

Title	Description	Student count	Supervisor
<p>Research on Knowledge Graph (Topic 1)</p>	<p>In the context of Semantic Web, a knowledge graph is a way of representing knowledge. In short, you start from a few triples and those triples are put in relationship to build a graph.</p> <p>Knowledge Graph organizes data for presentation around what it calls entities, which include individuals, places, organizations, sports teams, works of art, movies and so on. Information comes from both freely available and licensed sources including the CIA World Factbook, Freebase and Wikipedia.</p> <p>The concept of the knowledge graph was popularized by Google, which launched the Google Knowledge Graph back in 2012 in a bid to provide users with structured answers to their queries “ not just links.</p>	<p>3</p>	<p>Dr. Muhammad Masroor Ali (mmasroorali)</p>
<p>Research on Knowledge Graph (Topic 2)</p>	<p>In the context of Semantic Web, a knowledge graph is a way of representing knowledge. In short, you start from a few triples and those triples are put in relationship to build a graph.</p> <p>Knowledge Graph organizes data for presentation around what it calls entities, which include individuals, places, organizations, sports teams, works of art, movies and so on. Information comes from both freely available and licensed sources including the CIA World Factbook, Freebase and Wikipedia.</p> <p>The concept of the knowledge graph was popularized by Google, which launched the Google Knowledge Graph back in 2012 in a bid to provide users with structured answers to their queries “ not just links.</p>	<p>3</p>	<p>Dr. Muhammad Masroor Ali (mmasroorali)</p>

<p>Dealing with Big Data</p>	<p>There are several prospective research areas:</p> <p>Algorithmic Challenges: In many cases efficient deterministic algorithms are not feasible for big data. In those cases we need to design suitable distributed, approximation and randomized algorithms.</p> <p>Data Clustering: Data clustering is an effective technique for dealing with big data which has practical applications in facility assignment, privacy etc.</p> <p>Visualizing big data: We need to design suitable graph drawing algorithms for visualizing big data.</p> <p>Complex Network Analysis: Real world problems can be modeled by complex networks. By analyzing complex networks we can deal with disaster management, share market prediction prediction, marketing campaign for new products etc.</p> <p>Online Computation/Algorithms: Dealing with data stream where we get input in online fashion.</p> <p>Beyond Planar Graphs: Graph models dealing with big data are not necessarily planar. Algorithms for non-planar graphs has applications in big data analysis.</p> <p>Application Domains: Computer Networks, Bioinformatics, Computer Security, VLSI Layout etc.</p> <p>Thesis Assignment Policy: If you are interested to deal with any challenging problem mentioned above, you can</p>	<p>6</p>	<p>Dr. Md. Saidur Rahman (saidurrahman)</p>
------------------------------	---	----------	---

<p>Design of a Crypto Currency Platform</p>	<p>Cryptocurrency exchange platform is marketplace which matches buyers and sellers of various currencies like Bitcoin (BTC), Ethereum (ETH), Dash (DASH), Litecoin (LTC), Ripple (XRP) etc. Exchange will allow buyers and seller to specify a set price for their order, also process orders (instantly or limit the orders) at the market price (updated near real time).</p> <p>The following components can be considered for the marketplace.</p> <p>First of all trade engine. The trade engine should be the core of any exchange. It accesses the order book, matches buy/sell orders, executes transactions and calculates balances, currency exchange proxy APIs.</p> <p>Secondly, the user interface (UI) or front end of the exchange website; should enable the user to register and access an account, deposit, maintain and withdraw Crypto currencies, view current order book, past transactions, balances, statistics etc, view charts, place buy and sell orders.</p> <p>The third component is the wallet. This is where the currencies of the exchange operator and his traders are stored. Different wallet is used for different currency.</p> <p>The fourth component any exchange should possess is an admin panel. The functions like Altering liquidity, Editing the spread and trading fees, Approving user accounts for trading, after verifying Know Your Customer (KYC) or other compliance requirements, Managing currencies and markets like USD/BTC, BTC/ETH, etc., Crediting fiat deposits, or debiting fiat withdrawals requested by users help the exchange operator control and manage his exchange.</p> <p>Finally, the platform security. Personal data security should be placed in the first</p>	<p>3</p>	<p>Dr. Md. Mostofa Akbar (mostofa)</p>
<p>Finding security insights from Android device logs</p>	<p>It is a new topic. Currently, the plan is to collect operating system logs from multiple devices and use data mining/machine learning techniques to find out security insights automatically. The logs are in a text format and the knowledge of python should suffice.</p>	<p>1</p>	<p>Dr. Md. Mostofa Akbar (mostofa)</p>

<p>Topics in Bioinformatics</p>	<p>We want to work on different topics in BI ranging from a Prediction System based on Machine Learning Techniques to De-Bruijn Graph Based Genome Assembly. We also can work on algorithmic problems in BI. The actual problems will be finalized after discussing with the student.</p> <p>These topics would require learning tools and conducting extensive experiments.</p> <p>ANY NUMBER OF STUDENTS ARE WELCOME.</p>	<p>6</p>	<p>Dr. M. Sohel Rahman (msrahman)</p>
<p>Topics in Stringology</p>	<p>We can work on devising efficient algorithms for strings or make a journey in the realm of String Combinatorics. These works will mostly be theoretical but there could be options for conducting experiments using benchmark data.</p>	<p>2</p>	<p>Dr. M. Sohel Rahman (msrahman)</p>
<p>Solving Hard Problems using Metaheuristics</p>	<p>Most of the real life problems are actually hard and hence optimal algorithms are not that useful due to high computational complexity. We want to apply metaheuristic techniques to solve these problems. The problems may come from different domains including but not limited to Transportation, Computational Biology, Groundwater, Physics, other Engineering branches etc.</p> <p>ANY NUMBER OF STUDENTS ARE WELCOME.</p>	<p>6</p>	<p>Dr. M. Sohel Rahman (msrahman)</p>
<p>On detecting malicious nodes triggering DoS attack in Wireless Multi-hop Networks</p>	<p>Denial of Service attack (DoS) is a serious problem in both wired and wireless networks. In wireless multi hop networks, a DoS attack in the form of virtual jamming can be easily triggered by malicious nodes due to limitation of the four way handshake (RTS-CTS-DATA-ACK) mechanism used in the MAC layer. In this research we will address the following research questions:</p> <p>(a) How malicious nodes can be detected either statistically or using some machine learning algorithms.</p> <p>(b) How malicious nodes (once identified) can be isolated in order to mitigate DoS attack.</p> <p>Some knowledge of ns2- simulator will be an asset.</p>	<p>3</p>	<p>Dr. A.K.M. Ashikur Rahman (ashikurrahman)</p>

<p>Performance analysis of different Connected Dominating Set (CDS) algorithms for broadcasting in wireless multi hop networks</p>	<p>Broadcasting is a fundamental research problem in any networks. Due to shared channels, broadcasting creates serious redundancy, contentions and collisions in wireless multi hop networks which is known as BROADCAST STORM PROBLEM. Several Connected Dominating Sets have been proposed to mitigated broadcast storm problem. In this research work we will analyze the performance of several CDS algorithms using ns2 simulator. Some knowledge of ns2- simulator will be an asset.</p>	<p>3</p>	<p>Dr. A.K.M. Ashikur Rahman (ashikurrahman)</p>
<p>An AI based Medical Chatbot for Adolescence Health Support in Bangla Language</p>	<p>Researchers found that in Bangladesh have very limited or no access to sexual and reproductive health-related information and services, and face serious barriers to getting information and guidance regarding those issues. Several studies also indicated that social taboo and shyness discourage adolescents and their families from seeking treatment from the adolescent-friendly health corners at government facilities. The research also finds that shortages of medicine and lack of privacy for adolescents are one of the major challenges that should be resolved to make those centers functional. To overcome all the above limitations, in this project we aim to build an AI driven medical chatbot in Bangla language, where adolescents can interact with the chatbot and the chatbot can provide suggestions on different health-related issues.</p>	<p>2</p>	<p>Dr. Mohammed Eunos Ali (eunos)</p>
<p>A Deep-learning based Disease Prediction of Rice Corp</p>	<p>Rice is the major staple crop in our country, occupying about 70 percent of the grossed crop area and accounting for 93 percent of total cereal production. However, different disease outbreak in these corps can have devastating impact on farmers as well on the country's economy. Thus, timely prediction of disease outbreak of different types of rice disease is of major importance. In this project, we aim to build a deep learning based model that use different environmental factors such as weather, and soil, and images of of the corps to predict the outbreak of different types of diseases. The outcome of this model will be deployed in field level to test the efficacy of the proposed research.</p>	<p>2</p>	<p>Dr. Mohammed Eunos Ali (eunos)</p>
<p>Learning to Index Big Spatial Data</p>	<p>So far researchers in database and IR domains have heavily invested on devising built based on the available data, the search performance of the index varies across different distributions of queries. In this project, we will apply machine learning to learn different parameters of the index so that the index can be adapted to different query patterns and give desired performance outcome irrespective of the query patterns.</p>	<p>2</p>	<p>Dr. Mohammed Eunos Ali (eunos)</p>

Intelligent Caching on Path Queries	Caching has been used in improving query performance in almost all the fields of computer science. Most recently, path caching has been introduced to instantly answer fastest path queries on Google map or open street map. In this project, we will devise an efficient caching technique based on learning of the user movement in the space. We will exploit machine learning and spatial algorithms to design an effective strategy for path caching that can answer path queries instantly.	2	Dr. Mohammed Eunos Ali (eunos)
Blockchain and Attacks	Blockchain is a new technology which will play an important role for secure decentralization in such emerging fields as IOT, Cyber Physical Systems, Crowdsensing and next generation wireless communications. We will study different attacks for blockchain in the application areas like IOT, CPS, or etc.	2	Dr. Mahmuda Naznin(mahmudanaznin)
Image Segmentation, Noise Elimination and 3D image from 2D image	This is a collaborative work on Biomedical Imaging, a continuation work with Dr. Md. Tanvir R. Faisal, Postdoctoral Fellow on Northwestern University, USA where image processing like segmentation, image structure analysis will be done from dicom images. MATLAB knowledge is preferred, but not required.	2	Dr. Mahmuda Naznin(mahmudanaznin)
Content Based Networking	Content Based Networking or Name Based Networking is a new area where the IP is replaced with the name and the data content. There are many research challenges with secured content caching, name based data forwarding, IOT data handling, node storage management, trust management in NDN. ndnSim (NS3 based) has to be used for experiment. Senior research groups may help for learning simulator. (more groups can work in this area, but individual group will have at most 2 students)	2	Dr. Mahmuda Naznin(mahmudanaznin)
Document Image Binarization	Document image binarization is an essential step in any document analysis including OCR, preservation of historical and education documents, etc. Though there are a number of techniques available, each method targets to binarize a particular category of documents. This thesis will aim to develop a generalized solution for binarization of most of the categories of documents.	2	Dr. Md. Monirul Islam (mmislam)
Texture classification using local patterns		2	Dr. Md. Monirul Islam (mmislam)
supervised medical image segmentation/classification using deep neural network		2	Dr. Md. Monirul Islam (mmislam)

Blockchain vs. Hashgraph	<p>This work will be in collaboration with Blockchain at Berkley (https://blockchain.berkeley.edu).</p> <p>Blockchain is a decentralized, distributed and public digital ledger that is used to record transactions across many computers so that the record cannot be altered retroactively without the alteration of all subsequent blocks and the collusion of the network. Hence, Blockchains are secure by design. Decentralized consensus is achieved with a blockchain, thereby making blockchains suitable for the recording of medical records, identity management, transaction processing, documenting provenance, voting.</p> <p>However, this technology is quite slow and it has other issues. Hashgraph on the other hand can improve the transaction rate significantly compared to BLoackchain. Students will work on these two new technologies, explore issues and possible solution techniques.</p>	2	Dr. Md. Shohrab Hossain (mshohrabhossain)
WebRTC: security risks and challenges	<p>Web Real-Time Communication (WebRTC) is a recent web application technology, that enables real-time communication in the browser without the need for plug-ins. Applications of WebRTC include web conferencing, Real-time communication apps. However, there are a number of security threats of WebRTC based communication. This thesis work will focus on such security threats. Students will have to test and implement WebRTC-based communication software and analyze its various aspect with respect to security and privacy.</p>	2	Dr. Md. Shohrab Hossain (mshohrabhossain)
Mobile Malware analysis and detection	<p>This is a ongoing research work. Students will start from our previous two works (1. Behavioral Malware Detection Approaches for Android (IEEE ICC 2016) 2. another paper under review in IEEE GlobeCom 2018). Android development knowledge is a mandatory prerequisite of this topic.</p>	2	Dr. Md. Shohrab Hossain (mshohrabhossain)
Devising new road traffic policies for Dhaka through simulation	<p>Dhaka is known for its near-to-everyday traffic jam experiences. This could be lessen by effective policies on road networks. This thesis is intended for devising such effective policies. To do so, the road networks of Dhaka need to be simulated by our custom simulator with different possible alternatives.</p>	1	Dr. A. B. M. Alim Al Islam (razi)

Determining deployment topology for on-rail sensor networks in a railway system	Detecting uprooted rail blocks ahead from an approaching train has been investigated in recent times. For enabling a solution of this problem, it is important to determine effective deployment strategy of sensor nodes on a railway. This thesis aims at determining the topology based on our custom sensing modules. The determined topology needs to be assessed by real implementation and simulation.	1	Dr. A. B. M. Alim Al Islam (razi)
Security aspects in nanonetworks	Nanonetworks offer perhaps the tiniest possible objects, known so far, to work with for enabling networked communication. The communication is itself still at a rudimentary stage. Once enabled, its next challenge to be assessing its security aspects. This thesis aims for investigating the security aspects. Simulations are expected to be conducted in this study.	1	Dr. A. B. M. Alim Al Islam (razi)
Inducing signals in a brain	Brain is known for generating different signals to our body parts. On the other way, can we induce signals to a brain? This thesis is expected to investigate this aspect. To do so, real experimental setup will be needed to be built that can induce signals on live (or semi-live, just after death) brains. The brains can be harvested from animals such as from a goat.	2	Dr. A. B. M. Alim Al Islam (razi)
Creating emotions in machines	Detecting emotion from a machine is a well-researched topic in recent times. Its next step would be creating emotion from a machine. This thesis is intended for investigating this aspect. Here, real implementation of new machine learning technique will be needed in addition to user evaluation.	1	Dr. A. B. M. Alim Al Islam (razi)
Quantum computing in psychic phenomena	It is often believed that brain activities could be amenable to quantum computing, or vice versa. However, it is still less focused whether quantum computing could help in psychic phenomena. This thesis aims at investigating in this aspect. Building theoretical foundations along with simulations are expected to be conducted in this study.	1	Dr. A. B. M. Alim Al Islam (razi)

<p>Geo-Distributed Data Analytics and Machine Learning</p>	<p>The recent explosion of data volumes has reignited the focus on scale-out data analytics, and has fostered the world of Big Data systems. While these paradigms suffice for a single data center, we have reached a new inflection point where the combination of big and geographically distributed data requires new approaches for geo-distributed analytics processing and machine learning to minimize wide-area bandwidth costs. Centralized approaches together with heuristics such as data reduction or ad-hoc distributed querying may suffice in the short term. However, they are not sustainable as data volumes grow relative to transoceanic bandwidth and regulatory concerns become paramount.</p> <p>In this research project, we intend to develop an efficient training scheme for a machine learning algorithm that will work on Geo-distributed Big Data. Here, data are distributed among geographically separated data centers. Our goal is to train a distributed regression/support vector machine/neural network/etc. model that will use optimum communication (i.e. bandwidth) with the data centers as well as it will preserve the data privacy and converge quickly with a satisfactory accuracy.</p>	<p>2</p>	<p>Dr. Muhammad Abdullah Adnan (adnan)</p>
<p>A Deep Learning Approach for Gesturization of Bangla Language</p>	<p>The goal of this thesis is to research and develop an application that can 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.</p>	<p>2</p>	<p>Dr. Muhammad Abdullah Adnan (adnan)</p>

Online Computation Offloading Algorithm	Due to energy and resource constraints, mobile devices often offload tasks to nearby servers. Such offloading decisions must be taken in real time and in an online manner. In this thesis, we plan to investigate the real-time scheduling solutions for mobile cloud services that combine both mobile code offloading and mobile task delegation in which the dependency of the mobile tasks within a mobile application as well as the application deadlines will be considered when scheduling. Moreover, as energy constraint is vital for mobile devices, a joint optimization of makespan and energy consumption for mobile cloud offloading will also be studied.	2	Dr. Muhammad Abdullah Adnan (adnan)
AI Enabled IoT	In today's world everything is connected yet the connected IoT devices need human intervention to perform their designated tasks. Even the IoT devices sometimes become vulnerable as long as humans are in the loop. The goal of this research project is to make IoT devices intelligent so that human interaction with the IoT device is minimized. An IoT device will be able to learn itself, teach other IoT devices, co-ordinate with other IoT devices and take decision on what actions to perform all by itself. The outcome will be enhancement of the capabilities of IoT devices using AI and Cloud Computing.	2	Dr. Muhammad Abdullah Adnan (adnan)
AI Enabled IoT	In today's world everything is connected yet the connected IoT devices need human intervention to perform their designated tasks. Even the IoT devices sometimes become vulnerable as long as humans are in the loop. The goal of this research project is to make IoT devices intelligent so that human interaction with the IoT device is minimized. An IoT device will be able to learn itself, teach other IoT devices, co-ordinate with other IoT devices and take decision on what actions to perform all by itself. The outcome will be enhancement of the capabilities of IoT devices using AI and Cloud Computing.	2	Dr. Muhammad Abdullah Adnan (adnan)

<p>Energy Profiling of Different Programming Languages</p>	<p>This work will focus on power and energy profiling of different programming languages like Java, Python, JavaScript, etc. There are different methods to measure power and energy consumption with different level of accuracy. This work will start with a survey of different profiling already there for different languages and will continue exploring new applications and opportunities.
</p> <p>Related paper: http://webdocs.cs.ualberta.ca/~hindle1/2016/hindle-icse2016.pdf</p>	<p>2</p>	<p>Dr. Rifat Shahriyar (rifat)</p>
<p>Development of Applications for Health Data Acquisition</p>	<p>In this project, we will explore effects of environment on common health-related issues through participatory sensing. We will develop an application to acquire weather and other environmental data such as noise and information on health-related events such as flu, headache and analyze the data for contributing factors, detect seasonal patterns, etc.
</p> <p>* Collaboration with Dr. Atif Hasan Rahman</p>	<p>1</p>	<p>Dr. Rifat Shahriyar (rifat)</p>
<p>GPGPU to achieve Massive Parallelism</p>	<p>GPGPU is the use of a graphics processing unit (GPU), which typically handles massively parallel computation only for computer graphics, to perform computation in applications traditionally handled by the central processing unit (CPU). GPGPUs are used for tasks that were formerly the domain of high-power CPUs, such as scientific computations, physics calculations, encryption/decryption, etc. The use of GPGPU is not ubiquitous yet due to some challenges and overheads introduced by GPU. This work will focus on using GPGPU (preferably in Java) to understand the opportunities and limitations better and then accelerate massively parallel computation/workload related to scientific computing, AI, etc.
</p> <p>Further reading: https://medium.com/@jmaxg3/java-on-gppus-845d9ba58533</p>	<p>2</p>	<p>Dr. Rifat Shahriyar (rifat)</p>

Efficient archival system for genomic data	<p>DNA sequencing has had a major impact on life sciences, and led to many scientific breakthroughs. With the advent of high-throughput DNA sequencing technologies, there has been an exponential increase in the generation of DNA sequence data. This has enabled the meaningful use of genomic data in a wide range of healthcare and biomedical applications. However, this opportunity comes with a threat that increasing DNA sequence volumes will overcome our ability to store the sequences. This project aims to develop novel compression and storage techniques, explicitly targeted for DNA sequence, so we can accommodate the vast amount of genomic data we are witnessing these days. Although there have been quite a few works in this direction, the scientific community still has to go a long way in order to handle the explosive growth of sequencing data. This project aims at contributing towards this direction by developing efficient and novel data structures, file formats, and compression techniques for DNA sequencing data.</p> <p>Please have a look at the following article to get a sense of the motivation behind this project: https://www.technologyreview.com/s/542806/how-do-genome-sequencing-centers-store-such-huge-amounts-of-data/</p>	2	Dr. Md. Shamsuzzoha Bayzid (bayzid)
--	---	---	-------------------------------------

<p>Phylogenetic tree (evolutionary history) reconstruction using whole genome sequences</p>	<p>The overarching goal of this research project is answering impactful biological questions, especially those related to the study of evolution, by developing algorithms that can accurately analyze very large genome-scale datasets. The ongoing big data revolution in genomics can vastly increase our understanding of biology only if our computational toolkit can keep up with the pace of ever increasing abundance of molecular data.</p> <p>In 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, orthology detection, 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, and can arise from incomplete lineage sorting, gene duplication and loss, horizontal gene transfer, hybridization etc. 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>2</p>	<p>Dr. Md. Shamsuzzoha Bayzid (bayzid)</p>
---	--	----------	--

<p>Application of machine learning in predicting protein attributes</p>	<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 structures, foldings, protein-protein-interaction etc. With the success of 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 at applying machine learning techniques to predict protein attributes (in particular, protein structures and protein foldings) from protein sequences.</p>	<p>2</p>	<p>Dr. Md. Shamsuzzoha Bayzid (bayzid)</p>
<p>Identify authorship of binary code</p>	<p>Stylometry is a method for identifying anonymous authors of anonymous texts by analyzing their writing style. While stylometric methods have produced impressive results in previous experiments, the authorship of code is still a challenging problem. Code authorship is useful to identify the authorship of malware, group malware written by the same authors and identify plagiarism. Binary code authorship is challenging because many author specific attributes might get lost due to code obfuscation and compiler optimization. However, some characteristics might survive the compilation process and reveal authorship. The goal of the project is to identify the characteristics of a binary code that are hard to obfuscate and their usefulness for authorship attribution.</p>	<p>2</p>	<p>Dr. Anindya Iqbal (anindyaiqbal)</p>

<p>Fake news detection and attribution</p>	<p>Fake news in social media can have a powerful impact on shaping people's opinion. The goal of this project is to help users identify fake news by comparing similar news from alternative sources. The project will have four parts: 1. find relevant news, 2. identify the similarity or dissimilarity of the alternative news from the position of the current news, 3. determine a reality score to the current news based on various attributes including the facts presented in the news and the past reality score from the same source, and 4. perform a user study to understand the effectiveness of the tool. Identifying similarity of a fact can be challenging because one fact can be presented in many different ways. One way to solve this problem is to summarize news items into actors (people who are involved in the news), the actions they are performing and whether the action is the opposite of the action performed in the current news.</p>	<p>2</p>	<p>Dr. Anindya Iqbal (anindyaiqbal)</p>
<p>Machine learning and NLP based research on Bangla Language</p>	<p>Bangladesh Government has recently taken initiative to develop rich annotated corpus for developing different applications such as spell and grammar checker, sentiment analysis, etc. This project aims to apply state of the art techniques to design such applications using those corpus. There is likely to be opportunity to work in collaboration with industry.</p>	<p>2</p>	<p>Dr. Anindya Iqbal (anindyaiqbal)</p>
<p>Travel time prediction in Dhaka using ride sharing data</p>	<p>Using real world data, travel time will be predicted applying machine learning techniques. There will be opportunity to work in collaboration with ride sharing company.</p>	<p>2</p>	<p>Dr. Anindya Iqbal (anindyaiqbal)</p>
<p>Efficient approaches for DNA motif finding</p>	<p>Motifs are recurring patterns in genome sequences and motif finding is a widely studied problem in computational biology. It has diverse applications, for example in identifying co-expressed genes. Both statistical and combinatorial methods have been widely applied to find motifs in a set of genome sequences. One of the most widely used tools for motif finding is MEME which is based on the EM algorithm. Here we will explore approaches to make motif finding more efficient by pre-processing the genomic sequences without compromising accuracy. * No prior knowledge of biology is necessary but will be helpful</p>	<p>2</p>	<p>Dr. Atif Hasan Rahman (atif)</p>

<p>Genome assembly from second and third generation sequencing reads</p>	<p>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 second generation technologies, however second generation technologies have smaller error rates. We will explore alternate ways to combine strengths of the two types of technologies by</p> <ul style="list-style-type: none"> - constructing de Bruijn graph from second generation reads and "bridging" using third generation reads - merging third generation reads and then correcting them using second generation reads <p>and develop a method for genome assembly.</p> <p>* For some background on genome assembly and sequencing technologies please see http://escholarship.org/uc/item/6hx4q2gm</p> <p>** No prior knowledge of biology is necessary but will be helpful</p>	<p>2</p>	<p>Dr. Atif Hasan Rahman (atif)</p>
<p>Development of applications for health data acquisition</p>	<p>In this project, we will explore effects of environment on common health related issues through participatory sensing. We will develop an application to acquire weather and other environmental data such as noise and information on health related events such as flu, headache and analyze the data for contributing factors, detect seasonal patterns, etc.</p> <p>* Collaboration with Dr. Rifat Shahriyar</p>	<p>1</p>	<p>Dr. Atif Hasan Rahman (atif)</p>