Project Title: Data-driven Precision Medicine

Point of Contact: Purushotham V. Bangalore, Professor, Computer Science
Email: puri@uab.edu. Phone: 205 934-8604.

Project Synopsis:

The advances in science and engineering domains have also resulted in datasets generated at a much faster rate than any previous time in history. Not only has the volume of data increased in the recent years, but also modern scientific and engineering applications tend to use data from diverse sources to undertake multidisciplinary research. This data deluge is often characterized as the 5Vs (Volume, Velocity, Variety, Veracity, and Value) of "big data." In order to address the challenges posed by "big data" and enable data-driven discovery, especially in the emerging fields such as quantitative biology and precision medicine, it is imperative for future researchers and practitioners to have transferable interdisciplinary training and research experience in both data and life sciences. Particularly, in the area of precision medicine, it is projected that the volume of data is going to explode both in terms of data collected per individual as well as across demographic communities. For example, with all the genomic, proteomic, imaging, diagnostic, and other types of health and medical-related data could be in the order of multi-terabytes for an individual and could potential grow even larger over time as the resolution, type of data, and advances in technology such as smart sensors and IoT devices.

These advances will pose the following challenges in collecting, organizing, preserving, securing, analyzing, and decision-making to assist with detection, prevention, and treatment of various health-related issues:

- Data Collection: With the advent of ubiquitous smart devices that can be worn as well as part of the living environment (e.g., IoT devices) data is now being collected from multiple sources. There is a significant challenge in collecting and storing all these data in an interoperable format along with the necessary provenance.
- Data Organization: With the variety of data being collected, organization of these data to facilitate easier access and processing while preserving the privacy and confidentiality of the data will be a key challenge to overcome.
- Data Preservation: Considering that this data may have to be stored for the entire life-time of an individual (and may be beyond) how to preserve the data for such a long-time with ever changing technology and regulations is in itself a challenge that has not received much attention.
- Security and Privacy: Not only does this data needs to secured for compliance purposes but also privacy issues in terms of data-sharing and delegation with trusted parties as well as different types of health-care providers (from doctors to insurers to employers and relatives).
- Analysis: While individual datasets are in the order of terabytes (and growing), when performing any kind of demographic-wide analysis, this process could require enormous amount of computing power and novel machine learning algorithms.
- Decision-making: Today most of the final health-related decisions are done by the health professional in consultation with the patient and their relatives. However, with the growth of machine-learning and deep-learning, computer algorithms could assist with a more comprehensive diagnosis and allow the patient to explore alternative treatment options.

- Telemedicine: With the explosion of connected devices and a "flat-world" now healthcare could reach remote areas of the population through the use of technology and provide similar services remotely to people in rural or remote areas where they may not have in-person access to health services.

The proposed research is in alignment with state-wide initiatives such as the Alabama Genomic Health Initiative (AGHI) that is trying to sequence a large population in Alabama to understand the role genes play in the prevention and treatment of diseases. If the AGHI program were to be scaled to the entire state of Alabama or beyond, then all of the above discussed challenges have to addressed in order to provide a feasible solution. The scientific methodology and techniques developed as part of this effort could also be easily extended and applied to any other "big-data" problem such as autonomous transportation, smart cities, smart agriculture and food production, and advanced material science research.

The proposed research brings together faculty from diverse computational, engineering, and life science disciplines to work on the above-mentioned research questions. This will enable us to educate and train next generation scientists and engineers with advanced training and practical research experience needed to pursue a career in the emerging data-centric disciplines. This program will provide students with unique and innovative training and hands-on research experience not only in computer and data sciences including data mining, machine learning, and visualization, but also domain-specific expertise in an application area. One of the key foci of this program is to recruit and train a diverse student body, and long-term commitments by individual faculty involved on this proposal to training, mentoring, and recruiting women and underrepresented groups into science and engineering.

UAB is the perfect position to undertake this challenge and design a feasible solution as it has the largest life science research in the state along with the computational and big data analysis capabilities that have been built over the last several years. UAB also has the advantage of having access to the ever-growing volume and variety of data. The team members have expertise in the areas of precision medicine, high performance computing, cloud computing, IoT, big data analytics, security and privacy, large-scale databases, data mining and machine learning, and program languages.

List of potential team members:

Purushotham V. Bangalore (Professor, Computer Science)
Matthew Might (Professor, Medicine and Computer Science, Director PMI)
Jeremy Blackburn (Assistant Professor, Computer Science)
William E. Byrd (Scientist III, Computer Science and PMI)
Ragib Hasan (Associate Professor, Computer Science)
Peter Pirkelbauer (Assistant Professor, Computer Science)
Nitesh Saxena (Associate Professor, Computer Science)
Da Yan (Assistant Professor, Computer Science)
Chengcui Zhang (Professor, Computer Science)
Yuliang Zheng (Professor and Chair, Computer Science)
Others to be added at a later date.