

February 25th, 2021

Undergraduate Thesis Topic Allocation

All the prospective students of L4/T1 of January 2021 term are advised to look at the thesis topics published by the different thesis supervisors (See below). The students are encouraged to make a list of several topics that interest them. The allocation of topics will be done according to merit order of the students.

An online meeting to allocate the thesis topics will be held on February 28th, 2021 (Sunday) at 7:00 PM. The Zoom meeting link is:

<https://bdren.zoom.us/j/69041431587?pwd=bytDaTRGWEJSWHJ2dIBJYVQ0SIRWUT09>



Dr. Mohammad Saifur Rahman
Associate Professor & BUGS Secretary
Department of CSE, BUET

Thesis Topics

S L	Topic	Description	Students	Teacher
1	Knowledge graph representations of PCAP data, finding any hidden inferences	<p>In knowledge representation and reasoning, a knowledge graph is a knowledge base that uses a graph-structured data model or topology to integrate data. Knowledge graphs are often used to store interlinked descriptions of entities – objects, events, situations, or abstract concepts – with free-form semantics.</p> <p>In the field of computer network administration, PCAP is an application programming interface (API) for capturing network traffic.</p> <p>We want to find out if we can infer any issues which are hidden from this data using semantic techniques.</p>	1 or 2	Dr. Muhammad Masroor Ali (mmasroor ali)
2	Knowledge graph representations of a machine state where data are input from using osquery for monitoring system security	<p>In knowledge representation and reasoning, a knowledge graph is a knowledge base that uses a graph-structured data model or topology to integrate data. Knowledge graphs are often used to store interlinked descriptions of entities – objects, events, situations, or abstract concepts – with free-form semantics.</p> <p>osquery is a tool that exposes an operating system as a high-performance relational database. It enables developers to write SQL-based queries that explore operating system data. With osquery, SQL tables can be created to help represent otherwise fairly abstract concepts, such as: Running processes.</p> <p>We will want to explore the osquery results to find out whether any system security issues are present.</p>	1 or 2	Dr. Muhammad Masroor Ali (mmasroor ali)

S L	Topic	Description	Stu den ts	Teacher
3	<p>Knowledge graph representations of a machine state where data are input from using osquery. Can temporal knowledge graphs be used to find out any change in the system over time which signifies a security event?</p>	<p>In knowledge representation and reasoning, a knowledge graph is a knowledge base that uses a graph-structured data model or topology to integrate data. Knowledge graphs are often used to store interlinked descriptions of entities – objects, events, situations, or abstract concepts – with free-form semantics. In temporal knowledge graphs, relations between entities may only hold for a time,</p> <p>osquery is a tool that exposes an operating system as a high-performance relational database. It enables developers to write SQL-based queries that explore operating system data. With osquery, SQL tables can be created to help represent otherwise fairly abstract concepts, such as: Running processes.</p> <p>We will want to explore the temporal changes in osquery results to find out whether any system security issues are present.</p>	1 or 2	Dr. Muhammad Masroor Ali (mmasroor ali)
4	<p>Knowledge graph embedding of text corpus. Can we use this for sentiment or emotion detection?</p>	<p>In knowledge representation and reasoning, a knowledge graph is a knowledge base that uses a graph-structured data model or topology to integrate data. Knowledge graphs are often used to store interlinked descriptions of entities – objects, events, situations, or abstract concepts – with free-form semantics. Knowledge graph embedding aims to represent entities and relations in a large-scale knowledge graph as elements in a continuous vector space. Existing methods, for example, TransE, TransH, and TransR, learn the embedding representation by defining a global margin-based loss function over the data.</p> <p>We will want to explore whether we can use link prediction techniques in knowledge graphs for sentiment or emotion detection.</p>	2	Dr. Muhammad Masroor Ali (mmasroor ali)

S L	Topic	Description	Stu den ts	Teacher
5	Algorithmic Bioinformatics	The efficiency of algorithms or of data representation is often the difference between feasible and infeasible data analysis. By improving algorithms, we enable research that would not otherwise be possible. Our research work will be concerned with the development of efficient algorithms and efficient implementation of software for bioinformatics problems.	2 to 3	Dr. Md. Abul Kashem Mia (kashem)
6	Graph Algorithms	The final topic on graph algorithms will be fixed by discussing with students.	2 to 3	Dr. Md. Abul Kashem Mia (kashem)

S L	Topic	Description	Stu den ts	Teacher
7	Network Science	<p>I am copying some parts of the introduction chapter of the book Network Sciene. Read andd realize whether you are interested to on the area or not.</p> <p>http://networksciencebook.com/</p> <p><i>"I think the next century will be the century of complexity." Stephen Hawking</i></p> <p>We are surrounded by systems that are hopelessly complicated. Consider for example the society that requires cooperation between billions of individuals, or communications infrastructures that integrate billions of cell phones with computers and satellites. Our ability to reason and comprehend our world requires the coherent activity of billions of neurons in our brain. Our biological existence is rooted in seamless interactions between thousands of genes and metabolites within our cells.</p> <p>These systems are collectively called <i>complex systems</i>, capturing the fact that it is difficult to derive their collective behavior from a knowledge of the system's components. Given the important role complex systems play in our daily life, in science and in economy, their understanding, mathematical description, prediction, and eventually control is one of the major intellectual and scientific challenges of the 21st century.</p> <p>The emergence of network science at the dawn of the 21st century is a vivid demonstration that science can live up to this challenge. Indeed, behind each complex system there is an intricate network that encodes the interactions between the system's components:</p> <ol style="list-style-type: none"> 1. The network encoding the interactions between genes, proteins, and metabolites integrates these components into live cells. The very existence of this <i>cellular network</i> is a prerequisite of life. 2. The wiring diagram capturing the connections between neurons, called the <i>neural network</i>, holds the key to our understanding of how the brain functions and to our consciousness. 3. The sum of all professional, friendship, and family ties, often called the <i>social network</i>, is the fabric of the society and determines the spread of knowledge, behavior and resources. 4. <i>Communication networks</i>, describing which communication devices interact with each other, through wired internet connections or wireless links, are at the heart of the modern communication system. 5. The <i>power grid</i>, a network of generators and transmission lines, supplies with energy virtually all modern technology. 6. <i>Trade networks</i> maintain our ability to exchange goods and services, being responsible for the material prosperity that the world has enjoyed since WWII (Image 1.2). <p>Networks are also at the heart of some of the most revolutionary technologies of the 21st century, empowering everything from Google to Facebook, CISCO, and Twitter. At the end, networks permeate science, technology, business and nature to a much higher degree than it may be evident upon a casual inspection. Consequently, <i>we will never understand</i></p>	1 to 8	Dr. Md. Saidur Rahman (saidurrahman)

S L	Topic	Description	Stu den ts	Teacher
8	Drawing of Planar and Non-Planar Graphs	see https://cse.buet.ac.bd/research/group/gd/	1 to 8	Dr. Md. Saidur Rahman (saidurrahman)
9	Secret Sharing Scheme	A secret sharing scheme is a method of sharing a secret among a set of participants. This is done in such a way that only certain subsets of participants, called qualified subsets, can reconstruct the secret using their shares. Objective is to develop new secret sharing scheme.	1 or 2	Dr. Md. Saidur Rahman (saidurrahman)
10	Graph Theory, Graph Algorithms, Approximation Algorithms, Online Algorithms, Bioinformatics, Graph Neural Networks, Graph Data Mining, Heuristics and Metaheuristics	Actual topics will be fixed discussing with students according to their interest and capability.	1 to 8	Dr. Md. Saidur Rahman (saidurrahman)
11	Incremental Sampling for Combinatorial Optimization and Program Synthesis (Machine Learning)	Within machine learning, sampling is useful for generating diverse outputs from a trained model. Sampling without replacement is often desirable when duplicate samples are not beneficial. The aim of this research work is to devise a sampling algorithm that is applicable to many domains, e.g., program synthesis and combinatorial optimization. Two key components of the algorithm could be sampling without replacement and incremental sampling.	2/3	Dr. Md. Monirul Islam (mdmonirulislam)

S L	Topic	Description	Stu den ts	Teacher
1 2	Handwritten Bangla Character Recognition (NLP)	Automatic handwritten character recognition has immense academic and practical interests. The main challenge in such a recognition system is to deal with the enormous variety of handwriting styles by different writers. Bangla is one of the most spoken languages and ranked fifth in the world and spoken by more than 200 million people. Although there have been many attempts in devising algorithms for hand written English characters, there are only few works for Bangla. The aim of this research work is to develop deep learning-based algorithms for recognizing hand written Bangla characters.	2/3	Dr. Md. Monirul Islam (mdmonirulislam)
1 3	Evolving Neural Network Ensemble (Evolutionary Computing and Neural Network)	Using a coordinated group of simple solvers to tackle a complex problem could be traced back hundreds of years ago when ancient Chinese suggested a team approach to problem solving. For a long time, scientists/engineers have used divide-and-conquer strategies to decompose a complex problem into simpler sub-problems and then solve them by a group of solvers. However, knowing the best way to divide a complex problem into simpler ones relies heavily on the available domain knowledge. Fortunately, evolutionary computation provides some of the interesting avenues to automatic divide-and-conquer methods. The aim of this research work is to devise such an algorithm for some real-world complex problems.	2	Dr. Md. Monirul Islam (mdmonirulislam)
1 4	Machine Learning for Real-world Complex Problems	The aim here is to apply machine learning algorithms on interesting and challenging real-world problems.	2	Dr. Md. Monirul Islam (mdmonirulislam)

S L	Topic	Description	Stu den ts	Teacher
1 5	Study on the Technology for Guiding Visually Impaired Persons in Crowded Places	<p>Traditionally the visually impaired people use white stick while walking alone in the street. It only helps to identify the objects on the way. Informing about the location through technology must be a great help for the unlucky person. GPS can be easily used to find the location. Where GPS a good technology to inform a particular location it is also necessary to identify the objects around him so that the visually impaired person can identify the crowd. We can find two different objects; one is permanent structures which are not moving such as foot over bridge, shops, bus stops, foot paths etc, the other are moving objects such as moving persons and vehicles. A system of identifying known person through image recognition might give the feeling of having eyesight for a visually impaired person. The system might suggest the visually impaired person to find a safe and optimal path to the destination.</p> <p>Please find the following researches and products in this related field as follows:</p> <p>Ultrasound technology used to guide the blind</p> <p>https://www.youtube.com/watch?v=agR9YLio_00&feature=share&fbclid=IwAR0h2mne79LJr1bLyPMggGidi-s4qAe5ffwFtgYnnT9dMYo-AP9bgIOv0Ug</p> <p>Introducing Envision Glasses: AI-powered smartglasses for the Blind & Visually Impaired</p> <p>https://www.youtube.com/watch?v=9ehENnq2EFo&feature=share&fbclid=IwAR1-tVakbTt0-w5SeO1AY1uwi6oW9yFsfU_rDdH-9QyIbqJa2kZujPJakVY</p> <p>Visually impaired people are regaining their independence with talking glasses</p> <p>https://www.youtube.com/watch?v=LZYfhB68NzA&feature=share&fbclid=IwAR0F-VtbPkCzEmDhsfEAnvwlfsZ7BIRKB-4hVRgRpiM1IHEy2argwLHQfco</p> <p>Replacing white sticks with electronic devices: new technologies for the visually impaired</p> <p>https://www.youtube.com/watch?v=BVZSqtjggiI&feature=share&fbclid=IwAR0KYTbbCPHtKPIHEvJVqh12iSLz0t165NAj-ndL-w475cDU0XQnaOWIqek</p> <p>Seeing is believing: Glasses give sight back to the blind</p> <p>https://www.youtube.com/watch?v=RHRxiuxP3QU&feature=share&fbclid=IwAR2P37Hab53eo_q_lBiYwCJmV6g0G2ld7rcvrUAmE4jDH75cerrS-ei6UU</p> <p>Smart vest-wear tech helps the blind to walk unaided</p> <p>https://www.youtube.com/watch?v=by0aMBQDjyk&feature=share&fbclid=IwAR2MljceZOSd0vG1jybuS-tBfpEkfK2SMwkLjYpAY_TtBgdmu9DDqyZigp8</p>	1	Dr. Md. Mostofa Akbar (mostofa)

S L	Topic	Description	Stu den ts	Teacher
1 6	Designing Controller for Universal Display Systems using FPGA	<p>We see lots of display systems in commercial places and service centers. Configuration of display system makes it reusable and attractive as it serves the purpose of the customers in a greater extent. There must be a controller that drives this display system. There are several ways to build this controller. Arduino board and ATmega microcontroller systems can be easily programmed for such controller. These systems are reliable and robust but they have limitations too. They need more space and the processing power is limited due to sequential nature of processing. We are looking forward to a system that will be designed by using HDL to implement the desired controller circuit. The RTL (Register Transfer Level) coding using Verilog could be used to generate the gate level design of such systems. We can easily generate the FPGA implementation of the design by the designing tools and the burnt FPGA can be used directly as a controller of the system. Our target is to develop a controller that can be configured from cloud based system. Any serial interface or wireless interface could be used for establishing the communication with the system to the cloud. The research will address the issues faced in implementing and using these display systems in different environments.</p>	2	Dr. Md. Mostofa Akbar (mostofa)
1 7	Designing Environment Monitoring Systems (EMS) using FPGA	<p>Environment Monitoring System is generally used monitoring different environment related parameters such as temperature, motion, presence of water and humidity. Sensor will be used to collect information of such environment parameters. These data will be sent to a EMS server periodically. The communication between the sensor and EMS server will be done through wireless or wired interface. The main focus of the research is to design the sensors. The RTL (Register Transfer Level) coding using Verilog could be used to generate the gate level design the circuits to accumulate data in the sensor systems. The communication could be two way as the EMS server may pull the data from the sensors in a regular fashion. The main research challenge is to design a Environment Monitoring System that could be applicable in any environment through easy configuration. Use of FPGA makes the system more configurable. FPGAs can be programmed easily to change the system with less effort.</p>	1	Dr. Md. Mostofa Akbar (mostofa)

S L	Topic	Description	Stu den ts	Teacher
18	Designing Low Cost FPGA based Weighing Machine for Vegetable Market	<p>The weighing machines are heavily used in vegetable market where the shopkeeper can enter the price of the vegetable and it automatically calculates the sale amount by multiplying the unit price with the weight calculated. Such weighing machine has following shortcomings:</p> <ul style="list-style-type: none"> • The displays are not in the local language, i.e., in Bangla • You have to enter the price again and again if it is different in price • A receipt cannot be generated for the sale amount <p>Our target is to design a scale with the following features, which is equivalent to a mini POS (Point of Sale)</p> <ul style="list-style-type: none"> • Show the display in local language • Configuring the price in different buttons • Configuration from your mobile phone so the machine must be integrated with the cloud • By pressing a button it will show the unit price in the display and the image of the vegetable • You can increase or decrease the negotiated price • Storing the sale items in the memory • Displaying sale receipt when a sale is completed • Interfacing with the mobile phone to generate sales receipt. <p>The design will be done using RTL code of Verilog and the design will be implemented by burning into FPGA.</p>	2	Dr. Md. Mostofa Akbar (mostofa)
19	USE of AI, ML and IR Techniques for Intern Evaluation	<p>In Virtual Internship System, The interns work in a virtual industry environment solving industry level problems. The qualitative assessment like communication skill, leadership skill etc. of an intern has to be evaluated using AI, Machine learning and Information retrieval techniques on Interns performance data.</p>	2	Dr. Abu Sayed Md. Latiful Hoque (asmlatiful hoque)
20	Big Data Analytics on Health Data	<p>Data warehouse will be designed for national clinical data. Online Analytical Processing (OLAP) will be used to support decision queries for the development of better health care services.</p>	2	Dr. Abu Sayed Md. Latiful Hoque (asmlatiful hoque)
21	Comparison of High Performance Data Models	<p>Big data requires high performance in terms storage, query and transaction processing. Parallel and Distributed Database Architecture for relational, NOSQL and columnar models will be evaluated in the application of high-performance query and transaction processing</p>	2	Dr. Abu Sayed Md. Latiful Hoque (asmlatiful hoque)

S L	Topic	Description	Stu den ts	Teacher
2 2	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. Occasionally this can be supplemented with experimental validation.	1 or 2	Dr. M. Sohel Rahman (msrahman)
2 3	Applied Machine Learning	In this topic the goal is to apply appropriate machine learning (including deep learning) techniques on various problems in different domains including but not limited to: <ol style="list-style-type: none"> 1. Computer vision and image processing: here the primary focus would be on medical imaging. 2. Bioinformatics: Various classification/prediction problems in bioinformatics will be considered. 3. Interdisciplinary projects: Here the goal is to work in other domain (e.g., groundwater, climate, physics, medical domain etc.). 4. etc. 	Any	Dr. M. Sohel Rahman (msrahman)
2 4	Quantum Computing	We want to study quantum computing under this project and advance the current state of the art. Disclaimer: I am not an expert here. But we can work together to do something good in sha ALLAAH.	1-2	Dr. M. Sohel Rahman (msrahman)
2 5	Metaheuristics Optimazation	We want to tackle various optimization problems in different branches of science and engineering using metaheuristics techniques.	1-3	Dr. M. Sohel Rahman (msrahman)
2 6	Computational Modelin g	We want 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. We plan to use state of the art modeling paradigms (e.g., agent based models, SIR types models etc.).	1-3	Dr. M. Sohel Rahman (msrahman)
2 7	Botnet Detection	The botnet refers to a collection of bots (computers) controlled by a bot-master (attacker). Botnet detection is a promising research area that attracted many researchers in the past. Botnets pose major security threats such as accessing private sensitive information or denying legitimate users from accessing network resources. A recent study shows that (surprisingly) about 40% of computers connected to the Internet are infected by bots and controlled by bot-masters. In this thesis, we will apply machine learning techniques to detect Botnets.	2	Dr. A.K.M. Ashikur Rahman (ashikurrahman)
2 8	Accident Analysis of Dhaka City	Traffic accidents and injuries related to them are regular phenomena for the people of Bangladesh. Being one of the most crowded cities of the world, Dhaka city exhibits more accidents than other cities of the country. This thesis aims at identifying the most hazardous regions for such incidents within the Dhaka Metropolitan Region as well as assesses their influences.	2	Dr. A.K.M. Ashikur Rahman (ashikurrahman)

S L	Topic	Description	Stu den ts	Teacher
29	Deception Detection	Deception is the act of making someone believe something that is not true. If someone can through a lie by mixing it with some truth, then it becomes very easy to deceive others. However, during deception, human body's physiological parameters change which are very difficult to hide. In this thesis we will analyze video, audio, EEG, Eye Gaze etc for multimodal deception detection.	2	Dr. A.K.M. Ashikur Rahman (ashikurrahman)
30	Spatial Graph Embedding for Neighborhood Learning	<p>In this project, we will leverage the concept of graph neural networks to capture the geo-spatial context of the neighborhood of an entity. The Geo-Spatial Network Embedding (GSNE), that learns the embeddings of the neighborhood of an entity has a number of applications that include house price predictions, tourist spot recommendation, etc. For example, a house next to a good school will likely have a much higher price than the house which is far away from the school. We also plan to explore how other complex features such as textual description and images can be embedded in the multi-modal feature space to further enhance geo-spatial embedding.</p> <p>Research Areas: Deep Learning, Graph Embedding, Spatial Database</p>	1 to 2	Dr. Mohammed Eunus Ali (eunus)
31	Can You Make Your Chatbot Empathetic: Designing a Friendly Bengali Adolescent Chatbot	<p>Chatbots technologies are getting popular day by day in assisting people in numerous ways. Though this area has got much attention in the context of popular foreign languages (e.g., English, Chinese, etc.), research on Bengali chatbot is still at its infancy. In this project, we will build an Bengali chatbot particularly focusing on assisting adolescent health problems in Bangladesh. Building a chatbot is always challenging as it needs to consider users emotions, sentiments, local social context, etc. Thus, the main focus of this project will be developing an adolescent chatbot in Bengali language that addresses the following issues: developing users' empathy level, growing positive reinforcement, detecting users' sentiments and emotions, encountering regional or traditional non-bookish terms, etc. We will also figure out technical and implementation level challenges of building such technology in a specific sensitive health domain in Bengali throughout this study.</p> <p>Research Areas: HCI, Deep Learning, Bangla NLP</p>	1 to 2	Dr. Mohammed Eunus Ali (eunus)

S L	Topic	Description	Stu den ts	Teacher
3 2	Recommending Trips for Friends	<p>Location-based services (LBS) have developed a significant importance in recent years. Almost every modern social networking site, commonly referred as Location-based Social Networks (LBSNs), i.e., Foursquare, Yelp, Flickr, etc., have enabled their locations features. With the location enabled features, users can tag their daily activities in different locations through check-ins. These check-ins capture socio-spatial preferences of users and social groups that are being heavily used in many applications such as recommendation systems, location based advertising, and so on. However, not all the check-in locations are equally important to each member of socially cohesive user groups as the locations may reside far from their activity areas, or may not have considerable interests. Hence, it is essential to identify a selective set of co-engaged locations from the social network which are important to socially cohesive user groups. In this project, we will identify a set of co-engaged locations for a group of friends which can greatly enhance user experience and togetherness in a spatial region.</p> <p>Research Areas: Spatial Learning, Location Based Services</p>	1	Dr. Mohammed Eunos Ali (eunos)
3 3	Distinguishing Extremely Similar Rice Diseases	<p>An accurate and timely detection of diseases and pests in rice plants can help farmers in applying timely treatment on the plants and thereby can reduce the economic losses substantially. Though most of the diseases and pests are easily distinguishable by human or by traditional deep learning architecture, the appearance of some of the diseases and pests are so close that they are even hard for an expert to distinguish them. In that particular case, an expert needs to investigate different parts of plants to differentiate those diseases and Pests. In this work, we will work with BRRI's scientists to solve their problems by developing new deep learning models that can differentiate extremely similar rice disease from the captured images.</p> <p>Research Areas: Image Processing, Deep Learning</p>	1	Dr. Mohammed Eunos Ali (eunos)
3 4	Keyboard Agnostic Bangla Writing: A Personalized Phonetic Based Approach	<p>We all face difficulties in typing in Bangla using existing phonetic keyboards. In particular, if a user cannot memorize the rules and guidelines of the keyboard, the typing becomes erroneous. In this project, we want to build a keyboard agnostic Bengali writing where you don't need to remember any keyboard layout. This AI based keyboard will exploit the context and personalized writing patterns to assist you in hassle-free Bangla writing.</p> <p>Research Areas: Bangla NLP, Deep Learning (Number of students 1/2)</p>	1 to 2	Dr. Mohammed Eunos Ali (eunos)

S L	Topic	Description	Stu den ts	Teacher
3 5	Bangla NLP / Reinforc ement Learning / Bayesian Learning	This is a free topic. If you have any exciting topic in mind that matches with the title. You can choose this topic as your thesis.	1 to 2	Dr. Mohamme d Eunos Ali (eunos)
3 6	Study of Cognitiv e Science	Cognitive science is the interdisciplinary, scientific study of the mind and its processes. In this thesis, we will study intelligence and behavior, transformation process of information of human mind. Research Area: HCI, NLP, Health, Applied Machine Learning	2 to 4	Dr. Mahmuda Naznin (mahmuda naznin)
3 7	Network automati on and Quality of Services	In recent years, network automation and network resource management methods have become more data-driven, machine learning (ML) oriented. In this thesis, we will explore these area. Research Area: Cloud Computing, Data-Driven Model, Network Security, Applied Machine Learning	2 to 4	Dr. Mahmuda Naznin (mahmuda naznin)
3 8	Medical Imaging	Medical Imaging for early diagnosis. Research Area- Biomedical Applications, Applied Machine Learning	2	Dr. Mahmuda Naznin (mahmuda naznin)

S L	Topic	Description	Students	Teacher
39	Image processing and computer vision	<p>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.</p> <p>Image & video search and retrieval: this includes feature extraction, image representation and retrieval algorithm design.</p> <p>Security and surveillance: there are many biometric and non-biometric applications like face, iris, finger print, palm print recognition, crime scene analysis</p> <p>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</p> <p>Document Analysis/Classification/Preservation: Character recognition, document binarization Object and shape recognition, historical document enhancement, land map digitization, etc</p> <p>Intelligent transport system: license plate detection/recognition, road sign detection/recognition, etc</p> <p>Environmental monitoring and remote sensing/satellite imaging: land usage determination, changes in land usage, satellite image enhancement/de-noising</p> <p>Image enhancement: improve the visual quality of images using different methods</p> <p>Image de-noising: removing noises from images</p> <p>Image segmentation: dividing images into different semantically meaningful regions, e.g., object, etc</p> <p>Image Security: this is not human's security; rather the security of image itself. Some areas are: watermarking and registration, image steganography, image tampering/forgery detection, forgery detection of bank notes, etc</p> <p>Image compression: represent images compactly for transmission and storage.</p> <p>Image classification: too many applications, e.g., disease classification, OCR, leaf classification, crop disease classification</p>	1 to 6	Dr. Md. Monirul Islam (mmislam)

S L	Topic	Description	Stu den ts	Teacher
4 0	An Interpretable Deep Learning Model COVID Infection Rate Prediction	<p>Research Area: Deep Learning, Explainable AI</p> <p>Knowing the COVID infection rate in advance will allow the government to divide a city into zones based on the risk level and take preventive measures. Besides prediction accuracy, interpretability is also an important criterion for a COVID infection rate prediction model. However, most of the models that show high accuracy are not interpretable. Having the insights behind a prediction (e.g., a specific shopping center will be responsible for high infection rate) will enable the authority to take steps (i.e., restricting the access to the shopping center) to safeguard the people from the adverse effects of COVID pandemic.</p> <p>In this research project, we will develop an interpretable deep learning model to predict the COVID infection rate of an area.</p>	1	Dr. Tanzima Hashem (tanzimah ashem)
4 1	Transfer Learning for Crime Prediction in Multiple Cities	<p>Area: Machine Learning, Spatio-temporal Prediction</p> <p>Crime prediction and taking preventive measures play an important role to ensure the safety of the inhabitants of a smart city. Crime events exhibit spatial and temporal correlations; functionalities and urban characteristics of a region like points of interests (POIs) and traffic flow have direct influence on its crime occurrences. Crime prediction models capture the spatio-temporal correlations based on the past crime occurrences, and learn meaningful external features and their probable influence on crime. The accuracy of a model depends on the availability large scale training data. However, not all cities or all regions of a city have arrangements or facilities to collect relevant data for crime prediction (e.g., Dhaka city).</p> <p>In this research project, we aim to address the data scarcity problem of a city. We will develop a crime prediction model for a data insufficient city by learning from the urban dynamics of other cities that have abundant data for the crime prediction.</p>	1-2	Dr. Tanzima Hashem (tanzimah ashem)
4 2	Safe Path Learning in Road Networks	<p>Area: Location-based Services, Spatial Databases, Machine Learning</p> <p>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 paths for the travelers. However, the shortest or the fastest route is not always the best choice. While travelling on roads, people face many inconveniences like theft and pick-pocketing; women face harassment like eve-teasing. People would like to travel a little bit longer on a safer route that avoids those inconveniences. During the outbreak of an infectious disease like COVID-19, a pedestrian may want to avoid a crowded road to keep herself safe from infection.</p> <p>To meet the traveler's need on roads, in this research project we envision to build a model to learn safe paths between two locations on road by exploiting historical movement, accident, crime and harassment data.</p>	1	Dr. Tanzima Hashem (tanzimah ashem)

S L	Topic	Description	Stu den ts	Teacher
4 3	A Novel Dry Mucous Membrane Detection Technique using Artificial Intelligence	<p>Area: Ubiquitous Computing, Mobile Health, AI</p> <p>Medical conditions like fever, diarrhoea, vomiting as well as intense physical activities and poor intake of fluids may result in dehydration. Dehydration can cause different health complications that range from fatigue to brain damage and death. Dehydration results in dry mucous membranes in lip and tongue. In this research project, we envision to develop a diagnostic tool to detect dry mucous membrane using artificial intelligence (AI) based techniques. This tool will allow people to preliminary assess their hydration level. Knowing the level of dehydration early can help people to decide whether the dehydration can be treated at home or they need to go to the hospital immediately.</p> <p>No. of students: 2</p>	2	Dr. Tanzima Hashem (tanzimah ashem)
4 4	An Automated Alertness Assessment System	<p>Area: Ubiquitous Computing, Machine Learning, Image Processing</p> <p>Automated detection of a driver's alertness while driving may help to warn the driver and thereby reduce the road accidents is a well-studied problem. The alertness detection system can be also used to assess the general condition of a sick person (e.g., a person's general appearance may look as alert or restless or lethargic in case of no, mild and moderate dehydration, respectively) or identify the attention level of students/audience. The definition of the alertness measure criteria need to be varied across different application domain. In this research project, we envision to build a domain specific model to automatically assess a person's alertness state by exploiting the change in the facial features and movements.</p>	1	Dr. Tanzima Hashem (tanzimah ashem)
4 5	Security and Privacy in Machine Learning	<p>Machine Learning has now become pervasive. New systems and models are being deployed in every domain, leading to widespread deployment of software based inference and decision making. There is growing concern that ML exposes new vulnerabilities in software systems. ML models may be attacked by model inversion attacks, evasion attacks, poisoning attacks, model theft attack, etc.</p> <p>We will investigate ML attack surface, its security and privacy, training and inference in adversarial setting, focusing on possible attacks on these systems and available defenses.</p> <p>This work will be done in collaboration with Dr. ShaguftaMehnaz from Dartmouth college.</p>	2	Dr. Md. Shohrab Hossain (mshohrab hossain)

S L	Topic	Description	Stu den ts	Teacher
4 6	Intrusion Detection using Security Onion	<p>Log Management and Intrusion Detection solutions have been evolving for years. Security solutions are very expensive and sometimes does not meet the requirement of the organization. Therefore, it is essential to build custom built security solutions integrating different open source security tools.</p> <p>Security Onion is a Linux distribution for Intrusion Detection and Network Security Monitoring. It is based on Ubuntu and contains Snort, Suricata, Bro, Sguil, Squert, Snorby, ELSA, Xplico, NetworkMiner, and many other security tools.</p> <p>In this work, students will propose and build a custom open source intrusion detection system using various tools available in Security Onion.</p>	2	Dr. Md. Shohrab Hossain (mshohrabhossain)
4 7	Widening IoT Security: 3rd-Party Authentication in Federated Cloud, Edge, and Fog Systems	<p>This is an ongoing research work with Prof. Ying-Dar Lin from NCTU, Taiwan.</p> <p>Cloud and Edge computing paradigms provide storage and computing services to traditional and internet of Things devices. One computing platform is not suitable to fulfill the requirements of all IoT devices because of their heterogeneity. In this regard, a federation of various computing paradigms has been emerging, in which a user (first party) with an account on one computing platform (second party) can access the services provided by another computing platform (third party federated with the first computing platform). The user needs to authenticate itself with the third-party computing platform which does not have user credentials. In this work, we will propose and implement 3rd party authentication in federated Cloud, Edge, and Fog systems.</p>	2 to 3	Dr. Md. Shohrab Hossain (mshohrabhossain)
4 8	Computational Criminology	<p>Computational criminology presents different computational methods of criminology concepts, related complex phenomena of criminal activities, and generates solutions for the problems. It spans different areas such as forensic, criminal prediction, future crime forecasting, etc.</p> <p>In this thesis, our goal would be to devise new methodologies to facilitate criminological tasks such as identifying potential criminals/activities in/from a case study, predicting next step(s) of a criminal, identifying future trend of criminal activities, etc. Tasks to be done should cover real data collection, mixed method analysis, data mining, etc. Expected team size for this thesis will be 1-2.</p>	1	Dr. A. B. M. Alim Al Islam (razi)
4 9	Embedded Intelligence	<p>Artificial intelligence is now being in widespread usage, however, mostly at the cost of high resource utilization. There remain many applications (such as embedded system based applications) that cannot afford such high resource availability. Thus, research on enabling artificial intelligence sustaining resource constraints has recently emerged as a prominent research area.</p> <p>In this thesis, our goal would be to come up with effective mechanism(s) that can enable artificial intelligence in embedded devices having resource constraints. Attaining this goal would require development of new light-weight AI algorithm(s) and their device level experimentations. Expected team size for this thesis will be 1-2.</p>	1	Dr. A. B. M. Alim Al Islam (razi)

S L	Topic	Description	Stu den ts	Teacher
5 0	Quantum Intelligence	<p>Thinking classically is different from thinking quantumly, and so does in case of intelligence. In the realm of emerging quantum computing era, AI needs to be redesigned in many cases leveraging the core strengths of quantum computing sustaining its limitations. Thus, quantum intelligence would be substantially different from classical AI.</p> <p>In this thesis, our goal would be to come up with new quantum intelligence algorithms that can enable artificial intelligence in real quantum computers having special circuits and connections. Attaining this goal would require development of new quantum intelligence algorithm(s) and their experimentations in real quantum computers. Expected team size for this thesis will be 1-2.</p>	1	Dr. A. B. M. Alim Al Islam (razi)
5 1	Simulating Road Traffic of Dhaka for Effective Policy Making	<p>On-road experience in Dhaka city is perhaps one of the worst parts in lives of its dwellers. However, it is believed that policy making and their enforcement could substantially ameliorate the on-road experience. Attempting to do so directly in real settings may end up with backfiring consequences in case the enforced policy does not realize both micro-level and macro-level impacts, which are often very difficult to foresee. Here comes the role of simulation of to-be-attempted policies to understand their possible impacts even before their enforcements in real settings.</p> <p>In this thesis, our goal would be to come up with new effective on-road policies (such as making a road link unidirectional, removing slow vehicles from a link, etc.) that can engender better on-road experience. Attaining this goal would require development of new policies and their experimentations using a simulator named Dhakasim (developed in our research group). Expected team size for this thesis will be 1-2.</p>	1	Dr. A. B. M. Alim Al Islam (razi)
5 2	Computing Solutions to Serve the Under-served	<p>Digital divide has already segregated people between technology-literate and technology-illiterate sects. Such divide exhibits mourning consequences when a large part of whole population remains technology-illiterate as well as under-served from the perspectives of attaining the basic needs. This is even more prominent in the context of Bangladesh, where there remain many marginalized under-served communities (such as beggars, street children, blind people, etc.) that are mostly technology-illiterate. Thus, it is extremely challenging to devise technological solutions to serve these under-served communities.</p> <p>In this thesis, our goal would be to devise new technology-enabled solutions to facilitate serving some basic needs (food, shelter, education, etc.) of the marginalized communities. Attaining this goal would require development of new solutions following HCI-based studies such as VSD, participatory design, etc. Expected team size for this thesis will be 1-2.</p>	1	Dr. A. B. M. Alim Al Islam (razi)

S L	Topic	Description	Stu den ts	Teacher
53	Future Networking Technology: 5G and beyond 5G	<p>Different G's (2G, 3G, and 4G) have attained considerable successes paving the path of yet another "G", i.e., 5G, in near future, or it is already in. However, achieving success of the 5G, or beyond 5G, in specialized cases is still under research. An example of such cases is IoT, which demands a special treatment to 5G to get it rolling in its full potent in the system architecture of IoT. 5G security, 5G integration with SDN, etc., are some other related research areas.</p> <p>In this thesis, our goal would be to devise new protocol(s) or networking architecture to enable enhanced and secured performance of 5G or beyond 5G, perhaps in specific applications. Attaining this goal would require development of new protocol(s) or networking architecture and experimentation through simulation and/or testbed implementations. Expected team size for this thesis will be 1-2.</p>	1	Dr. A. B. M. Alim Al Islam (razi)
54	3D Modeling and Classification	<p>3D modeling and classification have diversified applications covering autonomous vehicles, animation, medical industry, and many more. These tasks demand various critical steps to be performed such as data acquisition, data preprocessing, model representation, object detection, etc.</p> <p>In this thesis, our goal would be to devise new methodologies to perform more accurate and efficient 3D modeling and classification. The methodologies can be automatic or semi-automatic – based on the application scenario(s) and deployment feasibility. Real experimentation is expected to be carried out here. Expected team size for this thesis will be 1-2.</p>	1	Dr. A. B. M. Alim Al Islam (razi)
55	Natural Language Understanding tasks for Bangla	<p>The recent transformer based models (such as BERT) have been proved to perform better than previous techniques for different downstream tasks such as fake news detection, hate speech detection, sentiment analysis, etc. for English and other languages. Very few works have been done to implement such models for Bangla. This project would explore different downstream tasks not implemented yet for Bangla utilizing a pre-trained Bangla BERT model recently developed at CSE, BUET.</p>	1 or 2	Dr. Anindya Iqbal (anindyaiqbal)
56	Automatic Code Improvement Suggestion based on Code Review	<p>Under tight deadline, the developers often fail to write robust code following recommended best practices. A recent and popular trend is to provide automatic code improvement suggestions applying different program analysis and/or machine learning techniques. Very recently, a couple of attempts have been made to improve such automatic suggestions utilizing associated code reviews one of which was done by our research group. In that work, a rich dataset of code change and associated review have been collected which will help this project by saving a lot of time? The target of this project would be to address the shortcomings of previous work and extend it to achieve better result.</p>	1 or 2	Dr. Anindya Iqbal (anindyaiqbal)
57	Automatic English Learning Helper for School Students	<p>Online-based automated teaching and assessment systems can help a lot for learning in places like the villages of Bangladeshi where qualified teachers are scarce. However, manually creating questions and answers for English content is cumbersome and expensive. In this project, we plan to create an AI-based automated tool for practice item generation for English learning school students in Bangladesh.</p>	1 or 2	Dr. Anindya Iqbal (anindyaiqbal)

S L	Topic	Description	Stu den ts	Teacher
58	Social media analytics	It will be more of an applied research project suitable for students who are interested in software development. State-of-the-art NLP techniques of Topic Modeling and Sentiment Analysis will be applied to classify posts in social media and news items for both Bangla and English. The insight gained will be usable for associating target news and post with a given content.	1 or 2	Dr. Anindya Iqbal (anindyaibal)
59	Natural Language Generation using GPT-2	In recent times, Transformer based language models have shown remarkable ability to generate human-like texts. Such models have been used to generate human-like dialogue, chat-bot, creative story generation, describing structured data, machine translation, and many others. With further fine-tuning, the same models have shown adeptness in text summarization, paraphrasing, and even poetry writing. They can also be used as writing helpers for writers. Although such down-stream tasks have been explored for English, none have been done for Bengali. This project will explore Bengali GPT-2 and its many down-stream tasks and applications.	1 or 2	Dr. Anindya Iqbal (anindyaibal)
60	Topics in low-resource NLP	We'll be tackling some fundamental problems in natural language processing (e.g., machine translation, text classification, text summarization, question answering) from a low-resource perspective. We shall explore different techniques like multilingual model training and transfer learning to overcome data shortage in low-resource setups, especially for languages like Bengali that are rather new to the web domain.	2	Dr. Rifat Shahriyar (rifat)
61	Scaling up inference speed for real-time model deployment	Though GPUs have helped accelerate training deep neural models, deploying them on GPU machines is often infeasible because of their high costs. In this work, we'll work with different techniques for scaling up real-time inference on CPUs. The topics will cover, but not be limited to, knowledge distillation, quantization, pruning, model caching, parallelization using multi-core processing.	2	Dr. Rifat Shahriyar (rifat)
62	Memory management for Python	We will study the existing garbage collection algorithm in Python to find its limitations. We will study different virtual machine runtimes for Python. We will work to design a high-performance garbage collector for Python.	1	Dr. Rifat Shahriyar (rifat)
63	Topics in empirical software engineering	We'll be tackling some problems in empirical software engineering. One possible topic is to study the developers' feedback on the different Java versions and their new features (Java 8 and onwards) by analyzing Stackoverflow and/or Github data. Another possible topic is to study the developers' feedback to the different popular Javascript frameworks by analyzing Stackoverflow and/or Github data.	2	Dr. Rifat Shahriyar (rifat)
64	Automated Filtering of Media Data Using Deep Learning	Now-a-days media data is being generated at a large scale in every corner of the world by every user. Management and analysis of this huge volume of data is creating concerns as manual endeavour for this task is prone to too many errors and takes a lot of time as well. Also, the lack of automated supervision is causing spread of harmful and obscene data around the electronic medium. The purpose of this research is to classify audio and detect anomaly for the classified audio. We'll also develop methods to identify keywords from audio conversations and identify the conversation topic in the process.	1 or 2	Dr. Muhammad Abdullah Adnan (adnan)

S L	Topic	Description	Stu den ts	Teacher
6 5	Machine Learning Centered Approach for Automated Rating of Applications	Recent regulations on data protection and privacy acts have led many popular applications to lose market share and face restrictions in operation in certain geographic areas. In this thesis, we intend to propose a method to rate applications based on their behaviors with user privacy and regional privacy regulations. We'll investigate the permissions an app asks for, check how relevant those permissions are given the type of the application, and finally provide a rating that best represents the app's behavior regarding user privacy. We'll use Machine Learning based models to determine the type of the application and the permissions each type of applications should ask for. We'll also propose unifying frameworks for data protection and privacy regulations and develop tools to guide developers during application development.	1 or 2	Dr. Muhammad Abdullah Adnan (adnan)
6 6	Geo-Distributed Data Analytics and Machine Learning	<p>The recent explosion of data volumes has reignited the focus on scale-out data analytics, and has fostered the world of Big Data systems. While these paradigms suffice for a single data center, we have reached a new inflection point where the combination of big and geographically distributed data requires new approaches for geo-distributed analytics processing and machine learning to minimize wide-area bandwidth costs. Centralized approaches together with heuristics such as data reduction or ad-hoc distributed querying may suffice in the short term. However, they are not sustainable as data volumes grow relative to transoceanic bandwidth and regulatory concerns become paramount.</p> <p>In this research project, we intend to develop an efficient training scheme for a machine learning algorithm that will work on Geo-distributed Big Data. Here, data are distributed among geographically separated data centers. Our goal is to train a distributed regression/support vector machine/neural network/etc. model that will use optimum communication (i.e. bandwidth) with the data centers as well as it will preserve the data privacy and converge quickly with a satisfactory accuracy.</p>	1 or 2	Dr. Muhammad Abdullah Adnan (adnan)
6 7	Security in Machine Learning Systems	With machine learning being integrated into real-life Systems, new avenues of attacks have opened up for attackers. Attackers have been working hard to figure out all the potential attacks an ML system could fall victim to. Hence, in order to secure a machine learning system, Security must be built into the model from the very beginning of the machine learning lifecycle. From the start of development through the production stages the entire time the model is in use, the system must be actively secured. In this research, we'll explore and analyze security aspects of ML systems such as Data confidentiality, System manipulation, Adversarial examples, Transfer learning attack, Data poisoning, etc. and innovate new techniques for securing machine learning systems.	1 or 2	Dr. Muhammad Abdullah Adnan (adnan)
6 8	Computationally Efficient Natural Language Processing	This research is focused on making NLP models more efficient both at training and testing time. This work will enable applications of such models at scale and potentially make them directly runnable on mobile devices without losing the quality of models run on servers. Model compression techniques such as quantization, knowledge distillation, model pruning, etc. and more efficient, modular, sparse architectures may need to be explored.	1 or 2	Dr. Muhammad Abdullah Adnan (adnan)

S L	Topic	Description	Stu den ts	Teacher
69	Neural Machine Translation for Low Resource Languages	Current Neural Machine translation models trained on little parallel data tend to produce poor quality translations. This challenge is exacerbated in the context of social media, where we need to enable communication for languages with no corresponding parallel corpora or unofficial languages such as romanized versions. In this research we will explore unsupervised, semi-supervised and multilingual neural machine translation for low resource languages.	1 or 2	Dr. Muhammad Abdullah Adnan (adnan)
70	A Deep Learning Approach for Gesturization of Bangla Language	The goal of this thesis is to research and develop techniques to generate gesture in real time from Bangla speech. We will formulate a deep learning based model on Bangla Sign Language and develop a tool that will translate verbal Bangla speeches into the corresponding gestures of Bangla Sign Language in real time. The outcome will be a portable application that will provide a bridge of communication between the hearing and hearing-impaired people.	1 or 2	Dr. Muhammad Abdullah Adnan (adnan)
71	Any Topic on Systems and Networking	Students can propose a topic and must be approved by me. Any number of students can work on any topic on different problems in groups or individually.	Any	Dr. Muhammad Abdullah Adnan (adnan)
72	Parkinson's Analysis with Remote Tasks	<p><i>This project is a collaboration with Dr. Ehsan Hoque, Associate Professor, University of Rochester.</i></p> <p>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. PARK instructs and guides users through six motor tasks and one audio task, and records their performance via webcam. Dr. Ehsan's team has published a sequence of papers using machine learning and deep learning, analyzing the dynamics of the facial expressions and the motion of hand gestures to automatically predict PD vs. non-PD. However, their model lacks data from Indian subcontinent.</p> <p>In this project, we will collect data from Bangladesh and train a model that is catered to the Bangladeshi population in predicting PD vs. non-PD. The work would require visiting several hospitals/clinics to interview Parkinson patients. Data will be collected using the Park platform. We will also collect data from elderly people without the disease to create a balanced dataset. We will then train a machine learning model on this dataset. We will also plug our data into the Park model to increase its effectiveness for a global population.</p> <p>Reference Material</p> <ol style="list-style-type: none"> https://parktest.net/index.html https://roc-hci.com/current-projects/park-parkinsons-analysis-with-remote-kinetic-tasks/ [See the reference papers] 	1 or 2	Dr. Mohammad Saifur Rahman (mrahman)

S L	Topic	Description	Stu den ts	Teacher
7 3	Eleven grand challenges in single-cell data science	<p><i>Research in these topics will be conducted in collaboration with Dr. Md. Abul Hassan Samee, Assistant Professor, Molecular Physiology and Biophys, Baylor College of Medicine</i></p> <p>Reference Paper: https://doi.org/10.1186/s13059-020-1926-6</p> <p>The recent boom in microfluidics and combinatorial indexing strategies, combined with low sequencing costs, has empowered single-cell sequencing technology. Thousands—or even millions—of cells analyzed in a single experiment amount to a data revolution in single-cell biology and pose unique data science problems. Single-cell RNA sequencing (scRNA-seq) enables transcriptome-wide gene expression measurement at single-cell resolution, allowing for cell type clusters to be distinguished, the arrangement of populations of cells according to novel hierarchies, and the identification of cells transitioning between states.</p> <p>In this paper (Linked above), the authors have outlined eleven challenges that will be central to bringing this emerging field of single-cell data science forward. For each challenge, the authors highlight motivating research questions, review prior work, and formulate open problems. We can work on one or more of these areas in collaboration with Dr. Samee.</p> <ol style="list-style-type: none"> 1. Challenge I: Handling sparsity in single-cell RNA sequencing 2. Challenge II: Defining flexible statistical frameworks for discovering complex differential patterns in gene expression 3. Challenge III: Mapping single cells to a reference atlas 4. Challenge IV: Generalizing trajectory inference 5. Challenge V: Finding patterns in spatially resolved measurements 6. Challenge VI: Dealing with errors and missing data in the identification of variation from single-cell DNA sequencing data 7. Challenge VII: Scaling phylogenetic models to many cells and many sites 8. Challenge VIII: Integrating multiple types of variation into phylogenetic models 9. Challenge IX: Inferring population genetic parameters of tumor heterogeneity by model integration 10. Challenge X: Integration of single-cell data across samples, experiments, and types of measurement 11. Challenge XI: Validating and benchmarking analysis tools for single-cell measurements 	1 to 2	Dr. Mohammad Saifur Rahman (mrahman)

S L	Topic	Description	Stu den ts	Teacher
7 4	Several topics in Bioinformatics	<p><i>Research in these topics will be conducted in collaboration with Dr. Md. Abul Hassan Samee, Assistant Professor, Molecular Physiology and Biophys, Baylor College of Medicine</i></p> <p>Study and design of algorithms to predict chromatin contact from DNA sequences: Genomic DNA folds systematically to ensure that specific pairs of loci remain in proximity (able to contact) in the three-dimensional space. Disruptions in these contacts have been associated with diseases. Datasets from chromatin conformation capture technologies, such as Hi-C, mark such pairs of genomic regions. Some questions we could ask: Can we predict the contact probability of two genomic regions? Which mutations in the human genome could potentially disrupt the contact between genomic regions?</p> <p>Study and design of algorithms to characterize silencer sequences: Silencer sequences are genomic regions where proteins bind and repress gene activity. Mutations in silencers could over-express genes and lead to diseases. Recent studies have started identifying and validating silencers in the human genome. Some questions we could ask: Can we identify silencers computationally? Are silencers evolutionarily conserved? What are the consequences of mutations in silencer sequences?</p> <p>Study and design of algorithms for spatial transcriptomics data: Spatial transcriptomics (ST) capture gene expression from intact human tissues at single-cell resolution. ST data reveal principles of tissue architecture, such as cellular colocalization patterns, and variations in gene expression across the tissue. Studying ST data from diseased samples could identify and characterize tissue regions for targeted therapy design. Some questions we could ask: Can we identify modules of genes that show systematic covariation of expression across the tissue? Can we segment an ST data into regions of high gene expression similarity?</p> <p>Study and design of algorithms for modeling high-throughput reporter assay data: A sequence's ability to act as an enhancer is defined by how efficiently it activates a gene (compare enhancers with silencers above). High throughput reporter assays (MPRA or CRISPRa/i-based assays) quantify thousands of sequences' enhancer activity in a single experiment. Some questions we could ask: Can we build regression models to understand why two different sequences would have different enhancer activities?</p> <p>Predicting the effect of methylation level changes in the human genome: Methylation is an epigenetic modification of genomic DNA that potentially alters gene expression levels. The exact effects of methylation and the mechanisms of how methylated regions function differently from unmethylated regions remain poorly understood. Some questions we could ask: Can we model the probability of a sequence becoming methylated? Can we model other epigenetic data (histone modifications) or transcriptomic data (gene expression) from methylation data?</p>	1 to 2	Dr. Mohammad Saifur Rahman (mrahman)

S L	Topic	Description	Stu den ts	Teacher
7 5	Actionable analytics of cancer	<p><i>This project is a collaboration with Dr. Abu Zafer Mohammed Dayem Ullah, UKRI/Rutherford Research Fellow, Health Data Research UK, Barts Cancer Institute,</i></p> <p>Objective:</p> <ol style="list-style-type: none"> 1. Health informatics approach to (any) cancer (or any disease of interest): Identifying association between cancer/disease incidence/mortality(survival) and clinical factors (demographics, medical history, lifestyle, etc.) 2. Integrative approach to (any) cancer: same as above but combining molecular/genomics data with clinical data 3. Risk score development for (any) cancer: machine learning-based approach to develop a scoring system for the probability of (any) cancer incidence/mortality in a specific period of time (5/10/20 year window) <p>Data source:</p> <ol style="list-style-type: none"> 1. TCGA data portal: https://portal.gdc.cancer.gov/ 2. ICGC data portal: https://dcc.icgc.org/ 3. NCBI Gene Expression Omnibus: https://www.ncbi.nlm.nih.gov/geo/ <p>These resources also come with a relevant simplified data analysis tool to do basic statistical analysis.</p> <p>Reference Papers:</p> <ol style="list-style-type: none"> 1. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5964983/ (relevant for the resource above) 2. Diabetes and cancer: Epidemiological and biological links https://www.wjgnet.com/1948-9358/full/v11/i6/227.htm 3. Deciphering the complex interplay between pancreatic cancer, diabetes mellitus subtypes and obesity/BMI through causal inference and mediation analyses https://gut.bmj.com/content/70/2/319 4. New-onset diabetes and pancreatic cancer https://www.eghjournal.org/article/S1542-3565(06)00676-8/fulltext <p>Reading the Method(s) section of the papers would be useful.</p>	1 to 2	Dr. Mohammad Saifur Rahman (mrahman)

S L	Topic	Description	Stu den ts	Teacher
7 6	Computational analysis of discussions related to the religion of Islam	<p>In this project we will computationally analyze audio/video/text data related to the religion of Islam. The data could be forum posts, audio or video lectures about Islamic topics, debates etc. We can apply topic modeling and/or other analysis approaches.</p> <p>Possible data sources</p> <ol style="list-style-type: none"> 1. Comments on Youtube videos 2. Facebook or Twitter posts 3. www.islamqa.info 4. Audio/Video lectures of different speakers <p>Questions we could ask</p> <ol style="list-style-type: none"> 1. Topic Distributions: What religious topics do people discuss about in popular Q&A sites (e.g. www.islamqa.info) 2. Question Types: What type of questions are asked by Muslims (or non-Muslims) in online Q&A forums? 3. Topic Difficulty: What Islamic religious topics are more or less difficult in online Q&A forums? 4. Topic Evolution: How do Islam-related discussion topics evolve in online forums? 5. Topic Complementarity: How do Islamic discussion topics distribute across different Q&A sites? 6. Using topic modeling analysis, can we identify the bias of a platform for or against Islam? 7. Using computational analysis, can we provide a summary description of different topics discussed in an Islamic lecture or lecture series? <p>Challenges</p> <ol style="list-style-type: none"> 1. In the English lectures or write-ups in Islamic context, some Arabic words, expressions may get introduced. Therefore a conventional language model may fail to perform well and retraining a model may be needed. 	1 to 2	Dr. Mohammad Saifur Rahman (mrahman)
7 7	Extraction of Information from low resolution scan of Bangladeshi Passports	<p>In this project, we will process low resolution scan images (less than 60 Kilobytes in size) of Bangladeshi Machine Readable Passport. We want to extract all useful information (Name, DOB, Gender, Issue date, Expiry date etc.) from the image. We have to achieve very high accuracy so that our model can be incorporated in various portals where applicants/users have to upload their passport scans. In a lot of these cases, the scans are manually verified against the information put in by user, which takes time. If we can provide a highly accurate model then much faster user registration process (with verification) can be incorporated in different portals. We will work with more than 1 lac passport scans to train necessary models.</p>	1	Dr. Mohammad Saifur Rahman (mrahman)

S L	Topic	Description	Stu den ts	Teacher
7 8	Efficient algorithms for reconstructing phylogenetic trees (evolutionary trees)	<p>The overarching goal of this research project is to answer 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.</p> <p>In this project, 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 this particular project, our research will contribute to the problem of fast and accurate species tree estimation from genes sampled throughout the whole genome, considering the presence of gene tree discordance.</p> <p>Prior knowledge in Biology is not required.</p>	1 ~6	Dr. Md. Shamsuzzoha Bayzid (bayzid)
7 9	Application of machine learning in predicting protein attributes	<p>Proteins are considered as 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 advancement 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 and protein foldings) from protein sequences.</p> <p>Please see the following articles:</p> <p>i) https://academic.oup.com/bioinformatics/article/36/17/4599/5841663?guestAccessKey=abd42e90-e1c1-429c-a073-a781893b765b</p> <p>ii) https://www.biorxiv.org/content/10.1101/2020.11.07.372466v3.full</p> <p>Prior knowledge in Biology is not required.</p>	1~6	Dr. Md. Shamsuzzoha Bayzid (bayzid)

S L	Topic	Description	Stu den ts	Teacher
8 0	Identifying Genomic Islands	<p>In this project, we envision to develop novel and highly accurate machine learning (ML) and statistical anomaly detection-based methods for identifying genomic islands (GIs) and horizontally transferred genes.</p> <p>Horizontal gene transfer (HGT), also known as lateral gene transfer (LGT), refers to the exchange of genetic material among organisms in a way other than the vertical transmission of genes from parents to offspring through reproduction. Genomic Islands (GIs) are clusters of genes that are mobilized through horizontal gene transfer. Such events are believed to be responsible for diverse features of bacteria including their resistance to antibiotics, ability to degrade artificially synthesized substances, flourish in novel environments, or to pulverize the threats of decimation or even extinction when faced with natural changes or targeted human efforts. Due to the profound impact of GIs on prokaryotic genome evolution, tremendous efforts have been made for decades to develop accurate methods for their detection. However, the current state-of-the-art in GI detection leaves much to be desired. The existing algorithms suffer severely from an imbalance in precision-recall trade-off, with precise methods tending to miss many GIs, whereas sensitive methods being prone to generating unacceptably high numbers of false positives. Moreover, several of them are supervised in nature and thus their applications to newly sequenced genomes are riddled with their dependency on the functional annotation of existing genomes. In this project, we will try to advance the state-of-the-art in GI prediction by developing highly accurate GI prediction methods.</p> <p>Prior knowledge in Biology is not required.</p>	1 ~6	Dr. Md. Shamsuzzoha Bayzid (bayzid)
8 1	Leveraging biomolecular and clinical data for developing better preventive care	<p>We envision developing effective computational, statistical, and machine learning based approaches for understanding and processing the ever-increasing amount of biomolecular and clinical data (electronic health records) available these days. The availability of this huge amount of data has opened the opportunity for investigating the genetic basis of various diseases and to develop better treatment and preventative measures accordingly. However, this opportunity is riddled with the fact that existing algorithms and “non-automatic” techniques involving manual intervention are prohibitively expensive for handling this Big-data. This emphasizes the need for developing appropriate computational, statistical, and machine learning techniques to build clinically translatable models to jointly analyze both clinical and genetic data in order to learn and predict various latent trends and genetic associations. Another important direction of this project would be building novel 3D deep learning models for medical image analysis.</p> <p>This project will be done in collaboration with:</p> <ul style="list-style-type: none"> i) Dr. Rezwana Reaz (CSE, BUET) ii) Icahn School of Medicine at Mount Sinai, New York, USA iii) Zahid Rahman, AWS at Amazon, Palo Alto, California, USA. 	1 ~6	Dr. Md. Shamsuzzoha Bayzid (bayzid)

S L	Topic	Description	Stu den ts	Teacher
8 2	Fake news/ Hate speech detection in social platform	<p>Online fake news , abusive or hateful speech detection in different languages on Social Networking Sites (SNS) have drawn the attention of researchers recently. Fake news or Hateful comments in public Facebook pages ignite social mishaps in Bangladesh as well. Our goal is to detect and classify fake news, hate speech in bengali language across the social network using data from CrowdTangle interface and API. (CrowdTangle- A tool from Facebook to help follow, analyze, and report on what's happening across social media).</p> <p>Researc Area: HCI, Machine learning, Network graph.</p>	2 to 3	Dr. Sadia Sharmin (sadia)
8 3	Graph Neural Network	<p>Graph neural networks (GNNs) are deep learning based methods that operate on graph domain. Recently, Graph Neural Network (GNN) has gained increasing popularity in various domains, including social network, knowledge graph, recommender system, and even life science. Researches of analyzing graphs with machine learning have been receiving more and more attention because of the great expressive power of graphs. The power of GNN in modeling the dependencies between nodes via message passing between the nodes of graphs enables the breakthrough in the research area related to graph analysis. Essentially, every node in the graph is associated with a label, and we want to predict the label of the nodes without ground-truth.</p> <p>GNNs have numerous application in social network analysis as social network have been developing based on the phenomenon that users usually acquire and disseminate information through those around them, such as classmates, friends, or colleagues, implying that the underlying social relations of users can play a significant role in helping them filter information. In terms of real-life application we have to design algorithms in such a way that can handle billions of nodes.</p> <p>Actual topic will be assigned based on discussion.</p>	2 to 3	Dr. Sadia Sharmin (sadia)
8 4	Attention Engineering for local applications	<p>Facebook engineers your attention. Facebook algorithms choose and schedule content and ads optimally. Based on your media use, cookies, browser history, and myriad other factors, they present you with the most engaging content at the best possible time. The addictive cycles that Facebook, YouTube, Snapchat, Reddit, Netflix, and others proliferate alter your perceptions and change your mental effectiveness. They cause people to easily consume hyperbolic journalism, biased information, misinformation, and disinformation. Individuals receive content that confirms their beliefs and ignores or chastises conflicting beliefs. This confirmation bias can lead to lapses in judgement and vapid, linear thought. Attention engineering strategies, ironically, contribute to scattered attention, less productive work, less critical thinking, and less self-control.</p> <p>We would like to investgate what is the case with new/ our local applications.</p>	2 to 3	Dr. Sadia Sharmin (sadia)

S L	Topic	Description	Stu den ts	Teacher
8 5	Understanding Users, Cognitive and Affective Factors	<p>Cognition and affection are important for interaction design. Cognition and affection has been applied to interaction design and cognitive research has led to the design of more effective interactive products, mental models, conceptual frameworks that are useful for interaction design.</p> <p>Research area: HCI, AI.</p> <p>Actual topic will decided based on discussion.</p>	2 to 3	Dr. Sadia Sharmin (sadia)
8 6	Alignment free phylogeny construction using maximum likelihood using k-mer counts	<p>A phylogeny is a tree that depicts the evolutionary relationships among various biological species. Traditional approaches for phylogeny constructions from genomic data require prior alignment of the sequences. Recently, alignment free algorithms have been developed but they are based on the distance based approach. Here, we will design and implement an alignment-free algorithm for phylogeny construction using maximum likelihood building on work by students of previous batches. This will be based on a birth-death-migration model for k-mer counts.</p> <p>No prior knowledge of biology is necessary but will be helpful</p>	2	Dr. Atif Hasan Rahman (atif)
8 7	Study and design of algorithms to analyze scRNA-seq data	<p>All cells in an organism share almost the same DNA sequence and the same set of genes. However, different cells look and behave differently as the sets of genes expressed vary across cell types. Single cell RNA-seq (scRNA-seq) data is used to study this. Here, we will study various challenges in this area and develop algorithms to address those challenges. Possible topics include</p> <ul style="list-style-type: none"> • Imputation of missing data using RNA velocity • RNA velocity estimation in isoform resolution <p>This will be a collaboration with Md. Abul Hassan Samee, Ph.D., Assistant Professor, Baylor College of Medicine</p> <p>No prior knowledge of biology is necessary but will be helpful</p>	1 or 2	Dr. Atif Hasan Rahman (atif)

S L	Topic	Description	Stu den ts	Teacher
8 8	Study and design of algorithms to analyze cancer genomic data	<p>Cancer is caused by a sequence of mutations (changes in DNA). Here, we will analyze cancer genome data (WGS) to find mutual dependencies</p> <ul style="list-style-type: none"> • between noncoding mutations • between coding and noncoding mutations <p>Mutual dependency could mean co-occurrence or mutual exclusion. Prior works have noted mutual exclusion between coding mutations. With an increased appreciation for the role of disrupted cis-regulation in cancer, we want to check for similar relationships involving noncoding mutations. We also want to test if these dependency patterns correlate with cancer phenotypes such as metastasis, survival, drug resistance, etc.</p> <p>This will be a collaboration with Md. Abul Hassan Samee, Ph.D., Assistant Professor, Baylor College of Medicine</p> <p>No prior knowledge of biology is necessary but will be helpful</p>	1 or 2	Dr. Atif Hasan Rahman (atif)
8 9	Topics in Computational Biology	<p>We will devise and implement a method for one of the following problems in computational biology:</p> <ul style="list-style-type: none"> • Genome assembly using third generation sequencing: Genome assembly is the process of merging 'reads' generated by sequencing technologies to construct the original genome. Third generation sequencing technologies including Pacific Biosciences and Oxford Nanopore can generate longer reads compared to earlier technologies at the expense of higher sequencing error rates. We will devise and implement methods to use third generation sequencing reads for scaffolding and other problems related to genome assembly. • Association mapping using k-mers: Association mapping is the process of finding which variants in genome are associated with diseases and traits. We will develop a method for finding relations between to traits and variations in genomes in a reference free manner. <p>No prior knowledge of biology is necessary but will be helpful</p>	1 or 2	Dr. Atif Hasan Rahman (atif)

S L	Topic	Description	Stu den ts	Teacher
90	Study of different firewall representations in the elasticity control of virtual network firewalls	<p>There are several approaches to represent network firewalls. The most common approach is to represent it as a sequence of rules. Firewall decision diagrams (FDD), binary decision diagrams (BDD), tree-rule firewalls are among other prominent approaches of representing network firewalls. Although the sequence of rule representation is the simplest way to represent a firewall, it might contribute to higher execution time compared to the other representations for large firewalls.</p> <p>Network function virtualization (NFV) has enabled network firewalls to be implemented on virtual machines. One of the main goals of designing virtual network firewalls is to achieve elasticity, i.e., the number of virtual instances can grow and shrink based on traffic load. The firewall execution time plays a crucial role to handle the traffic load effectively. The more frequently the virtual instances are overloaded, the more frequently new instances are created. In this project, we will adopt different approaches to build virtual firewalls and study their impact on elasticity control.</p>	1~3	Dr. RezwanaReaz (rezwana)
91	Application of blockchain in in fault tolerant state management for virtual network functions	<p>Network Functions can be stateless or stateful. Network function virtualization (NFV) has enabled network functions to be implemented on virtual machines. One of the main goals of designing virtual network functions (VNF) is to achieve elasticity, i.e., the number of virtual instances can grow and shrink based on traffic load.</p> <p>While it is easy to scale stateless network functions across multiple virtual NF instances, it is complicated when states are involved. Thus, one of the core challenges in achieving elasticity is to determine how NF states are stored and shared across multiple NF instances. State management can be even more complicated when NF instances can fail. In this project, we will explore the application of blockchain technology in NF state management under the presence of NF failure.</p> <p>Prior knowledge in blockchain is preferred.</p>	2~3	Dr. RezwanaReaz (rezwana)
92	Leveraging fog-cloud computing for implementing IoT firewalls	<p>IoT devices are vulnerable to many attacks. So it is important to implement IoT firewalls to filter out unwanted traffic before they can reach IoT devices. One way to implement an IoT firewall is to store and run a firewall in each IoT device customised for that device. However, most IoT devices come with very limited memory and computational capability. So storing and running a firewall within an IoT device is not a practical solution. Other options include implementing the firewall in the network hub, in a gateway router, or in a distant node like fog/ cloud.</p> <p>In this project, we will explore the fog-cloud based architecture to implement firewalls for IoT devices in a smart-healthcare scenario. Such an approach involves traffic redirection to a cloud or fog nodes and execution of the firewall in the cloud/fog nodes. While such an approach is beneficial in terms of scalability, economy, memory utilization, and computational capacity, it poses several challenges in terms of latency, stateful inspection of firewalls, and privacy. In this project, we will address these challenges.</p>	2~3	Dr. RezwanaReaz(rezwana)

S L	Topic	Description	Stu den ts	Teacher
9 3	Leveraging biomolecular and clinical data for developing better preventive care	<p>We envision to develop effective machine learning and statistical approaches for understanding and processing the ever increasing amount of biomolecular and clinical data (electronic health records) available these days. The availability of this huge amount of data has opened the opportunity for investigating the genetic basis of various diseases and to develop better treatment and preventative measures accordingly. However, this opportunity is riddled with the fact that existing algorithms and “non-automatic” techniques involving manual intervention are prohibitively expensive for handling this Big-data. This emphasizes the need for leveraging appropriate machine learning techniques to build clinically translatable predictive models to jointly analyze both clinical and genetic data in order to learn and predict various latent trends and genetic associations. Another important direction of this project would be building novel 3D deep learning models for medical image analysis.</p> <p>This project will be done in collaboration with:</p> <ul style="list-style-type: none"> i) Dr. Md. Shamsuzzoha Bayzid (CSE, BUET) ii) Icahn School of Medicine at Mount Sinai, New York, USA ii) Zahid Rahman, AWS at Amazon, Palo Alto, California, USA. 	2~3	Dr. Rezwana Reaz (rezwana)