings, desires, and queries to the world around us. Words, gestures and tone are utilized in union to portray a broad spectrum of emotion. The unique and diverse methods human beings can use to communicate through written and spoken language is a large part of what allows us to harness our innate ability to form lasting bonds with one another; separating mankind from the rest of the animal kingdom.

Communication of any kind is bi-directional. Bilingual translators are usually a two-way street: we can translate from one language, say English, to another, say Bangla and we can also do the reverse. That is the reason why we can see there are many interpreters in global conferences/meetings to easily translate between languages.

While this is true for people with speaking ability, it is not always the case for all people, especially for speech and hearing impaired persons. Mute people can only use sign language to express themselves. They can use some basic, predefined hand gestures for any words or sentences in order to communicate with others. However, the reverse is not always easy. By reverse we mean that when a person who is able to speak cannot easily communicate with mute and deaf. This is because mute people may have learnt sign language to express their emotions and feelings but the speaking people in general do not face the need to use sign language for communication. Consequently we can see when general people want to communicate with mute people, it is often challenging. This can be hugely frustrating, as it breaks the bridge between two persons.

To overcome the challenges mentioned above, we are in need of some kind of approach that can translate verbal Bangla speeches into Bangla Sign Language gestures in an automatic way. The goal | 1 or 2 | Dr. Muhammad Abdullah Adnan (adnan) |

SL	Topic	Description	Students	Teacher
		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. The outcome will be a portable application that will provide a bridge of communication between the hearing and hearing-impaired people.		
56	Geo-Distributed Data Analytics and Machine Learning	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.

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 machine learning 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. | 1 or 2 | Dr. Muhammad Abdullah Adnan (adnan) |
| 57 | Single-cell and Spatial Transcriptomics (ST) data analysis | 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.

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 | 3 | Dr. Mohammad Saifur Rahman (mrahman) |

SL	Topic	Description	Students	Teacher
		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.

Spatial transcriptomics (ST), on the other hand, captures gene expression from intact tissues at spot 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?

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 in the human DLPFC dataset. Noman et al. (CSE'17) experimented with further datasets involving skin cancer (melanoma) and breast cancer. He improved the pipeline to produce good spatial domains in diverse datasets. Their work is currently under revision in Bioinformatics. Asif et al., on the other hand, investigated the benefit of histology images in complementing the ST data in improving spatial domain detection processes. Further work could be done in this field to introduce and answer new research questions.

Fatima et al. (CSE'17) investigated cell to cell communication and identified that the communication scheme varies significantly across the spatial locations. Their analysis has also led to some candidates for cell non autonomous gene upregulation. The model they built could be used to analyze data from various diseases (such as alzheimer's) and identify novel genes related to these diseases. | | |

SL	Topic	Description	Students	Teacher
		I have 2 full-time post graduate students working in this area. Therefore, the undergraduate students are expected to get a great amount of mentorship. Each prospective student can expect to explore different ideas and problems. 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 Recorded presentations (work from CSE'17) 1. https://youtu.be/BvgHhIVM_z8 2. https://youtu.be/i9NYGgngz2s 3. https://youtu.be/8TiE8B5wCBw		
58	Parkinson's Analysis with Remote Tasks	This project is a collaboration with Dr. Ehsan Hoque, Associate Professor, University of Rochester. Our own Saiful and Tariq Adnan sirs (on Leave, pursuing PhD under Dr. Ehsan) are also collaborators in this project. Parkinson's disease (PD) is a neurodegenerative disorder that affects movement, speech, and coordination. Early diagnosis and treatment can improve the quality of life for people with PD. However, access to clinical diagnosis is limited in low and middle-income countries (LMICs). Park (Parkinson's Analysis with Remote Kinetic-tasks) is an online (https://parktest.net/index.html) system built by University of Rochester that enables the measurement of Parkinson's disease for anyone, anywhere - via webcam. Tithi and Noshin et al. (CSE'16) created a Bangla language version	1	Dr. Mohammad Saifur Rahman (mrahman)

SL	Topic	Description	Students	Teacher
		of the framework (https://parktestbangla.web.app/). Then they worked with collaborating Neurologists to collect data from patients and do some preliminary analysis. Zarif et al. (CSE'17) presented a novel machine learning method named PULSAR (Graph based Positive Unlabeled Learning with Multi Stream Adaptive Convolutions for Parkinson's Disease Recognition) for early diagnosis of PD using video data. The method was developed and evaluated using two datasets of PD patients and healthy controls: the Bangla PARK dataset, with 24 PD and 186 non-PD data, and the US PARK dataset, with 343 PD and 840 non-PD data. The method first extracts hand key points from the video data. Then, it constructs a spatio-temporal graph of the hand key points. Finally, it uses an adaptive graph convolutional neural network (AGCN) to learn features from the graph which is then classified to PD or Non-PD. The method was further extended to a multi stream approach by training 4 separate models - Joint Stream, Bone Stream, Velocity Stream and Acceleration Stream where the features are joint coordinates, bone lengths, velocity and acceleration respectively. The proposed model achieved 74.15% accuracy in the independent test set, which is at par with the accuracy of the expert neurologists. We would like to further investigate this dataset for an improved prediction performance. Zarif et al. only worked with the finger tapping task, but the dataset contains several other tasks which remain to be explored. Additionally, we have downloaded another large dataset - the TIM TREMOR Dataset. This dataset has sensor data along with the video stream, as well as depth map video along with the RGB data. Therefore, we would like to explore a multi-modal approach with this data to solve this prediction problem. So far, we have tried to detect PD from a single task. But a better model may be to utilize the different tasks combined as that is what the neurologists do. This should be explored as well. [Note: Zarif will also continue with this project. Therefore the prospective new student can expect to get good amount of mentorship]		

SL	Topic	Description	Students	Teacher
		Reference Material 1. https://youtu.be/G215GW3QyU8 2. https://parktestbangla.web.app/ 3. https://www.facebook.com/toKnowParkinsons 4. https://roc-hci.com/current-projects/park-parkinsons-analysis-with-remote-kinetic-tasks/ [See the reference papers]		
59	Phylogeny estimation	Phylogeny refers to the evolutionary relationships among a set of entities. phylogeny represents the evolution of a particular gene within a group of species. Quartet-based summary methods for phylogeny reconstruction have gained substantial interest. Examples include QMC, QFM and their weighted counterparts wQMC, wQFM etc. In this work we will investigate QFM and wQFM to improve their running time and memory footprint. We will also investigate ways to improve the quality of the tree produced by these methods. We will also explore some alignment free methods for phylogeny reconstruction from whole genomes. 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. Reference material 1. https://youtu.be/t9kWAncFbc8	1-2	Dr. Mohammad Saifur Rahman (mrahman)

SL	Topic	Description	Students	Teacher
60	Computer aided decision for a Shariah-compliant Investment Crowdfunding Platform	biniyog.io is a Shariah-compliant investment crowdfunding platform that connects small and medium businesses with retail investors for short-term debt financing. Business (SME) applies for funding and shares relevant documents. biniyog.io then assesses the SME's documents, financials and credit risk. If eligible, the SME is then featured on the platform with proper justification to help investors decide. Investors invest and also receive repayments through the platform. The platform would like to maximize accuracy in risk analysis. As it grows, it is going to be expensive (both from temporal and financial angle) to have humans painstakingly go over the financial and other documents submitted by the SMEs alongside their request for investment. Therefore, we need to develop algorithms to automatically conduct necessary analysis of the documents in hand. This includes analyzing sales and expense data, understanding transaction volume and patterns, assessing verifiability of given data, calculating optimal amount vs time frame of short-term external capital that the business can safely handle etc. However, SMEs typically do not have any uniform way of maintaining their daily financials, despite the rise of different digital solutions. This means we have to deal with sources as diverse as Vendor invoices (expenses), Carbon copies of client invoices (sales), Courier company reports (sales), Instant messaging (whatsapp/messenger) conversation history (sales), Handwritten (only sales), Handwritten (sales, expenses and/or credit sales – custom format), Handwritten (register book with alphabetically sorted pages to segment clients by), Excel sheets (custom format) Ecommerce order data, Point-of-sale (POS) transaction reports, Custom software reports etc. Another challenge is to verify the authenticity of the documents. Currently the platform relies on industry benchmarks and anomaly detection to ensure that the data has not been tampered with. Typically this is done by some custom excel formula that would vary from business to business. But given the variety of trends needed to be checked, there is always room for improvement over manual efforts.	1-2	Dr. Mohammad Saifur Rahman (mrahman)

SL	Topic	Description	Students	Teacher
		We will get first hand data from the company and conduct our own analysis to come up with novel solutions to the problem.		
61	Highly accurate and scalable algorithms for reconstructing phylogenetic trees (evolutionary trees)	(At least two projects on phylogenetic tree estimations will be assigned) The overarching goal of these 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 the ever-increasing abundance of molecular data. 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. In these particular projects, 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. We will contribute toward developing novel theoretical/algorithmic results as well as building fast, scalable, and efficient tools for estimating species trees. Prior knowledge of Biology is not required.	1-6	Dr. Md. Shamsuzzoha Bayzid (bayzid)

SL	Topic	Description	Students	Teacher
62	Predicting protein attributes using machine learning and language models	Proteins are considered 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, and protein-protein interactions) from protein sequences. Please see the following articles: 1. https://doi.org/10.1093/bioinformatics/btaa531 2. https://doi.org/10.1093/bib/bbab578 3. https://doi.org/10.1093/bioadv/vbad042 Prior knowledge of biology is not required.	1-6	Dr. Md. Shamsuzzoha Bayzid (bayzid)
63	Racially Diverse Cellular Models of Alzheimer's Disease for Better Diagnosis and Cure	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	1-6	Dr. Md. Shamsuzzoha Bayzid (bayzid)

SL	Topic	Description	Students	Teacher
		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. We will leverage CRISPR-Cas9 genome editing in this regard. **Hypothesis** Previous studies tested the associations of APOE ε alleles with sAD risk among Caucasians, African-Americans, Hispanics, and East Asians [1-7]. 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, 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. This project will be done in collaboration with: i) Icahn School of Medicine at Mount Sinai, New York, USA ii) Tokyo Metropolitan Institute of Medical Science, Japan.		
64	Study and design of methods to analyze scRNA-seq and spatial transcriptomics data	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	2	Dr. Atif Hasan Rahman (atif)

SL	Topic	Description	Students	Teacher
		• RNA velocity estimation in isoform resolution • Predicting cell location from scRNA-seq data • 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		
65	Structural variant calling in genomes using deep learning	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 small sequence variations as well as structural variations such as reversals, translocations, deletions, duplications, etc. Many of them are related to 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	2	Dr. Atif Hasan Rahman (atif)
66	Reference-free methods in computational biology	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 genome are associated with diseases and traits. We will develop a space efficient 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	2	Dr. Atif Hasan Rahman (atif)

SL	Topic	Description	Students	Teacher
67	Machine Learning Model for Prediction of Post Traumatic Stress Disorder	Mental health is of utmost importance as it directly influences our overall well-being and quality of life. The study of trauma highlights the critical link between mental health and traumatic experiences. Trauma can have long-lasting effects on an individual's psychological well-being, often resulting in conditions such as PTSD, anxiety, and depression. Understanding trauma and its impact allows mental health professionals to provide effective support and interventions to help individuals heal and recover. By addressing trauma-related issues, we can promote better mental health outcomes, enhance resilience, and restore a sense of safety and well-being in those who have experienced traumatic events.	1-2	Dr. Sadia Sharmin (sadia)
68	Authentication of Food reviews on Social Media. Influencing the Influencers ?	Social media analysis is not new; many people have already created exciting and innovative algorithms to study this. However, it is still a great data science research topic because it allows us to understand how people interact on social media. This is done by analyzing data from social media platforms to look for insights, bots, and recent societal trends. Once we understand these practices, we can use this information to improve our marketing efforts. For example, if we know that a particular demographic prefers a specific type of content, we can create more content that appeals to them. Social media analysis is also used to understand how people interact with brands on social media. This allows businesses to understand better what their customers want and need. Overall, social media analysis is valuable for anyone who wants to improve their marketing efforts or understand how customers interact with brands.	1-2	Dr. Sadia Sharmin (sadia)
69	Community Detection /Clustering: Principles of Big Graph	The concept of community detection has emerged in network science as a method for finding groups within complex systems through represented on a graph. In contrast to more traditional decomposition methods which seek a strict block diagonal or block triangular structure, community detection methods find subnetworks with statistically significantly more links between nodes in the same group than nodes in different groups. Clustering (also known as community detection in the context of graphs) methods for graphs/networks are designed to locate communities based on the	1-2	Dr. Sadia Sharmin (sadia)

SL	Topic	Description	Students	Teacher
		network topology, such as tightly connected groups of nodes. Before performing the community detection algorithm in graph-based clustering, the data is represented by a graph. Edge weight in the graph can be calculated using a variety of similarity measures. A community is formed when nodes in a network are of the same type.		
70	Data Analytics and Visualization	Data visualization techniques can be used to create charts, graphs, and other visual representations of data. This allows us to see the patterns and trends hidden in our data. Data visualization is also used to communicate results to others. This allows us to share our findings with others in a way that is easy to understand.	1-2	Dr. Sadia Sharmin (sadia)
71	An Experimental Evaluation of Label Generation Techniques	Learning complex machine learning (ML) models often requires large amounts of labeled data, which can be costly to acquire in absence of application feedback loops. Recently, we see a trend toward data programming or weak supervision in general to automatically generate noisy labels for large, unlabeled datasets as a means of pretraining, for example, deep neural networks. Common techniques include traditional boosting, label generation from conflicting labeling functions, the automatic generation of such labeling functions, and domain-specific object composition (e.g., placing a known object on a random background). This topic aims to conduct a systematic comparison of these techniques on different datasets, and ML algorithms in order to gain a better understanding under which settings these techniques yield overall improvements of model accuracy.	1-2	Dr. Sadia Sharmin (sadia)
72	Predictive modeling / Predictive Maintenance	Predictive modeling is a significant portion of data science and a topic you must be aware of. Simply put, it is the process of using historical data to build models that can predict future outcomes. Predictive modeling has many applications, from marketing and sales to financial forecasting and risk management.	1-2	Dr. Sadia Sharmin (sadia)

SL	Topic	Description	Students	Teacher
73	How Social Media Trends and offline Behavioral correlates	Social media have undeniably become a principal means of daily communication, information, socialization, personal expression, and entertainment. As such, the content shared on social media can have great resonance and potential to be popularized worldwide, affecting offline behaviors. We will study how Social Media Trends correlates with offline	1-2	Dr. Sadia Sharmin (sadia)
74	Online Optimization	Online optimization refers to the process of optimizing a system or making decisions in real-time as new data becomes available. Unlike offline optimization, where all the data is available upfront, online optimization deals with dynamic and evolving environments where decisions must be made continuously based on incoming data. In online optimization, the goal is to make effective and efficient decisions in a timely manner, considering the trade-offs between exploration and exploitation. The decisions are typically made iteratively, where at each step, the optimizer must consider the current state of the system, available information, and predefined objectives or constraints. Online optimization has applications in various domains, including finance, e-commerce, supply chain management, and resource allocation, etc. e.g. https://pubsonline.informs.org/doi/abs/10.1287/educ.1100.0072 https://pubsonline.informs.org/doi/10.1287/opre.2018.1763	1-3	Dr. Muhammad Ali Nayeem (ali_nayeem)
75	Simheuristics	Simheuristics is a field that combines simulation and metaheuristic optimization techniques to solve complex real-world problems. It involves using simulation models to represent and analyze the behavior of complex systems, and applying metaheuristic algorithms to optimize the decisions or parameters within those systems. In simheuristics, simulation provides a way to model and	1-3	Dr. Muhammad Ali Nayeem (ali_nayeem)

SL	Topic	Description	Students	Teacher
		understand the dynamics, constraints, and interactions of real-world systems. It allows for the evaluation of different scenarios and the generation of data that can be used to guide optimization. By simulating the behavior of the system under consideration, simheuristics enables the exploration of various decision options and the assessment of their impact. Metaheuristic algorithms, on the other hand, provide efficient and robust optimization strategies to search for high-quality solutions within complex and large solution spaces. These algorithms, such as genetic algorithms, particle swarm optimization, or simulated annealing, explore the space of possible solutions iteratively, utilizing heuristic techniques to guide the search process. e.g. https://www.sciencedirect.com/science/article/pii/S221471601500007X		
76	Leveraing Pareto Solutions to infer accurate Alignment/Phylogeny	We can generate a lot of alternative alignments/phylogenetic trees by employing multi-objective optimization. Can we leverage these solution space to find better solution compared to single-objective based methods? e.g https://www.sciencedirect.com/science/article/pii/S147692712200041X	1-2	Dr. Muhammad Ali Nayeem (ali_nayeem)
77	ROADEF challenge	We will participate this yearly competition and work based on the given task https://www.roadef.org/challenge/2022/en/index.php	1-2	Dr. Muhammad Ali Nayeem (ali_nayeem)
78	NeurIPS competition	We will participate this yearly competition and work based on the given task https://euro-neurips-vrp-2022.challenges.ortec.com/ This is a collaboration project with Dr. M Sohel Rahman.	1-2	Dr. Muhammad Ali Nayeem (ali_nayeem)

SL	Topic	Description	Students	Teacher
79	GECCO competitions	We will participate this yearly competition and work based on the given task https://gecco-2023.sigevo.org/Competitions	1-2	Dr. Muhammad Ali Nayeem (ali_nayeem)
80	Predictive maintenance of power stations at North-West Power Generation Company	Predictive maintenance of power stations using sensor data involves leveraging advanced analytics and machine learning techniques to monitor, analyze, and predict the maintenance needs of critical equipment within the power station. Sensors installed throughout the power station continuously collect data on various parameters such as temperature, pressure, vibration, and performance indicators of machinery and components. By applying predictive maintenance, power station operators can proactively identify and address potential equipment failures or maintenance requirements before they cause significant disruptions or costly downtime. By analyzing historical sensor data and patterns, machine learning models can be trained to detect early signs of anomalies, deviations from normal behavior, or indicators of equipment degradation. e.g. https://www.sciencedirect.com/science/article/pii/S0360835219304838 https://www.sciencedirect.com/science/article/pii/S095741742030525X	1-3	Dr. Muhammad Ali Nayeem (ali_nayeem)
81	Surrogate-assisted optimization leveraging deep learning	Designing surrogate-assisted optimization leveraging deep learning involves using deep learning techniques to enhance the efficiency and effectiveness of optimization algorithms. Traditional optimization methods can be computationally expensive and time-consuming, especially for complex and high-dimensional problems. Surrogate-assisted optimization addresses this challenge	1-3	Dr. Muhammad Ali Nayeem (ali_nayeem)

SL	Topic	Description	Students	Teacher
		by creating a surrogate model that approximates the objective function or constraints, allowing for faster evaluations and optimization iterations. e.g. https://www.sciencedirect.com/science/article/abs/pii/S0020025522012658		
82	Hybridizing Reinforcement Learning and Metaheuristic to design robust optimization framework	Hybridizing reinforcement learning and metaheuristic techniques involves combining the strengths of both approaches to design a robust optimization framework. Reinforcement learning (RL) focuses on learning optimal decision-making policies through trial and error interactions with an environment, while metaheuristic algorithms provide effective search and optimization strategies for complex problems. In many optimization scenarios, finding robust solutions that perform well across different environments or problem variations is crucial. However, traditional optimization methods often struggle to handle uncertainty or changing conditions. By combining RL and metaheuristics, we can create a framework that leverages the adaptability and exploration capabilities of RL with the global search and exploitation abilities of metaheuristic algorithms. e.g. https://www.sciencedirect.com/science/article/abs/pii/S0950705121003075	1-3	Dr. Muhammad Ali Nayeem (ali_nayeem)
83	Real-world Optimization using Metaheuristics	We want to tackle various optimization problems in different branches of science and engineering using metaheuristics techniques. No prior knowledge is necessary.	1-3	Dr. Muhammad Ali Nayeem (ali_nayeem)

SL	Topic	Description	Students	Teacher
84	Multi-objective Portfolio optimisation	Portfolio optimization using multi-objective optimization involves the process of designing an investment portfolio that maximizes returns while simultaneously minimizing risks. Traditional portfolio optimization focuses on finding a single optimal solution that balances risk and return. However, multi-objective optimization expands the scope by considering multiple conflicting objectives simultaneously, allowing investors to explore and select from a range of diverse portfolio options. e.g. https://www.sciencedirect.com/science/article/pii/S0305054821003427 https://www.sciencedirect.com/science/article/pii/S2210650218309702 https://www.kaggle.com/code/ashbellett/portfolio-optimisation	1-3	Dr. Muhammad Ali Nayeem (ali_nayeem)
85	Autonomous drone swarm intelligence	Optimizing swarm behaviors of autonomous drones involves enhancing the collective performance and coordination of a group of drones working together to achieve a common goal. Swarm robotics draws inspiration from the behavior of social insect colonies, such as ants or bees, where individual agents interact locally with their neighbors to accomplish complex tasks as a collective entity. In the case of autonomous drones, optimizing swarm behaviors typically involves addressing various aspects, including communication, task allocation, path planning, and obstacle avoidance. The goal is to improve the efficiency, robustness, scalability, and adaptability of the swarm, enabling it to handle complex missions or scenarios more effectively. e.g. https://www.science.org/doi/10.1126/scirobotics.aat3536	1-3	Dr. Muhammad Ali Nayeem (ali_nayeem)

SL	Topic	Description	Students	Teacher
86	Development of innovative consumer applications on top of recent research outcomes	We will Identify emerging technologies or advancements and brainstorm application ideas. Then we will define project scope and goals, providing a structured framework for student involvement. We will evaluate completed applications based on functionality, usability, and innovation.	1-3	Dr. Muhammad Ali Nayeem (ali_nayeem)
87	Application of Metaheuristics Algorithms in Leader Election in Distributed Systems	Leader election is an important problem in distributed computing systems. In this problem, when the leader is crashed, other nodes must elect another leader. Leader election is a challenging task in the cloud computing environments due to the occurrence of deadlock and node failures in resource sharing. Most of the existing algorithms do not consider maximum available resources while electing leaders due to increased complexity. The leader election problem can be seen as an optimization problem considering several factors like maximum available resource, minimum communication overhead etc. A recent work [1] applies genetic algorithms to the leader election problem. There exists an opportunity to apply other metaheuristic algorithms. In this project, we would like to apply metaheuristic algorithms like crow search, monarch butterfly to solve the leader election problem considering resource optimization.	2	Dr. Rezwana Reaz (rezwana)
88	Secure Task Scheduling in fog-cloud platform for IoT applications	Task scheduling is an important problem for IoT applications as some IoT applications require immediate response. Fog-cloud based framework is a convenient solution to design task scheduling algorithms for IoT applications. In a fog-cloud based architecture, some delay sensitive tasks are executed in fog, whereas some delay tolerant tasks are executed in the cloud. For a task to be executed in a fog node, the corresponding network layer traffic is redirected to the fog node and for a task to be executed in the cloud, the corresponding network layer traffic is redirected to the cloud. To enhance the security of IoT applications, IoT traffic needs to be filtered. Traffic filtering can be done either in fog or cloud. IoT task scheduling and IoT firewall implementation are distinct research	2	Dr. Rezwana Reaz (rezwana)

SL	Topic	Description	Students	Teacher
		problems with different goals. In this work, we would like to merge these two problems and obtain a single solution that addresses both. Initial Reading [1] IoT Task Scheduling [2] IoT firewall		
89	IoT-Blockchain based Remote Patient Monitoring	Internet of Things (IoT) and Blockchain technologies have been utilized heavily in diverse domains, including remote patient monitoring (RPM). IoT data is centrally collected, processed, and stored. But this centralization may face different problems including single-point failure, data tampering, privacy issues, etc. Blockchain can address these problems using its decentralized architecture. Hence, the combination of IoT and blockchain is a plausible solution to design a smart RPM system. F. Ashraf et al. [1] proposed a comprehensive IoT-blockchain based RPM framework that addresses all aspects of designing an RPM system. In this project, we would like to design the access control mechanisms and smart contracts for this framework. Initial Reading: [1] IoT Blockchain in RPM [2] Attribute based access control	2	Dr. Rezwana Reaz (rezwana)