PROJECTS MENTORED BY DR. MAHIMA KAUSHIK AT CIC

1. **Image Captioning and Android App for Visually Impaired persons**
   Students - Rddhima Raghunand, Shobhit Maheshwari

   Computer vision has become ubiquitous in our society, with applications in several fields. In this project, we focussed on one of the visual recognition facets of computer vision, i.e image captioning. The problem of generating language descriptions for visual data has been studied from a long time but in the field of videos. In the recent few years emphasis has lead on still image description with natural text. Due to the recent advancements in the field of object detection, the task of scene description in an image has become easier.

   The aim of the project is to improve the current standard architectures available for the task of Image Captioning and integrate the model with an Android Application, so as to make an app that can work as real-time scene descriptor. The app can be of great utility to Visually Impaired persons, since it can help them navigate, browse and perform other day-to-day activities. In this project, we are trying to improvise on the deepnet, which was based on a combination of VGG16 encoder and an LSTM decoder.

Generated Caption
A group of people are walking on a busy street.

Human provided caption
A crowd of people walk down a busy sidewalk

2. **Generating dataset for wearable biosensors**
   Students: Simran Bansal, Surya Soni, Pranjal Kumar

   Biosensors have found applications in a variety of fields like wound healing, for finding glucose levels and electrocardiogram monitoring. The development of these biosensors can be done using different techniques which involve both micro and nano level processing. Biosensor is a device which detects, transmits and records the information using a bio-analytes. These devices are quick yet efficient and help us to work at a faster rate than the conventional methods of going to the centre, giving sample

and then, finally getting report, which uses a considerable amount of time in this fast-paced world.

The most promising application of nanotechnology is in the field of point-of care diagnostics, which will enable the primary-care physician and patients to perform assays even at their door steps. Nanotechnology though, shows interesting applications in biosensing, the danger of metallic contamination and the safety of chemically developed particles is still questionable. To reduce these risks, the introduction of green synthesis of nanoparticles is being done. In this project, we aim at learning and making nanoparticles using green synthesis, followed by their characterization using various techniques. The next step would be to study their interaction with DNA to generate a dataset for the same.

3. DengAI: Predicting Disease Spread overview

Student: Anustha Kalia

Dengue fever is a mosquito-borne disease that occurs in tropical and subtropical parts of the world. It has become a global problem after the world war two and is common in more than 110 countries. Because it is carried by mosquitoes, the transmission dynamics of dengue are related to climate variables such as temperature and precipitation. Although the relationship to climate is complex, a growing number of scientists argue that climate change is likely to produce distributional shifts that will have significant public health implications worldwide.

DataDriven.org has rolled out a dataset which consists of **climate related features** and the task is to figure out an algorithm to **predict expected number of cases** of dengue. The data pertains to two cities namely, *San Juan* and *Iquitos* along with 20 features pertaining to climatic conditions like temperature, humidity etc.

This data was collected by various U.S. Federal Government agencies—from the Centers for Disease Control and Prevention to the National Oceanic and Atmospheric Administration (NOAA) in Department of commerce.

The aim of this project was to develop an understanding of relationship between **climate** and **dengue dynamics** through extensive literature survey. Also, using insights gained from (1) to develop an algorithm which is able to predict number of cases of dengue on basis on climate and environment related features.

4. **Brain Computer Interface: Applications specifically for paralyzed persons**
   Student: Pankaj baranwal

Electroencephalography or EEG is an electrophysiological monitoring method which is widely used to monitor and record the brain wave patterns. These patterns can then be used to analyse the brain activity and derive meaningful information about a person's mental state. A brain computer interface (BCI) is a direct communication pathway between an enhanced or wired brain and an external device which monitors and analyses the EEG signals. Hence the aim of this project was to utilize a BCI to analyse brain waves and find out unique patterns so that physical functions can be mapped to brain signals. In the world, there exist many BCI-enabled devices but there are many limitations of such devices.

Hence, it was proposed to develop algorithms which are better equipped at dissipating noise and reading the actual useful part from the entire EEG signal. Once this task is complete with satisfactory results, the study of the field of neuroprosthetics would be done, which is still in an infancy stage and develop a device to send artificial sensory information to the brain by using appropriate neural prosthesis.

Such a device would be able to read neural signals from the brain and use them as instructions by a paralysed patient to show actual movement in their limbs. It could also be used to detect certain diseases which affect brain waves in particular pattern.

5. **Small interfering RNA (siRNA) and its varied applications: An extensive Review**
   Students - Rddhima Raghunand, Shobhit Maheshwari

Since the Nobel-prize winning discovery of the RNA interference (RNAi) technique in 2006, several attempts have been made to use it to design and develop drug treatments for a variety of diseases, including cancer. In this review, we have focussed on the potential of small interfering RNAs (siRNA) in anticancer treatment. The significant barriers that exist on road to clinical application of siRNA drugs have been

described. Current delivery approaches of siRNAs using lipids, polymers and in particular, polymeric carriers that overcome the aforementioned obstacles have also been added. Additionally, siRNA databases along with a qualitative and/or quantitative summary of the data in each database have been collated.
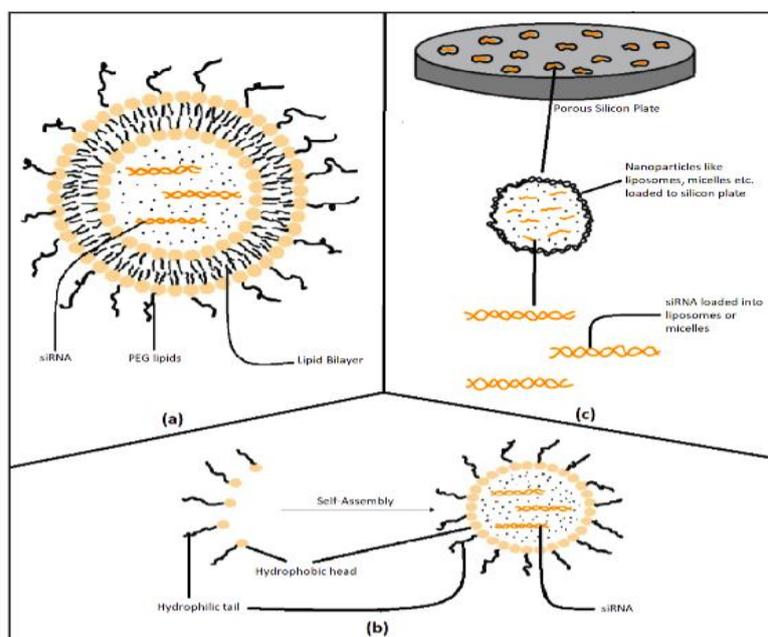

Fig.   1:
(a) Liposomes, (b) Polymeric micelles nanoparticle,(c) Multi-stage delivery vector for siRNA

6. **Integrated bio utility tool**

   Students - Anushtha Kalia, Rddhima Raghunand, Shobhit Maheshwari

   This project dealt with the study and implementation of various algorithms related to the field of computational biology. The aim is to enhance the understanding of the algorithms related to biology. There are various online tools available for the analysis of DNA, RNA and proteins but as such an integrated platform for the tools does not exist. Various basic biology problems are available on the website of Rosalind which is a coding platform for learning bioinformatics and programming through problem solving. Bioinformatics tools aid in the comparison of genetic and genomic data and more generally in the understanding of evolutionary aspects of molecular biology. Thus, the idea was to develop a single tool, which can work for multiple uses.

7. **Predicting Secondary Structure of Proteins**

   Student: Tanishq kumar dangar

   Computer Science as a tool to make human lives easy has always been an effort of various leading institutes throughout the globe. Statistical methods have been of great

help when it comes to analyzing data. Machine learning is one of the leading domains in data analysis. Now, the later comes the question that 'can these tools be used to ease biology problems?' Well bioinformatics is one such domain helping biology solve some mysteries. Computer Science have been continuously trying to mimic nature and human body so as to modify computer systems in terms of space and time efficiency. Artificial Neural Networks (ANN) are one such example. Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) also follow the same theory. Another question which raised in my mind are 'can this biology inspired systems solve biological problems itself?'

This project aims to get an answer to both the questions. A question many computer scientists have been trying to solve is 'Predicting Secondary Structure of Proteins' and in this review, an answer to the above-mentioned questions with the help of this problem of predicting secondary structures for protein have been explored.

8. Finding and analyzing the repeating sequences in a Proteome
Student - Shobhit Maheshwari

Various protein sequences in a proteome have a very similar chain of amino acids. The slight changes among these sequences may be a result of either a mutation or the similar properties of the changed amino acids, for example in case of *Mycobacterium Tuberculosis* EsxP and EsxK gene only the third amino acid from the start varies, while all the other parts of the sequence are exactly the same. In case of the EsxP the third amino acid is Threonine (T), while in case of EsxK third amino acid observed was Serine(S). Threonine and Serine have almost the same physical and chemical properties. Thus, we can draw a conclusion that both these proteins have almost similar functionalities, which can be verified from the vast database of functions of proteins. The first step towards the problem is obviously to find the recurring sequence. The recurring sequence in this project were found using a self-written code. The next step was to find the protein sequences corresponding to those recurring sequences and analyze them using the tools and databases available online. The study involved the properties of the changes in amino acids and their functionality or in case if the changed proteins have dissimilar properties etc. In case of the dissimilarity of the changed proteins, we might make a guess that the sequence is mutated and get it validated from databases available online.
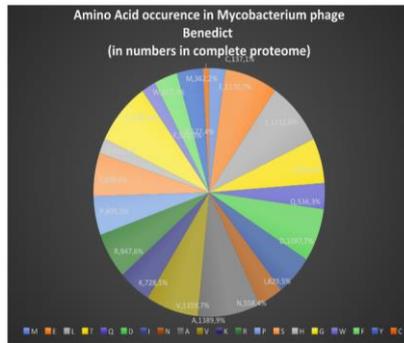
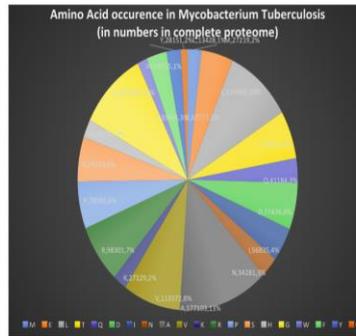Figure 3: Pie chart of amino acid in Mycobacteria phage benedict

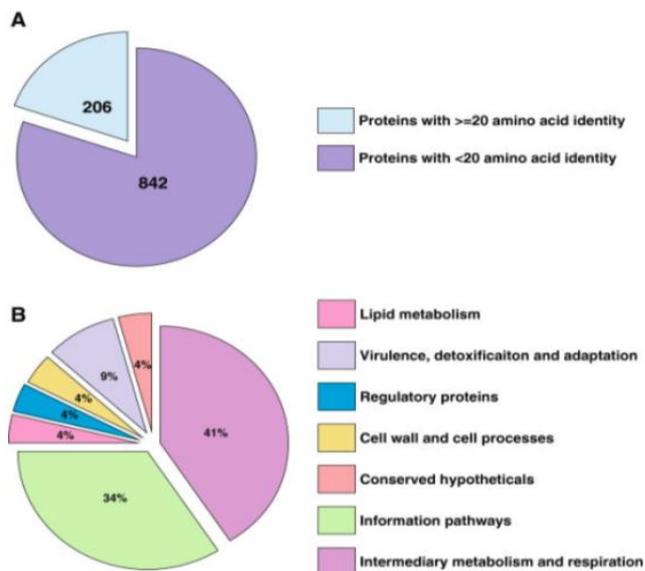Figure 4: Pie chart of the amino acid occurrence in proteome of mtb

Figure 5: Analysis of the proteome of mtb

9. RNA secondary structure prediction algorithm

Student - Shobhit Maheshwari

Objective - Given a sequence of nucleotides of a RNA sequence predict the secondary structure of the sequence and to interpret the dot-bracket notation returned by the algorithm so as to predict bulges along with the hairpin loops.

Approach (Nussinov-Jacobson Algorithm)- The algorithm is a dynamic programming approach to predict the secondary structure of the RNA sequence. The stability of the sequence is directly proportional to the negative energy of the structure. The bulges and loops in the structure contribute to positive energy. Thus, the structure should not contain too many loops and bulges. The nucleotide sequence present inside the loop does not interact when and RNA sequence combines with other molecules. Another algorithm named Zuker's algorithm takes the stacking energies into

consideration hence predicting the most suitable structure for the sequence. NJ algorithm takes into account the G-U wobble base pair by assigning weights to all the interactions i.e. A-T bonding is given a weight of 2, since G-C bond is a stronger bond due to the presence of triple H-bond thus the weight assigned is 3 and the G-U base pair is assigned as 1.

The output for the algorithm is a dot-bracket notation which is a sequence of dots and bracket which can be interpreted as follows:

i) (........) :- The starting and the opening bracket signify the starting and ending point of the loop. The number of dots means the number of nucleotides inside the loop.

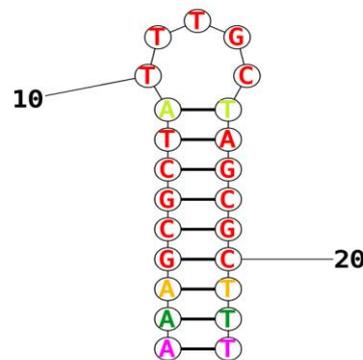ii) (.......(:- The opening bracket followed by another opening bracket means a bulge formation.

Fig.: RNA structure for the sequence AAAGCGCTATTTGCTAGCGCTTT

10. Brain Computer Interface: Applications specifically for paralyzed persons
   Student: Pankaj baranwal

Electroencephalography or EEG is an electrophysiological monitoring method which is widely used to monitor and record the brain wave patterns. These patterns can then be used to analyse the brain activity and derive meaningful information about a person's mental state. A brain computer interface (BCI) is a direct communication pathway between an enhanced or wired brain and an external device which monitors and analyses the EEG signals. Hence the aim of this project was to utilize a BCI to analyse brain waves and find out unique patterns so that physical functions can be mapped to brain signals. In the world, there exist many BCI-enabled devices but there are many limitations of such devices.

Hence, it was proposed to develop algorithms which are better equipped at dissipating noise and reading the actual useful part from the entire EEG signal. Once this task is complete with satisfactory results, the study of the field of neuroprosthetics would be done, which is still in

an infancy stage and develop a device to send artificial sensory information to the brain by using appropriate neural prosthesis.

Such a device would be able to read neural signals from the brain and use them as instructions by a paralysed patient to show actual movement in their limbs. It could also be used to detect certain diseases which affect brain waves in particular pattern.