Artificial Intelligence / Machine Learning for Healthcare

Delhi Research Implementation and InnoVation (DRIIV)

Delhi Science & Technology Cluster

Table of Contents

Executive Summary

COVID-19 pandemic has spurred the much-needed transformation of healthcare systems by forcing us to think about end-to-end to end solutions. This proposal aims to bring together the Delhi Cluster of institutions (and beyond) to engage and create dovetailed solutions from laboratory research to point-of-care solutions for healthcare. Six academic institutions along with industry partners, Startups, NGOs, and a wider collaborative network have joined hands to create **D**elhi **R**esearch **I**mplementation and **I**nno**V**ation (DRIIV) Cluster with AI/ML in Healthcare as one of the thematic areas. The long-term (10-year) focus of the cluster is to enable health systems transformation through collaborative technology building. The near-term (3-year) focus areas of the AI/ML in Healthcare theme are diagnostic, predictive, and prognostic solutions for Public health challenges including COVID-19, Antimicrobial Resistance, and Tuberculosis, Maternal and Child health**,** Non-Communicable Diseases, through health systems solutions and technology development. The cluster will also engage with other verticals such as effective education to develop curricula for skilling and capacity building in AI/ML for healthcare.

Problem Statement

Health systems reform is a systemic change and cannot be envisaged through isolated research and development. The ongoing COVID-19 pandemic has exposed vulnerabilities in the health systems all across the globe. On the other hand, the threat has also provided an extra-ordinary opportunity for a collective multi-sectoral response achieved with systems thinking and coordinated local execution. **The overarching goal of this theme is to enable a transformative change in healthcare through Artificial Intelligence and Machine Learning collaborations for (i) targeted public health challenges and epidemiological surveillance (ii) data, modeling and predictive models, (iii) digital tools, Apps, and platforms including conversational AI, (iv) diagnostics and prognostic solutions using computational and biotechnological solutions (Figure 1).** This theme will actively engage with other themes of the Delhi Science and Technology Cluster such as Effective Education, Mobility, Air and Water sub-themes for collaborative progress towards healthcare systems improvement. **The key areas in healthcare that will be addressed include**:-

1. **Public health challenges.**
    - **Infectious diseases** including COVID-19, Antimicrobial Resistance, neonatal sepsis and Tuberculosis.
    - **Non-Communicable Diseases** including Cancer, Rare Neuromuscular Disorders, and Spinal Disorders
2. **Health systems challenges.**
    - Federated AI and data sharing platform for operationalizing a cluster response.

- Data exchange, interoperability, and security, including tele-medicine and tele-diagnostics.
- Analytics and predictions
- Resource allocation and supply chain optimization

3. **Technology development.**
   - AI/ML based predictive models
   - Databases
   - Novel diagnostics including point of care pore based solutions, chemoinformatics, and biological assays.

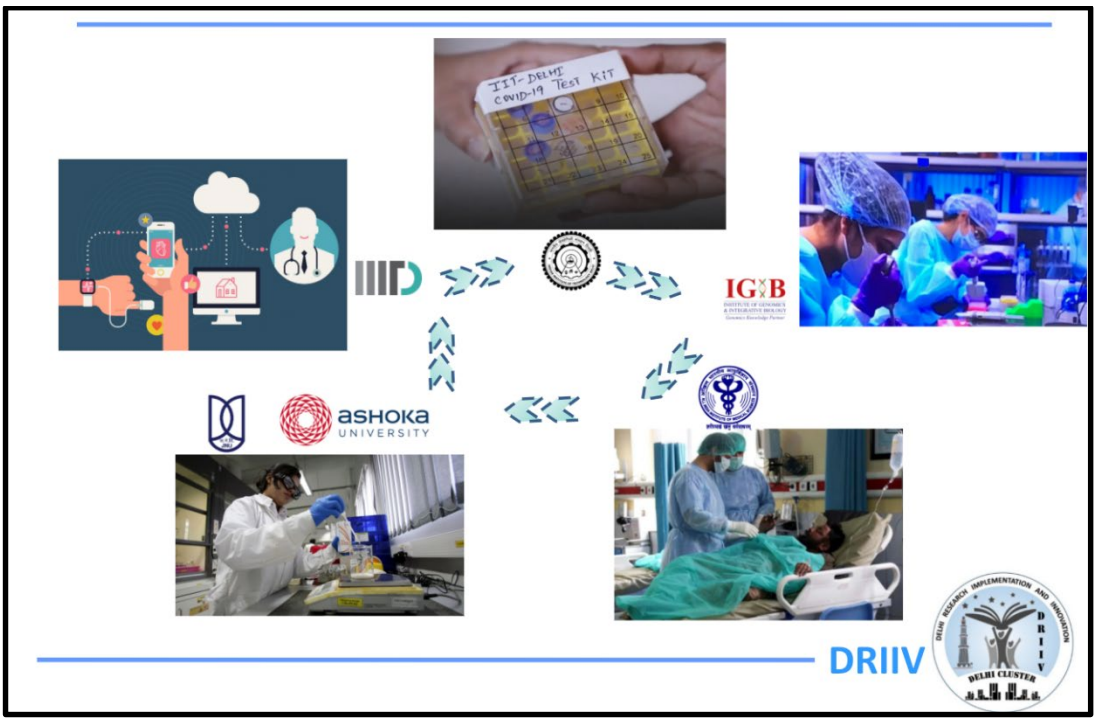


Figure 1. A cluster approach to healthcare solutions with technology. A dovetailed approach combining clinical, computational, biotechnological, and implementation science expertises is proposed for infectious diseases, non-communicable diseases, and public health challenges.

Specific Projects.

The following sub-projects will be undertaken within the AI/ML for healthcare theme with the involvement of cluster members.

1. **WP1. An AI enabled Federated platform for data exchange, analytics and predictions in infectious diseases.**
    - **Focus areas.**

      COVID-19, Antimicrobial Resistance and Neonatal Sepsis, Tuberculosis.

    - **Investigators and Institutions**

      PI: Dr. Tavpritesh Sethi (IIIT-Delhi), Co-PIs: Dr. Raghava Mutharaju (IIIT-Delhi), Prof. Ponnurangam Kumaraguru (IIIT-Delhi), Dr. Harpreet Singh (ICMR), Dr. Rakesh Lodha (AIIMS), Dr. Ahmadulla Shariff (AIIMS), Prof. Anurag Agrawal (IGIB), Shama Karkal (Swasti).

    - **Goal.**

      To build a real-time, deployable, and scalable data exchange and analytics solution for COVID-19, AMR, neonatal sepsis and antibiotic resistance predictions across different organizations.

    - **Approach.**

      Methodology. A novel multimodal AI approach to integrate and analyze heterogeneous epidemiological data will be used. Our integrated platform for epidemiology will be the first of its kind in the world that will combine syntactic and semantic interoperability layers to suggest mappings of different datasets for deriving combined insights. We will build novel natural language processing based intelligent middleware for query translation and ease of use by non-specialists in the form of conversational agents. The applications developed will be **NDHM compliant, hence scalable across the country** by leveraging the Consent Manager (CM), Health Information Provider (HIP) and Health Information User (HIU) APIs provided by NDHM sandbox. **A pilot version of COVID-19 data integration and OTP based authentication framework has already been developed at the investigator's lab using publicly available datasets and deployed using a multi-tenant architecture (Figure 2)**. A key AI/ML feature of the application will be NLP based intelligent middleware for semantic information to allow easier mapping of data sources and for ease of use by non-experts in hospitals by providing query translation interfaces.

    - **Specific Outcome/Technology.**
        - An AI enabled data port, analytics platform and federated AI platform for predictive modeling in (re)emerging epidemics

- An android App for active surveillance of infectious diseases,
- Focus on COVID-19, Antimicrobial Resistance and Tuberculosis.

- **Budget.**

  5.964 Cr

Broad Budget for WP1

	Year 1 in Lakhs	**Year 2 in Lakhs**	**Year 3 in Lakhs**	**Year 4 in Lakhs**	**Total in Lakhs**
Capital Expenditure	40	40	0	0	80
Operating Expenditure	174.8	174.8	166.8	0	516.4
Totals	214.8	214.8	166.8	0	**596.4**

Detailed Budget for WP1

Head	**Year 1**	**Year 2**	**Year 3**	**Total**
Capital Equipment	40	40	0	80
Consumables	25	25	25	75
Manpower	51	51	51	153
Data collection field studies in hospitals and Primary Health partners	50	50	50	150
Travel & Training	3	3	3	9
Contigencies	10	10	10	30
Overheads	35.8	35.8	27.8	99.4
Total	**214.8**	**214.8**	**166.8**	**596.4**
Grand Total (in figures)	**596.4 Lakhs**			

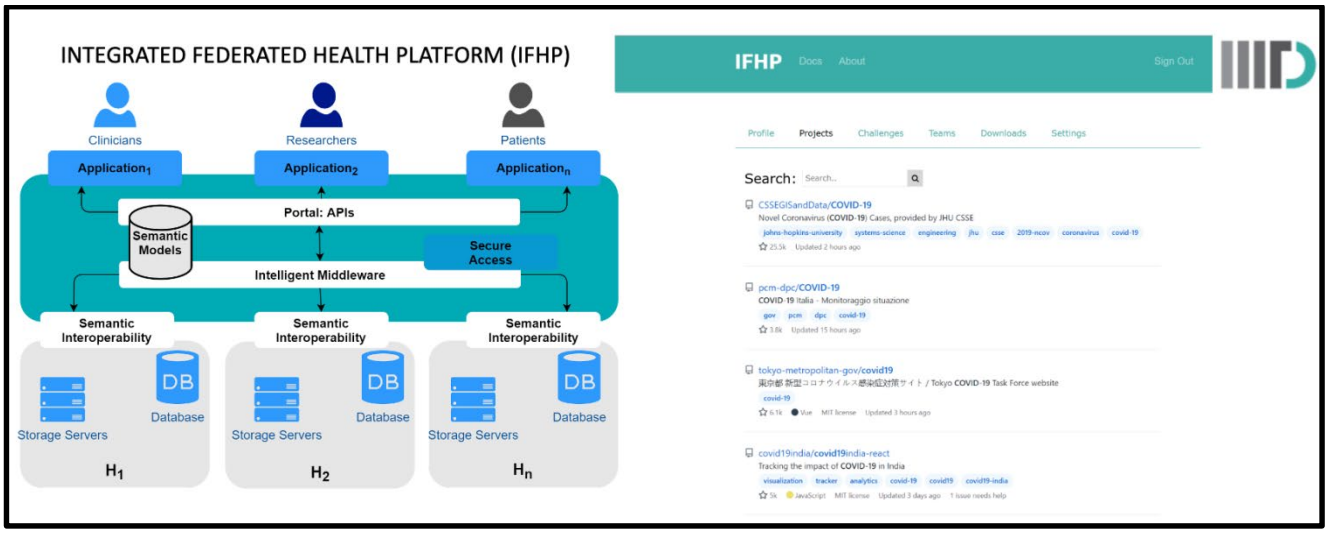


Figure 2. Proposed architecture for federated interoperable, intelligent, and secure platform for data exchange, analytics and predictions for epidemics. Investigators have created an NDHM compliant pilot platform with OTP authentication and analytics carried out on publicly available datasets. The deployment version of the platform will have integrated intelligent middleware for NLP based query translations, analytics and computational pipelines for predicting emerging trends from structured and unstructured data[1,2,3,4].

2. WP2. Deep phenotyping for long-term sequelae of COVID-19.

- **Focus areas.**

  COVID-19

- **Investigators and Collaborators**

  PI: Prof. Anurag Agrawal (CSIR-IGIB), Co-PIs: Dr. Rajesh Pandey (CSIR-IGIB), Dr. Tavpritesh Sethi (IIIT-Delhi), Dr. Shinjini Bhatnagar (THSTI), Dr. Ramachandran Thiruvengadam, Dr. Koundinya Bapu Desiraju

[1] "Symptoms and medical conditions in 204 912 patients visiting ...." 1 Dec. 2015, https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(15)00152-7/fulltext. Accessed 12 Mar. 2021.
[2] "Predicting Hemodynamic Shock from Thermal Images using ...." 14 Jan. 2019, https://www.nature.com/articles/s41598-018-36586-8. Accessed 12 Mar. 2021.
[3] "Generalized Prediction of Shock in Intensive Care Units using Deep ...." 11 Jan. 2021, https://www.medrxiv.org/content/10.1101/2021.01.07.21249121v2.full. Accessed 12 Mar. 2021.
[4] "Predicting Emerging Themes in Rapidly Expanding ... - medRxiv." 15 Jan. 2021, https://www.medrxiv.org/content/10.1101/2021.01.14.21249855v1. Accessed 12 Mar. 2021.

- **Goal.**

  To phenotype COVID-19 patients for long-term outcomes and multi-omic profiling with care pore based sequencing.

  Scale-Speed-Sensitivity-Sample-cost (4S) has been the need of the day during COVID-19 pandemic when central dogma regarding Samples to the Sequencer needs revisiting with an alternative of *Sequencer to the Samples*. Towards this we need balance between the above 4S aspects. It would be enabled by the sequencing ability within the minimum available infrastructure. Nanopore based sequencing would be able to meet all the above criteria regarding the SARS-CoV-2 genome sequence. CSIR-IGIB has been leading many COVID-19 initiatives, nationally and internationally, wherein we have developed SOPs (both experimental and analysis) for genomic surveillance of SARS-COV-2. We intend to develop a hub and spoke model wherein MicroLabs would be sourcing data to the central server. The clinically important samples would be taken through multi-omic profiling which may include Global Screening Array (GSA), RNA-Seq and holo-transcriptome. In combination they would help understand the regulation at DNA, RNA and microbial interaction, with a modulatory role in clinical outcome. This would augment integrative genomics of SARS-CoV-2.

- **Approach.**

  Deep phenotyping of 2500 COVID-19 samples at diagnosis and follow up for characterization of long term outcomes and predictions.

  Given the diversity of the symptoms, severity and clinical outcome; it is important to capture multimodal data from the SARS-CoV-2 positive individuals inclusive of age, gender, co-morbidities, surgical history, medication, respiratory support, X-ray, blood parameters, hospital stay and clinical outcome. The integration of this multimodal dataset with the genome sequence is important to understand the causal reason/s for the differential clinical history and outcome. The clinical partners would be the valued partner and their effort towards capturing the longitudinal dataset would be at the core.

  COVID-19 has also highlighted the missing link wherein same virus clade with differential clinical severity and clinical outcome. Thus, it is important to understand the host for its phenotypic diversity representation and investigate the possible

reasons explaining the sub-phenotype. This can be captured using three complementary methods inclusive of:

- Genome architecture of the pathogen (SARS-CoV-2),
- Host response captured through transcriptome, and
- Possible role of co-existing/co-infections in clinical outcome.

All the above would be elucidated in combination with the clinical data on a longitudinal scale. Important publications highlighting the investigator contribution in COVID-19 genomic surveillance includes:

We published the first pan-India SARS-CoV-2 integrative genomic study and other important ones in this field are:

- *Integrated genomic view of SARS-CoV-2 in India*. Wellcome Open Research, 2020.
- *Clinical, Serological, Whole Genome Sequence Analyses to Confirm SARS-CoV-2 Reinfection in Patients From Mumbai, India*. Frontiers in Medicine, 2021.
- *Threading the Pieces Together: Integrative Perspective on SARS-CoV-2*. Pathogens, 2020.
- *Nature and dimensions of the systemic hyper-inflammation and its attenuation by convalescent plasma in severe COVID-19*. Journal of Infectious Diseases, 2021.
- *Noncoding RNAs: modulators and modulatable players during infection-induced stress response*. Briefings in Functional Genomics, 2020.
- *Next generation sequencing for pandemic preparedness*. Indian Chemical Engineer, 2020.

- **Outcomes.**

Database and models for multi-omic profiles of COVID-19 patients in follow-up in order to characterize long term progression.

COVID-19 pandemic challenged us in many ways but also provided an opportunity to come together between stakeholders with common gold of better COVID-19 surveillance (molecular and genomic) and parallel efforts towards pre-emptive identification of sub-groups requiring focused medical attention. The combination of both the above efforts would lead to containment of diseases spread and reduced mortality. The clinical data in combination with the blood parameters and the genomics data would be used to harness the strength of AI-based models for risk stratification and mortality prediction. We recently submitted the first paper on the clinical and laboratory data analysis to Scientific Reports and have posted them in medRxiv (https://www.medrxiv.org/content/10.1101/2020.12.19.20248524v1).

- **Budget**

502.5 Lacs

Detailed Budget

				All Values are Rs. in lacs
Head	**Year 1**	**Year 2**	**Year 3**	**Total**
Capital Equipment	60	20	20	100
Consumables	100	80	40	220
Manpower	43	43	43	129
Travel & Training	2	2	2	6
Contingencies	1	1	1	3
Overheads	20	14.5	10	44.5
Total	226	160	116	502.5
Grand total (in figures)	**502.5**			

3. **WP3. AI-guided liquid biopsy as a mass screening assay for cancer: Drug response prediction and monitoring leveraging gene expression readouts of blood platelets**
   - **Focus areas.**

     Cancer

   - **Investigators and Collaborators:**

     PI: Dr. Debarka Sengupta (IIIT-Delhi), **Collaborators:** Prof. Dr. Colleen Nelson (Queensland University of Technology), Prof. Dr. Stefanie Jeffery (Department of Surgery, Stanford University), Dr. Naveen Ramalingam (Fluidigm Corp., USA), Prof.

Dr. Ritu Gupta (AIIMS, New Delhi), Mr. Pallab Mitra (CareOnco Biotech Pvt Ltd), Dr. Gaurav Ahuja (IIIT-Delhi), Prof. Jayadeva (IIT-Delhi)

- **Goal.**

  1. Developing a diagnostic kit basis a comprehensive platelet-gene panel[5] for early and affordable blood based cancer detection. 2. Exploring the potential of platelet gene expression in treatment decision making and therapeutic response monitoring.

- **Approach.**

  A platelet gene panel[5] has been developed by the investigators that along with a pre-trained AI model predict the existence of cancer with an accuracy of 97% (Figure 3). The aim for this proposal is this gene panel and bringing it to bear as a mass-screening liquid biopsy assay. As part of the mega cluster project proposal we aim to evaluate potential technologies to transform the gene panel into a diagnostic kit. The technologies we plan to evaluate include droplet digital PCR (ddPCR) and nanopore based RNA-sequencing technologies. We are also collaborating with Fluidigm Corp. to help us evaluate the Biomark HD system to enable low cost mass screening of cancers. The proposed platelet based diagnostic assay will allow serial profiling and monitoring throughout disease life cycle. We will evaluate the worth of platelet gene expression readouts for personalised drug response prediction (along the lines of https://thesenguptalab.shinyapps.io/Precily/, a prototype developed by Dr. Debarka Sengupta's group). RNA-seq data suffer from batch effects. Our preliminary plan is to use Empirical Copula (and Copula Entropy) based (Pattern Recognition Volume 112, April 2021, 107697) methods within the Bayesian (hierarchical) framework for making drug response predictions. Notably, empirical copula relies on order statistics, thus unaffected by batch differences to a large extent.

- **Outcomes.**

  1. A mass screening assay for cancer. 2. RNA-seq data repository comprising serial profiling of platelet gene expression in multiple patients, along with their treatment history.

- **Budget.**

  2.26 Cr.

[5] "Molecular signature comprising 11 platelet-genes enables accurate ...." 27 Oct. 2020, https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-020-07147-z. Accessed 12 Mar. 2021.

Detailed Budget

				(All Values are Rs. In lacs)
Head	Year 1	Year 2	Year 3	Total
Capital Equipment	70	0	0	70
Consumables	25	15	5	45
Manpower	18	18	18	54
Travel & Training	3	3	3	9
Contingencies	2	3	5	10
Overheads	23.6	7.8	6.2	37.6
Total	141.6	46.8	37.2	225.6
Grand Total (in figures)				

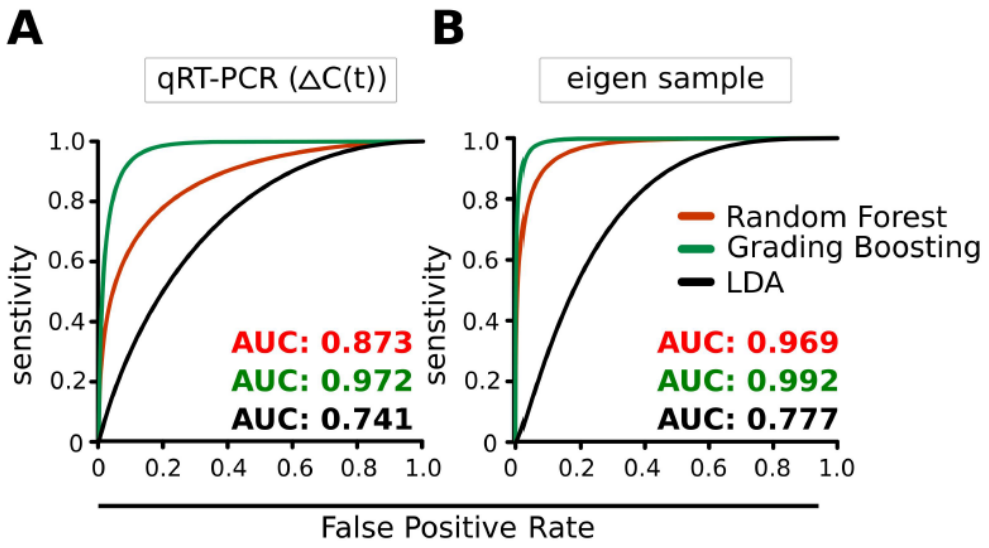


Figure 1: (A) AUC (Area under the curve) plot representing the performances of three independent classifiers i.e. Gradient Boosting Machines (GBM), Random Forest (RF), and Linear Discriminant Analysis (LDA) in distinguishing tumor and healthy samples using Cq values of 11 genes from 10 NSCLC patients and 7 healthy controls. **(B)** AUC plot depicting the improvement in the classification accuracy by augmenting the data-points with artificial samples, using the EigenSample technique.

4. WP4. Development of models for prediction of disease severity and progression for a few genetic neuromuscular disorders.

- **Investigators and Collaborators:**

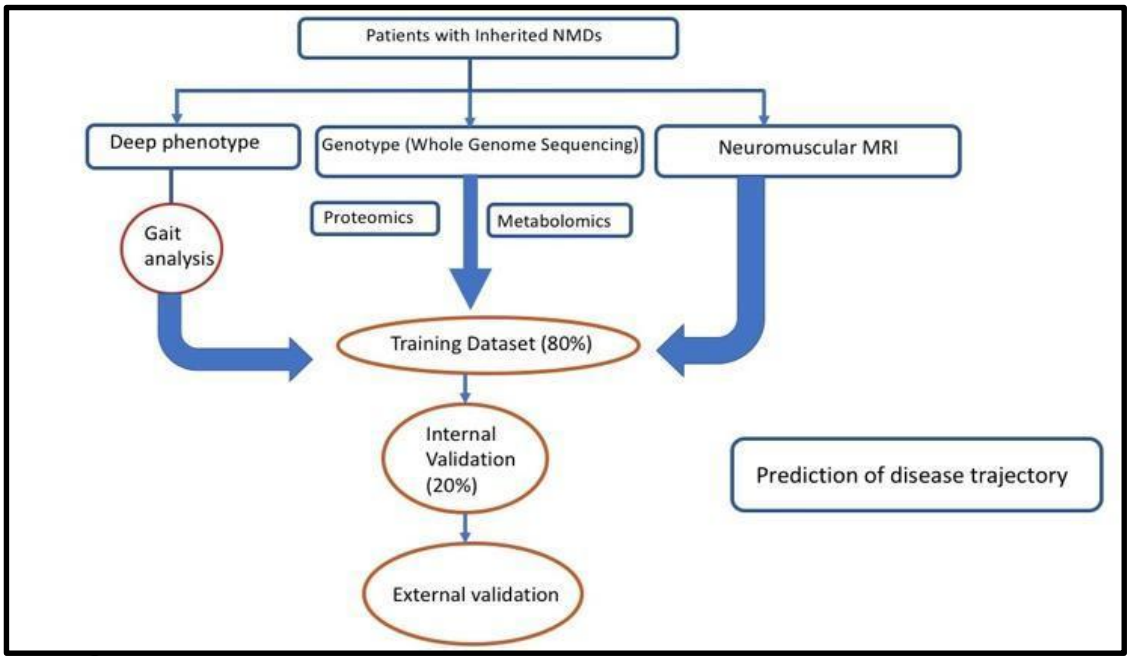
  Alok Bhattacharya (Ashoka University), Gautam Menon, (Ashoka University) Manu Awasthi (Ashoka University), VY Vishnu, (AIIMS, New Delhi), MV Padma Srivastava(AIIMS, New Delhi), Ajay Garg, (AIIMS New Delhi), Ashwin Srinivasan, (BITS-Pilani K.K.Birla Goa Campus), Tanmay Verlekar (BITS-Pilani K.K.Birla Goa Campus)

- **Goals:**

  Our main goal is to find clinical, environmental and genetic factors that influence disease severity and progression. A large number of patients suffer from nearly 1000 genetic disorders that affect muscle, quite often leading to either major disability or loss of life in the early ages. There is no cure for almost all these diseases. For management of these diseases including finding new therapies, we need to understand disease progression in Indian patients. The major objective of the proposal is to develop machine learning models from a large data set of clinical parameters (including gait analysis), genomic data (whole genome sequencing, protein and metabolic profiles), and Neuromuscular MRI to diagnose Inherited Neuromuscular disorders, differentiate between various subtypes and predict disease trajectory of the patients.

- **Approach:**

  We intend to construct models that will allow us to identify features that are predictive of the disease, its severity, and progression at 3 levels of increasing complexity. 1) The models that simply identify the (probability of) presence or absence of a disease. At this level of detail, we expect the genotype data would be sufficient. 2) The models that identify the stage of the disease, categorised in some clinically meaningful manner (for example: early onset, intermediate, advanced). At this level of detail, we expect that both genotype and phenotype data would be needed. 3) These models (regression models) will forecast progression from one stage to another, in terms of months or years. At this level of detail, we expect genotype, phenotype and MRI data would be needed.



edictive, Diagnostic and Prognostic models for NMDs; A simple image-based gait analysis tool that will help to identify foot drop in the early stages of the disease and measure progression; Models that will accurately model disease progression and severity providing prognostic opportunity and target for new therapy development; Patient database that will have invaluable information for future work and drug trial.

- **Budget.**
  5.1 Cr

Detailed Budget

Overall budget details					All values in Rupees
Head	Year1	Year2	Year3	Year4	Total
Capital Equipment	4,140,000	800,000	0	0	4,940,000
Consumables and services	6,390,000	7,450,000	8,650,000	7,790,000	30,080,000
Manpower	2,040,000	2,040,000	3,000,000	3,000,000	10,080,000
Travel and training	210,000	150,000	150,000	210,000	720,000
Contengency	300,000	300,000	300,000	300,000	1,200,000
Overhead	900,000	900,000	900,000	900,000	3,600,000
Total	13,980,000	11,640,000	13,000,000	12,200,000	50,820,000

5. WP5. A secure learning framework for latency-constrained tele-surgery and tele-diagnostics.

- **Investigators and Collaborators.**

  Dr. Harshan Jagadeesh (IIT-Delhi), Dr. Ranjitha Prasad (IIIT-Delhi)

- **Goals.**

  To develop privacy preserving federated learning based secure protocols for tele-presence based surgery and low-latency tele-diagnostic applications.

- **Approach.**

  Data-driven secure privacy-preserving approach that works over wireless networks that fulfil the latency and other QOS constraints for Tele-surgery and emergency tele-diagnostics. This requires that the network fulfils latency based quality-of-service (QOS) requirements such as latency requirement of less than one millisecond between devices. In addition, these links need to be secure, due to the sensitive nature of the information exchanged.

- **Outcomes.**

  Data-driven secure privacy-preserving protocols over wireless networks that fulfil the latency and other QOS constraints for the above mentioned applications; Demonstration of the above data-driven secure protocols for different types of medical signals/data using an in-situ lab setup using wireless devices; Technical papers/journals in transactions related to wireless communications and machine learning.

  Budget. 121.664 Lakhs

Detailed Budget

Head	Year 1	Year 2	Year 3	Total
Capital Equipment	10	0	4	14
Consumables	1	1	1	3
Manpower	9.24	9.24	11.24	29.72
Travel & Training	2	2	2	6
Contigencies	2	2	2	6
Overheads	4.848	2.848	4.048	11.744
Total	29.088	17.088	10.288	69.464

Grand Total (in figures)	69.464			
Head	Year 1	Year 2	Year 3	Total
Capital Equipment	10	0	4	14
Consumables	1	1	1	3
Manpower	4.62	4.62	5.26	14.5
Travel & Training	2	2	2	6
Contigencies	2	2	2	6
Overheads	3.924	1.924	2.852	8.7
Total	23.544	11.544	17.112	52.2
Grand Total (in figures)	52.2			

6. WP6. AI and chemoinformatics-based solutions in diagnostics, therapeutics and healthcare

- **Investigators and Collaborators: PI:** Dr. Gaurav Ahuja (IIIT-Delhi); **Collaborators:** Asst. Professor, Dr. Debarka Sengupta (IIIT-Delhi), Dr. Subrata Chattopadhyay (IIT Patna), Dr. Srivatsava Naidu (IIT Ropar), Prof. Jayadeva (IIT-Delhi)

- **Focus Area**

  Diagnostics, Therapeutics, and Healthcare

- **Goals**

  To develop diagnostic, therapeutic and repellant odorants using AI driven chemoinformatics

- **Approach.**

  Smell is phylogenetically the most ancient sense, and has been exploited to induce distinct behavioral responses, such as aversive response (repellents), or anti-stress ambience (Aromatherapy), and most recently as diagnostic for COVID-19 infection. Dr. Gaurav Ahuja's laboratory was the first to explain the cellular basis of the loss of smell in COVID-19 infected individuals (Gupta et al. 2020). After more than a year of this pandemic, reports are highlighting a long-term impact of infection on olfactory health, and it is emerging that a periodic olfactory exercise could potentially accelerate recovery. The investigators will leverage AI to synthesize odorants cocktail, which will be deployed as an easy-to-use kit for the faster recovery of

olfactory performance in COVID-19 patients. Similar AI-driven strategy will be employed to design and devise odorant cocktails for anti-depression treatment (Aromatherapy) or for mosquito repellents. Of note, using AI and chemoinformatics methods Dr. Ahuja's laboratory identified an array of Natural Phytochemicals whose predicted repellent efficacy is as good as that of DEET (a gold standard mosquito repellent).

- **Outcomes**

  1.Odorant cocktail-based ***portable tool-kit*** for human nasal exercise for the faster regaining of olfactory loss in COVID-19 patients. 2. Development of prototype solutions for natural, eco-compatible, and health-friendly mosquito repellent solutions. 3. Development of odorant-based anti-depression solutions (Aromatherapy).

- **Budget.**

  2.56 Cr.

				(All Values are Rs. In lacs)
Head	**Year 1**	**Year 2**	**Year 3**	**Total**
Capital Equipment	50	40	0	90
Consumables	25	25	5	55
Manpower	18	18	18	54
Travel & Training	3	3	3	9
Contingencies	2	3	5	10
Overheads	23.6	7.8	6.2	37.6
Total	121.6	96.8	37.2	255.6
Grand Total (in figures)	2,55,60,000 (2.55 Cr)			

7. WP7. AI enabled detection and intervention of future variants of SARS-COV2

- **Focus areas**

  COVID-19

- **Investigators and Collaborators:**

  PI: Dr. Arjun Ray (IIIT-Delhi), Co-PIs: Dr. Tavpritesh Sethi (IIIT-Delhi)

- **Goal**

  To characterize all the possible genetic variants in the S-protein of SARS-COV2 and its efficacy of interaction with existing drugs, for enabling a timely AI-enabled detection of future outbreaks and speculative intervention.

- **Approach**

  Utilizing protein modeling and virtual screening, all the genetic variants of S-protein shall be screened against pharmacologically active molecules. Post experimental validation, using positive examples' geographical origin, along with AI-derived signals. The interaction database will not only be used to detect emerging mutants but also propose interventions based upon pharmacologically active molecule interacting with drugs.

- **Outcomes.**
  - Development of a comprehensive database of correlating leads of drug-variant interactions.
  - Development of pipeline to learn social media trends for outbreaks, vaccine hesitancy and adverse effects monitoring.
- **Budget. Rs 1,27,15,200**

1	Equipment -> Workstation - 3	3000000
A'	***Total - Capital***	3000000
B	**Recurring Items (All calculated for three years total)**	
1	Manpower -> Junior Research Fellow - 2	22,32,000

2	Consumables (Regular consumables + experimental validations)	5000000
3	Travel	300000
4	Contingencies	300000
5	Other Cost (Publication costs etc)	1000000
6	Scientific Social Responsibility	0
	General - I (Manpower, Consumables, Travel, Contingencies, Other Cost, SSR)	**88,32,000**
	General - II Overhead Charges (@10%)	**8,83,200**
B'	***Total - Recurring***	***97,15,200***
C	**Total cost of the project** (without overhead)	**1,18,32,000**
D	**Total cost of the project (A' + B')**	1,27,15,200

8. WP8 Test for monitoring health effects of pollution using 3D microscopy of unstained RBCs from peripheral smear

1. **Investigators: Prof. Kedar Khare (IIT Delhi)**
2. **Co-investigators: Dr. Ritu Gupta (AIIMS), Prof. Sagnik Dey (CAS, IIT Delhi)**
3. **Focus area: Diagnostic imaging/ microscopy, environmental health effect monitoring**
4. **Goals: We propose a research program aimed at developing a simple test involving 3D imaging of blood cells from peripheral blood smear using digital holographic microscopy (DHM) technology for personalized assessment of the impact of environmental pollution.**
5. **Approach: While air pollution has become a part of public discourse, there is no simple test for an individual to understand the short- and long-term effects of environmental pollution on health. As a result, people do not take pollution related health issues seriously enough. We plan to employ the sensitive digital holographic microscopy (DHM) technology for 3D imaging of unstained blood cells collected from a drop of blood as in a peripheral smear. The individualized short-term effect of pollution that this methodology can potentially detect is novel in the area of environmental health. The DHM technology generates 3D cell images using natural refractive index contrast and as a result, changes in 3D morphological features of blood cells can be detected at nano-metric precision with minimal wet lab processing. The methodology requires routine blood smears without any staining. Further the extensive work of the investigator on this topic has made this technology affordable so that it can be employed for addressing a public health problem.**
6. **Outcome: The main outcome of the proposed first-of-its-kind systematic study is to develop understanding of 3D blood cell morphology-based markers for pollution exposure, which will further allow us to establish a generic low-cost method for assessing impact of pollution on an individual's health.**
7. **Budget: INR 50 L, 3 years**

---

9. WP9. Novel AI-based genomic solution for precision oncology in breast cancer

a. Investigators: Dr. Ishaan Gupta (IIT-Delhi), Dr. Atul Batra (AIIMS)

b. Collaborators: Professor SVS Deo (AIIMS), Professor Sandeep Mathur (AIIMS), Dr Daya Nand Sharma (AIIMS), Dr. Pranay Tanvar (AIIMS)

c. Focus Area: Diagnostics, Therapeutics, and Healthcare

d. Goals: To develop novel therapeutic paradigm for triple negative breast cancer treatment

e. Background. Approximately 50,000 women are diagnosed annually with TNBC and around 30,000 lives are lost at a particularly younger age in India. Approximately 29% and 57% of patients with breast cancer present with local and locoregional disease, respectively, while the remaining are diagnosed with metastatic disease. The current standard of care of patients with TNBC is neoadjuvant anthracycline- and taxane- based chemotherapy followed by surgery, and radiotherapy. Around one-third of patients will have pathological complete response in the post neoadjuvant chemotherapy surgical specimen, and this subpopulation has a 3-year overall survival reaching 90%, while those with residual disease fare much worse with recurrence rates reaching 50%. Recently, capecitabine administration in those with residual disease was associated with improved overall survival and is a new standard of care for residual disease. Despite this, patients with TNBC continue to present an extremely challenging entity to treat with short-lived responses to chemotherapy and rapid development of chemoresistance.

f. Approach. We aim to identify the evolution of genomic and transcriptomic landscape with neoadjuvant chemotherapy in residual tumor and in recurrent tumors in a cohort of 300 patients with TNBC in a recently built network (by Dr. Atul Batra) of breast cancer centers across India. Such a dataset currently doesn't exist anywhere in the world. Initial results identified targetable mutations in genes such as ADAMTS8 and suggest that in a larger cohort we will identify the profile of TNBC in Indian patients and the changes in DNA repair pathways (eg BRCA1, BRCA2, RAD1, etc), growth factor pathways (IGF1/2, EGFR, NGF, MET, WNT, etc), immunomodulation (CTLA4, IL12, NK cell, BCR, JAK/STAT, etc), and metabolic (tyrosine, steroid, etc synthesis) to dissect the mechanisms of chemoresistance including acquisition of new mutations and activation/deactivation of these pathways after the administration of neoadjuvant chemotherapy and at relapse. Data from these multi-omic, multi-modal assays will be used to develop a point-of-care diagnostic test coupled with an AI-based analytic platform to stratify patients in a larger validation cohort with targetable interventions.

g. Outcomes. We anticipate that residual disease will reveal mutations in targetable pathways eg PIK3CA/PTEN and inform novel therapeutic interventions, currently not in use to treat breast cancer, such as PARP inhibitors and immunotherapies. We also expect to identify a genetic prognostic signature predict the sensitivity to the above-mentioned drugs in the initial biopsy sample that can significantly improve triple negative breast cancer survival across India and the world. The diagnostic test and the AI-based platform thus developed will be made available at cost across government hospitals across India and the *IP generated will commercialized* for availability elsewhere.

h. Budget: 5 crores, 3 years

---

10. WP10. Medical Imaging and AI for diagnosis, monitoring and prognosis of Cancer

1. Investigators: Prof. Anup Singh (IIT Delhi), Prof. Amit Mehndiratta (IIT Delhi),

2. Co-investigators: Clinical partners from hospitals in NCR Delhi (AIIMS, Fortis, Medanta, etc.)

3. Focus area: Diagnostic imaging for the diagnosis, monitoring and prognosis of Cancer disease.

4. Goals: We propose a research program aimed at developing diagnostic systems based upon medical images and AI and evaluating their clinical applications for Cancer diagnosis, monitoring and prognosis. In addition to the research work, activities on training/workshops of manpower (academic & industry persons) in Medical Imaging and AI shall also be carried out.

5. Approach: Our group has been pursing research work related to diagnostic imaging for the diagnosis, monitoring and prognosis of Cancer disease in collaboration with various clinical partners at Delhi, AIIMS Delhi, Fortis hospital and Medanta hospital. We have made substantial progress and contributions in this area. However, there is still need of more work to further develop the methodology as well as evaluating their clinical potential on large number of patients. Moreover, different types of diagnostic systems shall be developed depending upon the type of cancer. So far, we have been working on brain tumor, breast cancer, prostate cancer, and osteosarcoma. In the proposed research study, medical imaging data from cancer patients shall be acquired from different hospitals in NCR Delhi. Quantitative analysis of medical images shall be performed for extracting multiple parameters related to physiology. Extracted physiological parameters shall be used for diagnosis, monitoring and prognosis of Cancer disease. AI shall be used to develop classifiers/predictive models based upon these physiological parameters for the accurate diagnosis, monitoring and prognosis of Cancer disease.

Our group has been successfully conducting annual workshop "MedImg: Medical Imaging Techniques and Clinical application" since 2015. In last 6 years, approximately 350 participants (from 17 states of India), including faculty, PhD students and MD (radiology) fellows has been trained under this workshop. Detailed content and course workshop of the workshop is available at http://medimg.iitd.ac.in/. The workshop has been conducted jointly by IITD and AIIMS, New Delhi. With the support of the proposed project, we like to continue this workshop/training activities and for this we would also require to purchase some medical imaging accessories.

6. Outcome: Our research group strongly believes in developing open access tools for non-commercial application and to promote research. Some of our research outcome has been already shared at GitHub for open access, our DCE analysis tool and texture analysis tools are shared with clinical partners already for free use (under research agreement) and available for research usage. Diagnostic imaging systems based upon imaging & AI for the diagnosis, monitoring and prognosis of Cancer disease. Trained manpower in medical imaging and AI with focus on Cancer diagnosis, monitoring and prognosis.

- **Budget: INR 749 L, 3 years**

Budget for AI & Medical Imaging Facility for cancer diagnosis, monitoring and prognosis.				
	(All Values are Rs. In lacs)			
Head	**Year 1**	**Year 2**	**Year 3**	**Total**
Capital Equipment	**200**	**50**	**0**	**250**
Consumables	**15**	**20**	**20**	**55**
Manpower	**57**	**61**	**65**	**183**
Travel	**2**	**2**	**2**	**6**
Training	**15**	**15**	**15**	**45**
Workshop	**15**	**20**	**20**	**55**
Contingencies	**10**	**10**	**10**	**30**
Total				**624**
Overheads				**125**
Total + Overheads				**749**
Grand Total (in figures)	**Seven Hundred Forty Nine Lakhs**			

- **PIs research publications:**
  - **Prof. Anup Singh (https://scholar.google.co.in/citations?user=qyd5fCkAAAAJ&hl)**
  - **Prof. Amit Mehndiratta (https://scholar.google.com/citations?user=rqiKH2AAAAAJ&hl)**

---

11. WP11.Project title: Identification of epigenetic signatures in AMR using AI and ML

1. Investigators and collaborating Institutes: IIT Delhi, AIIMS, MAMC, CMC Vellore: Prof. Vivekanandan Perumal (IITD), Prof. Jayadeva (EE), Prof. Ravikrishnan Elangovan (IITD), Prof. Sarita Mohapatra (AIIMS, New Delhi) and Prof. Vikas Manchanda (MAMC, New Delhi) and Prof. V. Balaji (CMC Vellore)

2. Background and objectives: Bacteria have acquired resistance to many of the antibiotics currently in use. Most health-care related funding / policy making agencies across the world including WHO, Welcome Trust, NIH, CDC, ICMR and Ministry of Health, India (cseindia.org/userfiles/inap_amr_20170420.pdf) etc. have recognized antimicrobial resistance (AMR) as one of the most important problems of this century. The availability of bacterial genomics data is increasing at an unprecedented manner, necessitating newer methods of analysis.

In several clinical isolates, genetic changes do not explain AMR. It is also becoming increasingly clear that (a) genetic changes do not always explain antibiotic resistance and (b) a sizable proportion of mismatches between the genotypic profiles and the AST phenotypic profiles (Reference 1). In addition, recent studies suggest a conclusive role for epigenetics in AMR (Reference 2); but epigenetic signatures of AMR remain poorly understood. There are not many labs in India that can perform epigenetic analysis of complete bacterial genomes at a single-base resolution. Our lab is equipped to perform bacterial epigenetics at a single-base resolution. We have recently mapped m6A and 5-mC on whole genomes of MRSA (ATCC) [Please see Figure 1]

Understanding epigenetic signatures in AMR is now a key research area and is expected to aid clinical decision-making and the development of new diagnostic assays. We believe that understanding bacterial epigenetics at single-base resolution in terms of AMR may be best provided by an interdisciplinary team of innovators and clinicians.

3. The approach: Whole-genome sequencing (WGS) has been increasing realized as an important tool for the identification of antimicrobial resistance in bacteria. We propose to perform WGS (primarily focused on the epigenome) on 400 clinical isolates from the ESKAPE group of pathogens known for their antimicrobial resistance. This data will then be systematically catalogued in an electronic, cloud-based database that contains (a) the phenotypic information of the isolate (including AST profile), (b) the whole-genome sequences (genotypic profile) and (c) epigenetic profile at single-base resolution in the form of electronic medical records, including the sample type, clinical presentation, drugs used for treatment and response to antimicrobial therapy. The creation of the proposed database will help understand the genotypic coordinates of antimicrobial resistance. These epigenetic coordinates of antimicrobial resistance we determine here will help minimize the gap between phenotypic and genotypic resistance.

We plan to collect isolates from three large tertiary care hospitals (AIIMS, New Delhi, MAMC Delhi and CMC Vellore) for this study. Methods to assess epigenetics at a single-base resolution in bacterial genomes have evolved and are becoming more and more affordable. Currently, epigenetics signatures of AMR are not established. We will perform WGS of 400 clinical isolates from the ESKAPE group of pathogens and characterize the epigenetic signatures of AMR isolates using artificial intelligence (AI) and machine learning (ML)-based approaches. IITD is already collaborating with AIIMS, MAMC and CMC Vellore in this area. For all the 400 isolates we will have the complete genotypic and phenotypic profiles along with the epigenetic profiles determined using AST data (Vitek), genotypic data (i.e sequence and AMR gene information for the whole genome) and epigenomic profile at single-base resolution (using oxford nanopore).

4. Expected outcomes: The proposed work will create one of its kind bio-repository for 400 isolates from the ESKAPE group of pathogens which will contain (a) phenotypic profile (AST profile) (b) genotypic profile (complete sequence genome information and information of AMR genes) and (c) epigenetic information on whole bacterial genomes at single-base resolution. There is no such repository anywhere in the world currently (the one or two repositories that exist currently do not contain any epigenetic information).

Single-base resolution bacterial epigenetics is a new area of research still in its infancy. We have the necessary equipment and the expertise to perform bacterial epigenetics. We have a single-base resolution map of m6A and 5-mC in the genome of Methicillin resistant *Staphylococcus aureus* (MRSA) from ATCC (Please see Figure 1). In all, there are less than a dozen publications with single-base resolution bacterial epigenetics in AMR; none of these are from India. We have some leads of differences between MRSA and MSSA.

In sum, this work will lay the foundation for a plethora of studies on AMR and epigenetics and the availability of phenotypic, genotypic and epigenome profiles of clinical isolates will help understand yet known mechanisms in AMR and aid in the development of newer diagnostic assays. The multi-disciplinary team assembled for this alliance and the institutions represented herein provide the range and depth of expertise and resources required for end-to-end translation of these ideas to the clinical setting.

References:

1. Neuert S, Nair S, Day MR, et al. Prediction of Phenotypic Antimicrobial Resistance Profiles From Whole Genome Sequences of Non-typhoidal *Salmonella enterica*. *Front Microbiol*. 2018;9:592. Published 2018 Mar 27.

2. Ghosh D, Veeraraghavan B, Elangovan R, Vivekanandan P. Antibiotic Resistance and Epigenetics: More to It than Meets the Eye. *Antimicrob Agents Chemother*. 2020;64(2):e02225-19.

Figure 1: Preliminary data

Figure 1: Mapping of m6A and 5-mC in MRSA (ATCC) at single-base resolution. This was performed by direct sequencing (no PCR amplification) of MRSA harvested from mid-log phase using Grid-ion (oxford nanopore) sequencer. Megalodon and in-house codes were used to map m6A and 5mC information at single base resolution along the full-length genome of MRSA.

Budget:

	Year 1	Year 2	Year 3	Year 4	Total
Capital Expenditure	70 lakhs	30 lakhs	NIL	NIL	100 lakhs
Operating Expenditure	50 lakhs	50 lakhs	35 lakhs	35 lakhs	170 lakhs
Totals	120 lakhs	80 lakhs	35 lakhs	35 lakhs	270 lakhs

---

12. WP12. Development of gigapixel imaging system and automated histopathological analysis system

Investigators: Ravikrishnan Elangovan (DBEB), Kedar Khare (KSBS), Jayadeva (EE), Prabhat Mallik, AIIMS,

Focus Area: Diagnostic assay, imaging, automation

Background: Microscopy evaluation of biological samples remains routine diagnostics methods for many communicable and non-communicable diseases. One of the major challenges in histopathology diagnostics is variability from operator to operator. Combination of fluorescence imaging and traditional white light microscopy has scope to improve the specificity and sensitivity of many existing diagnostics methods. In this project, we aim to develop a giga pixel, multi-mode imaging system to scan histopathology slides in short duration. Image acquired will be processed and classified using deep learning image analysis methods.

Business case: There is a significant untapped market for research products for use by research labs, both in industry and the public sector. The project aims to design value added products that make use of cutting edge AI/ML developed at IIT Delhi and to apply it in the imaging domain. Products envisaged are smart imaging systems that can provide annotation, recognition, and feature recognition driven insight that can accelerate discovery and understanding for research lab personnel.

Objectives:

1. **Development of gigapixel imaging hardware**
2. **Integrated AI-based automation (analysis to recognition, annotation, and eventially, reporting)**

Approach:

Objective 1: Development of gigapixel imaging hardware

Current imaging solutions offer few tens of million pixels per frame of imaging. For a typical scanning of smear slide or histopathology slides, one would need to scan multiple hundreds of images. In currently available automated scanning systems, these images are collected via automation of stage and analyzed for various features. One of imaging limitation of the current system is area of CCD and lens used for magnification. These automated systems are expensive and takes few minutes to hour to completely acquire required number of images. We would like to use micro-lens geometry to have miniaturized collection of emission from sample and collect as much few mm of area in single shot. Group has required optics expertise to design, develop and demonstrate prototypes of these systems.

Modes of imaging: Brightfield, fluorescence modes, multi-spectral

Objective 2: Integrated AI-based automation (analysis to reporting)

a. Data collection and curation: With collaborators in AIIMS and MAMC Delhi, we will be curating sample collection and images. We have recently started an imaging facility in IIT Delhi with high-through put imaging system. We will be able to setup a mechanism for online curation of these images and verified by two human experts. We will target following four disease condition as pilot system demonstration.

- **Sputum smear slides for Tuberculosis diagnosis**
- **Blood smear slides for Malaria diagnosis**
- **Cervical pap smear for cervical cancer diagnosis**
- **Tissue biopsy for Lung cancer diagnosis**

b. AI models for object identification: Learning small models, and with small data: Most clinical datasets are heavily skewed, with only a few labelled samples from ailing subjects. The cost of identifying and labelling subjects is high, and studies are usually limited to a small number of subjects. Our work on learning models with a small VC dimension has led to the Minimal Complexity Machine (MCM) family [1-12]; these models are usually 3x to 10x smaller, and often over 100x smaller than state of the art. In deep learning settings, they are often smaller by factors of 300x or more. Small models tend to generalize better, which is important for noise removal in the context of non-idealities. Minimizing model complexity helps us find the most relevant and minimal set of features, work with little data, and generalize better. The identification of relevant features, and small models are very valuable in interpreting machine learning results. In the case of medical data, this is a crucial advantage. We will also use MCM deep learning methods.

c. Integrated hardware-AI solution: We will demonstrate hardware-AI integrated solution in clinical setting in year 3.

Outcomes:

- **Novel platform gigapixel imaging technology for modernizing microbiology imaging and pathology imaging**
- **High performance AI with integrated hardware for efficient clinical decisions**

Reference: Greenbaum, A., Luo, W., Su, TW. *et al.* Imaging without lenses: achievements and remaining challenges of wide-field on-chip microscopy. *Nat Methods* 9, 889–895 (2012). https://doi.org/10.1038/nmeth.2114

Budget:

Development of gigapixel imaging system and automated histopathological analysis system				
	(All Values are Rs. In lacs)			
Head	**Year 1**	**Year 2**	**Year 3**	**Total**
Capital Equipment	**270[a]**	**50**	**0**	**320**
Consumables	**20**	**20**	**20**	**60**
Manpower	**54**	**54**	**54**	**162**
Travel	**2**	**2**	**2**	**6**
Training	**15**	**15**	**15**	**45**
Contingencies	**10**	**10**	**10**	**30**
Total				**623**
Overheads				**124.6**
Total + Overheads				**747.6**
Grand Total (in figures)	**Seven Hundred Forty seven Lakhs and sixty thousand**			

[a] Spectral confocal imaging system

1. Jayadeva. "Learning a hyperplane classifier by minimizing an exact bound on the VC dimension." Neurocomputing 149 (2015): 683-689.

2. Jayadeva, Sumit Soman and Amit Bhaya. "The MC-ELM: Learning an ELM-like network with minimum VC dimension." 2015 International Joint Conference on Neural Networks (IJCNN). IEEE, 2015.

3. Jayadeva, Suresh Chandra, Sanjit S. Batra, and Siddharth Sabharwal. "Learning a hyperplane regressor through a tight bound on the VC dimension." Neurocomputing 171 (2016): 1610-1616.

4. Pawas Gupta, Sanjit S. Batra, Jayadeva, "Sparse short-term time series forecasting models via minimum model complexity", Neurocomputing, Volume 243, 21 June 2017, Pages 1-11, ISSN 0925-2312.

5. M. Sharma, Jayadeva, S. Soman, H. Pant, "Large-Scale Minimal Complexity Machines Using Explicit Feature Maps," IEEE Transactions on Systems, Man, and Cybernetics: Systems 47.10 (2017): 2653-2662.

6. Jayadeva, Sharma, Mayank, Sumit Soman, and Himanshu Pant. "Ultra-Sparse Classifiers Through Minimizing the VC Dimension in the Empirical Feature Space." Neural Processing Letters (2018): 1-33.

7. Sharma, Mayank, Jayadeva, Sumit Soman, and Himanshu Pant. "Non-Mercer Large Scale Multiclass Least Squares Minimal Complexity Machines." In 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1-8. IEEE, 2018.

8. Soman, Sumit, Jayadeva, Rajat Thakur, Mayank Sharma, and Suresh Chandra. "Sparsity in function and derivative approximation via the empirical feature space." Information Sciences 512 (2020): 402-415.

9. Soman, Sumit and Jayadeva. "Learning from Low Training Data using Classifiers with Derivative Constraints." In Proceedings of the ACM India Joint International Conference on Data Science and Management of Data, pp. 86-93. ACM, 2019.

10. Jayadeva, Soman, Sumit, Himanshu Pant, and Mayank Sharma. "QMCM: Minimizing Vapnik's bound on the VC dimension." Neurocomputing 399 (2020): 352-360.

11. Sharma, Mayank, Sumit Soman, and Jayadeva. "Minimal Complexity Machines Under Weight Quantization." IEEE Transactions on Computers 01 (2021): 1-1.

12. Pant, Himanshu, Sumit Soman, Jayadeva and Amit Bhaya. "Neurodynamical classifiers with low model complexity." Neural Networks 132 (2020): 405-415.

13. Jayadeva, Khemchandani R, Chandra S. Twin support vector machines for pattern classification. IEEE Transactions on pattern analysis and machine intelligence. 2007 Mar 19;29(5):905-10.

14. Jayadeva, Pant H, Sharma M, Soman S. Twin neural networks for the classification of large unbalanced datasets. Neurocomputing. 2019 May 28;343:34-49.

15. Jayadeva, Soman S, Saxena S. EigenSample: A non-iterative technique for adding samples to small datasets. Applied Soft Computing. 2018 Sep 1;70:1064-77.

---

13. WP13: Online Breath analysis for clinical diagnosis using an AI enabled low cost mass spectrometer

Investigators and Collaborators: Prof. Bhaskar Mitra (IIT Delhi), Prof. Jayadeva (IIT Delhi),

Goals: We plan to establish on-line analysis of breath using a low cost mass spectrometer as a tool for clinical diagnosis. Getting insight into a person's metabolic status via on-line analysis of exhaled breath is significant and advantageous, because a wide range of compounds can be detected, because sampling is easy and completely noninvasive, and as opposed to sending samples to a lab, the results are available on the spot. Breath is available in nearly unlimited quantities, its analysis presents no burden to the subject being measured, and its on-line analysis allows continuous monitoring of metabolic health, disease progression, and medication in short time intervals. Moreover, this approach might become indispensable in certain critical situations, e.g., in the emergency room, for correct treatment of a dangerous infection such as pneumonia, when consumption of potentially dangerous party drugs must be ascertained, in doping control, or in other situations where immediate action is necessary.

Approach: In our initial approach, we will focus on diagnosis and management of asthama and COPD. Currently, the biomarker FeNO used in clinical routines for diagnosis and management of asthama. A recent review including 175 studies evaluated the diagnostic value and the clinical utility of FeNO for the management of asthma in adults.The diagnostic value is influenced by atopy, asthma-therapy, age, and smoking habits. It may be remarked that traditional mass spectrometry data was very effective in identifying Covid; this may be explored as well.

COPD is a highly prevalent lung disease which is the fourth leading cause of death and is expected to increase to the third leading cause of death in the world by 2030. As it is primarily a disease of the lungs, a high number of breath-analysis studies are available for COPD; however, only a minority of the studies follow on-line methods. We plan to use online analysis of breath to identify biomarkers for COPD and a monitoring protocol using our low cost mass spec system with AI/ML for pattern recognition.

Online breath analysis requires a sensitive tool to be effective. Our research group has developed a low-cost 3D printed quadroupole mass spectrometer that is suitable for detection of VOC's and low molecular weight compounds. For online breath analysis we plan to use the device as a Proton transfer reaction – Mass spectrometer (PTR-MS), which uses ionized water vapor as the ionization source.

To utilize MS as a clinical diagnostic technique, it requires the use of AI/ML. Manual analysis of spectra can be cumbersome, especially in cases of mass spectrometers operating in mTorr regime. This is due to the large number of chemical species present in the ambient, as well as fragmentation of the target molecule itself. Our approach will be to create training data from known cases and then use the AI/ML system to identify patterns which can help detect and quantify disease progression.

	Top: Design and Fabrication of the 3D printed mass spec. Figure on the right shows they typical trajectory of a trapped ion. Left: Mass spec in action with a hollow cathode plasma discharge.

Outcomes.

1. **AI enabled low cost 3D printed mass spectrometer for online breath analysis**

2. Detection and monitoring protocol for Asthma and COPD using the system

Budget

	Year 1	Year 2	Year 3	Total
Capital Expenditure	100 lakhs	NIL	NIL	100 lakhs
Operating Expenditure	22 lakhs	30 lakhs	30 lakhs	82 lakhs
Totals	122 lakhs	30 lakhs	30 lakhs	182 lakhs

The Role of AI/ML in the project

The miniaturization of the mass spectrometer introduces new challenges. The effects of non-idealities in fabrication, and deviation in particle motion from an idealized situation are among many factors that can lead to errors. We propose to use pattern recognition and machine learning to reduce the error rate and improve identification of sample constituents. In a real world setting, the pattern corresponding to one ailment may be masked or confused by other prevailing conditions. Therefore, we need to use machine learning to improve recognition in the presence of noise.

At EE, IIT Delhi, we have pioneered three novel technologies that are of relevance here. The references cited by number are listed under project 11.

1. Learning small models, and with small data: Most clinical datasets are heavily skewed, with only a few labelled samples from ailing subjects. The cost of identifying and labelling subjects is high, and studies are usually limited to a small number of subjects. Our work on learning models with a small VC dimension has led to the Minimal Complexity Machine family[1-12]; these models are usually 3x to 10x smaller, and often over 100x smaller than state of the art. In deep learning settings, they are often smaller by factors of 300x or more. Small models tend to generalize better, which is important for noise removal in the context of non-idealities.

2. Learning from unbalanced data: In clinical settings, the number of samples from subjects with an ailment or condition is usually much smaller than samples from normal subjects. Often, this skew is handled by using an equal sized control group, but this distorts the distribution of samples and is not the best approach in a machine learning setting. Two methods developed by us include the Twin SVM [13] - that has nearly 1200 citations and is widely used to learn from unbalanced data, as well as the Twin Neural Network [14]; these are not sensitive to the relative sizes of individual classes.

3. Augmenting Small Datasets: Learning outcomes can be improved by adding samples to small datasets. This is particularly useful for very small datasets. Our technique, Eigensample [15] generates novel samples that perturb the original distribution the least; in fact, the first few principal components are unchanged. This is superior to techniques like GANs in that the novel samples generated are nearly noise free, and the labels of new samples are also available.

---

14. WP14: Indigenous affordable Spinal Epidural Electrical Stimulator (sEES) with AI/ML in the loop, for motor recovery in spinal cord injury (SCI)

Soutik Betal[1], Jayadeva[1], Shaunak Sen[1], Bhaskar Mitra[1], Dhiman Mallick[1] and H. S. Chhabra[2]

1. **Department of Electrical Engineering, IIT Delhi, New Delhi 110016**
2. **Indian Spinal Injuries Centre, Vasant Kunj, New Delhi 110070**

Abstract

Developing a low cost and improved design of epidural stimulator is the need of the hour especially in emerging nations to promote research for development of an affordable treatment to regain motor function after SCI. There is a significant market for research products in this developing field, for use by researchers and clinicians engaged in developing protocols and treatment modalities. The cost of individual implants runs into several lakhs, even for limited trials or experimental investigations that physicians would like to conduct. This effort aims at filling this gap and accelerating development in this domain. We aim to work towards value added products that can use sensor data in a smart AI/ML driven loop, to help physicians determine optimal treatment protocols.

Application Summary:

Spinal cord injury (SCI) is often a catastrophic condition requiring chronic care. It can cause permanent loss of sensation and motor function below the level of injury, thus causing severe physical disability and psychological distress. Standing and walking are highly desirable goals of SCI individuals[1]. However, despite an enormous amount of research in SCI, the neurological prognosis for individuals with complete and severe SCI remains dismal. Activity-based locomotor training, the only practice used to enhance recovery at present, does not yield significant outcomes in individual with severe SCI[2–4]. Spinal Epidural Electrical Stimulation (sEES) has been shown to produce rhythmic motor activity patterns in spinalised rats[5] as well as humans with complete SCI[6]. Voluntary control of flexion and extension leg movements has been noticed in two individuals with complete SCI after sEES[6,7]. Subsequently, independent stepping was noticed in an SCI individual after task-based training with sEES[7]. More recently, intentional over ground walking has been noticed in two American Spinal Injury Association Impairment Scale B (AIS-B) SCI individuals after sEES with locomotor training[8]. An unexpected voluntary motor control of lower limbs and independent standing without sEES was also reported in one of these patients[8]. These studies suggest that the spared structural connections at the site of injury which were functionally inactive, may be capable of modulating the excitability of the sensori-motor networks below the level of injury[6]. Also, the plasticity in the local spinal circuits and/or the descending pathways with axonal sprouting was proposed as a hypothesis for the voluntary control of leg movements observed without sEES.

From https://academic.oup.com/brain/article/144/2/420/6046571

sEES is like a pacemaker for spinal cord stimulation for various applications like lower back pain relief, locomotion of paralysed lower limbs, etc. The principle of sEES[9] is that low power stimulatory pulses/waveforms are converted into high power signals that are applied to neuromuscular tissues of patients with SCI. As a result these signals generate electrical impulses which artificially evoke action potentials in the desired nerve fibres (APs) that are responsible for triggering muscular contractions. . sEES consists of the stimulator and the electrodes as its two main components. The instructions from the patient or physician will be fed to an embedded system software application which it will translate into a series of parameters (pulse shape, activation duration, stimulating frequency, amplitude, inter-pulse duration)[10]. Using these parameters, the pulse generator or battery, synthesizes pulse-like waveforms or signals. These waveforms can be implemented by digital signal processing or microcontroller-based systems[11,12]. The design of the sEES should be such that it obtains high time resolution, adaptive pulse generation based on positioning/posture of the patient and also achieve more flexible closed loop controller feedback system[13].

The Indian Spinal Injuries Centre (ISIC) has performed a prospective pilot clinical study on SCI patients to establish the role of sEES in improving the motor abilities and functional status of SCI individuals and explore the role of sEES in neurological recovery in such individuals.The outcomes of this pilot study are encouraging. Our presentations on the work were awarded the best paper awards at the Korean Spine Surgeons Society and the Association of Spine Surgeons of India's annual conferences. However, the cost of the implant used for epidural stimulation is prohibitive. We strongly believe that an affordable design of epidural stimulators could help the individuals with severe SCI lead a better quality of life. We also feel that the design could be improved through developing smarter ways of stimulation. We hence, aim to develop a low cost and improved design of epidural stimulation. This will serve as an affordable opportunity for a better quality of life for all those SCI patients or lower limb paralysed individuals, particularly in the emerging nations like India. The proposed innovation has a great scaling potential as the demand of a low cost and improved spinal cord stimulator is expected to be substantial not only in a densely populated country like India but also globally. This is especially important since spinal cord injury generally affects the lower social economic strata of the society, has a substantial life long socio-economic impact on not only the individual but the whole family and most persons with spinal injury need to be suitably vocationally rehabilitated. Our team leading this innovation will comprise of experts on SCI in clinical aspects from Indian Spinal Injuries Centre (host institute) and technical aspects from IIT Delhi, both being premiere institutes of the country.

Business case:

The work will involve MEMS design of of electrode array implants, and a sensor array. Data from sensors is planned to be used in developing a AI/ML predictive model that will help learn the relationship between optimal stimulus patterns and regeneration, and help accelerate the development of customized treatment protocols for individual patients. The value addition is at the hardware, software, protocol development and treatment levels.

Non-Fiscal Support Needed

Support from the public and industry funding partners for facilitating scale-ups through stakeholder meetings, facilitating access to Business APIs

Intellectual Property conditions

Each of the work packages has a clear IP potential for commercialization. IP will be managed on a work-package basis the proposal stakeholders are agreed on.