Module 1: Data Science for Healthcare

1.1: Introduction, Healthcare Data Sources and Data Analytics for Healthcare, Applications and Practical Systems for Healthcare.

1.2: Electronic Health Records(EHR), Components of EHR, Benefits of EHR, Barriers to Adopting EHR, Challenges of using EHR data, Phenotyping Algorithms

Data Revolution in Healthcare

Potential of Healthcare → $ 300 Billions.

Importance of Data

In the digital age, healthcare data plays a vital role in transforming the industry. The collection of various healthcare data types is made easier due to technological advancements.

Potential of Data

Recent reports indicate that the potential worth of healthcare data is estimated to be around $300 billion. This highlights the significant value and impact of data in the healthcare sector.

Data Collection

Hospitals and healthcare institutions are increasingly collecting vast amounts of patient data. This data collection trend is driven by advancements in data sensing and acquisition technologies.

Challenges and Opportunities

Analyzing Healthcare Data

The sheer volume and heterogeneity of healthcare data pose significant challenges in processing and analysis. Advanced analytical techniques are required to make sense of this data efficiently.

Proactive Healthcare Delivery

Data analytics presents the opportunity to shift healthcare delivery from reactive to proactive. By leveraging insights from health data, clinicians can predict and prevent potential medical complications.

Integrated Data Analysis

Integrated analysis of diverse healthcare data sources is unlocking new insights and capabilities. This approach enables a comprehensive understanding of patient profiles and medical trends.

Nature of Interdisciplinary Challenges

Diverse Disciplines

Healthcare advances come from databases, data mining, and medical research. Challenges arise due to varied backgrounds and expertise.

Skills Gap

Computer scientists lack medical knowledge, while practitioners may lack statistical skills. Coherence in data analysis is hindered by this disparity.

Independent Work

Different perspectives lead to isolated efforts without practical applications. Real domain-specific problems are overshadowed by technical solutions.

Importance of Bridging Communities

Collaborative Integration

Bridging disciplines is essential for utilizing advanced analysis techniques effectively. The fusion of expertise unlocks the vast potential of healthcare data.

Advancing Data Analysis

Aligning diverse communities enables a comprehensive discussion of relevant contributions, fostering innovation in healthcare data analysis.

Data Privacy Gap in Healthcare Domain

Data Privacy Gap in Healthcare Domain

Sensitive Nature of Healthcare Data

Healthcare data is highly sensitive, containing personal and compromising information about individuals. Sharing this data poses significant risks to privacy and confidentiality.

Challenges in Sharing Information

The gap between medical researchers and computer scientists in sharing healthcare data complicates efforts towards collaboration and innovation. Bridging this divide is essential for advancements in healthcare technology.

Access to Healthcare Data

Medical researchers naturally have access to healthcare data due to their work, raising concerns about maintaining data privacy and compliance with legal regulations.

Laws and Safeguards

Health Insurance Portability and Accountability Act (HIPAA)

HIPAA in the United States prohibits the release of medical information without safeguards to preserve privacy. Compliance is crucial for protecting individuals' medical data.

Importance of Safeguards

Safeguards play a vital role in protecting individuals' medical information from unauthorized access or disclosure. They ensure compliance with data privacy laws and ethical standards.

8

Healthcare data sources and data analytics for healthcare

This part discusses the details of various healthcare data sources and the basic analytical methods that are widely used in the processing and analysis of such data. The various forms of patient data that is currently being collected in both clinical and non-clinical environments will be studied. The clinical data will have the structured electronic health records and biomedical images. Sensor data has been receiving a lot attention recently. Techniques for mining sensor data and biomedical signal analysis will be presented. Personalized medicine has gained a lot of importance due to the advancements in genomic data. Genomic data analysis involves several statistical techniques. These will also be elaborated. Patients' in-hospital clinical data will also include a lot of unstructured data in the form of clinical notes. In addition, the domain knowledge that can be extracted by mining the biomedical literature, will also be discussed. The fundamental data mining, machine learning, information retrieval, and natural language processing techniques for processing these data types will be extensively discussed. Finally, behavioral data captured through social media will also be discussed.

Types of Data sources:-

① Electronic Health record (EHR):-

Types of EHR: full range of data relevant to a patient's care such as demographics, problems, medications, physician's observations, vital signs, medical history, laboratory data, radiology reports, progress notes, and billing data
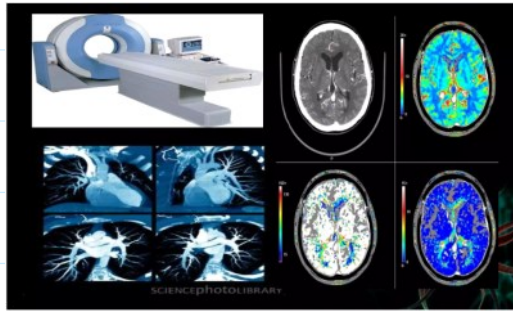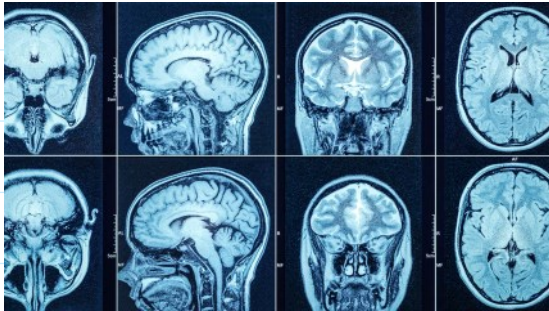
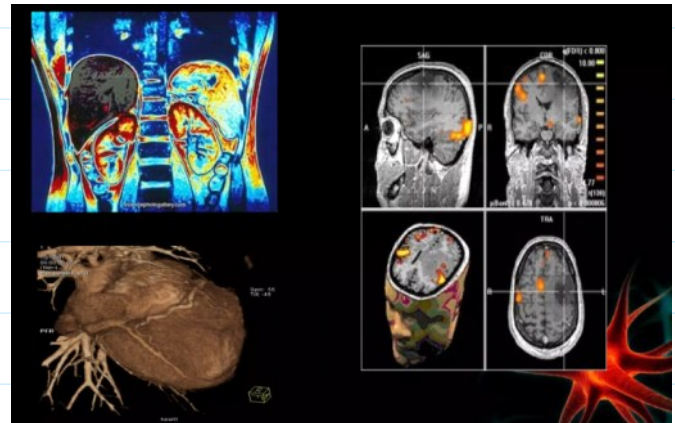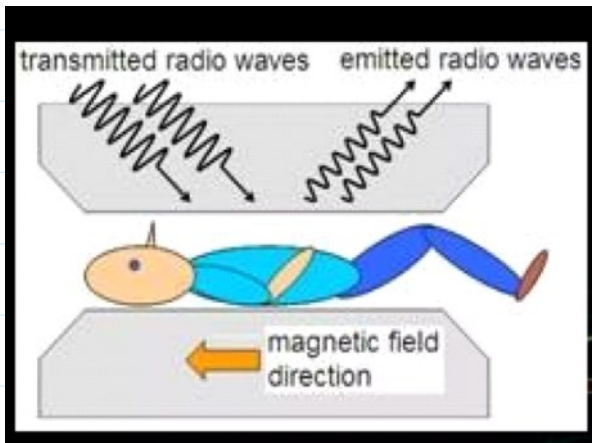## Biomedical Image Analysis

categories of medical Image analysis → Anatomy
→ Functions.

Radiological Imaging.

ultrasound Images.

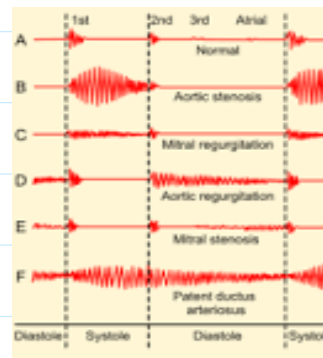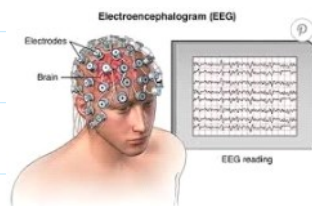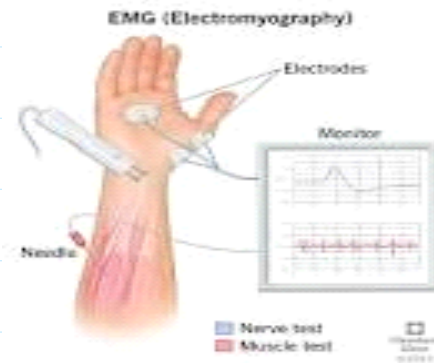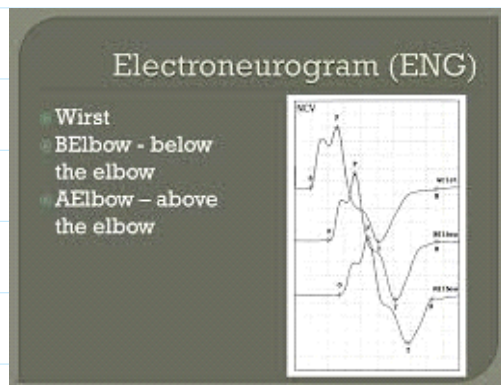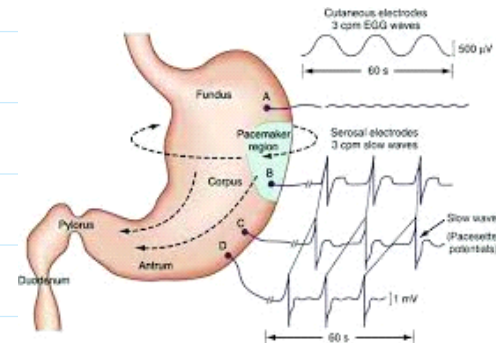CT scan

MRI images

PET scan images
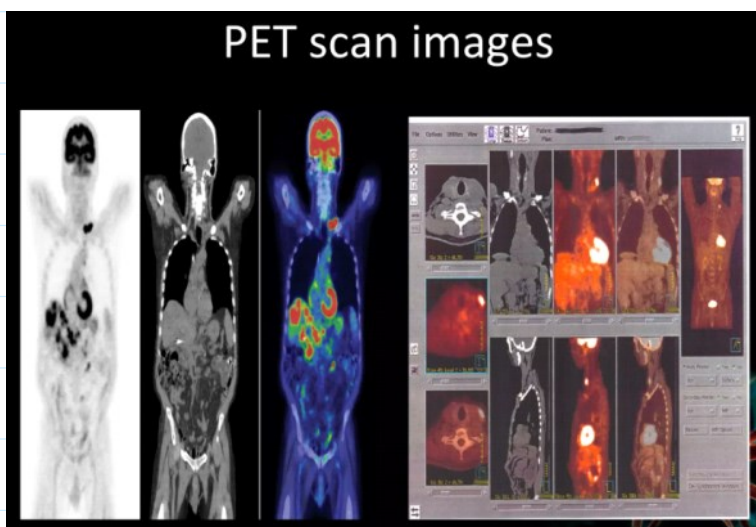
Examples of such signals include the electroneurogram (ENG), electromyogram (EMG), electrocardiogram (ECG), electroencephalogram (EEG), electrogastrogram(EGG), phonocardiogram(PCG),

Phenocardiogram

Genomic Data Analysis

Clinical Text Mining