# From Data to Cure: Leveraging Artificial Intelligence and Big Data Analytics in Accelerating Disease Research and Treatment Development

**Chaitran Chakilam[1]*, Sambasiva Rao Suura[2], Hara Krishna Reddy Koppolu[3], Mahesh Recharla[4],**

[1]*Validation Engineer, chaitrann.chakilam@gmail.com, ORCID ID: 0009-0008-3625-4754
[2]Sr Integration Developer, suurasambasivarao@gmail.com, ORCID ID: 0009-0006-9917-6648
[3]Data Engineering Lead, harakrishnareddyk@gmail.com, ORCID ID: 0009-0004-9130-1470
[4]Oracle EBS Onsite Lead, recharlamaheshh@gmail.com, ORCID ID: 0009-0008-3636-4320

***Corresponding Author:** Chaitran Chakilam
*Validation Engineer, chaitrann.chakilam@gmail.com, ORCID ID: 0009-0008-3625-4754

**Abstract**
The emergence of novel genotyping, big data, data sharing, open-source algorithms, and powerful computing resources has facilitated the integration of artificial intelligence (AI) and big data to elucidate disease mechanisms, find implements for drug design and development, and enhance precision medicine. The parallel explosion of published studies containing large quantities of biomolecular data with accompanying clinical or pathology information has indeed benefited disease studies. However, similar to the biases that plagued many of the datasets used for the development of commercial AI algorithms, diagnosing radiological images, biological oddities or data imbalances could mislead the outcome of AI and big data studies. There is a high need of standardization and heterogeneous curation practices in the input data for AI and big data prometheus. Therefore, a systematic overview of the key technological advances, challenges, and emerging solutions for data mining is warranted for benchmarking and motivating further development of AI and big-data-guided disease research.
As the window of opportunity for the rapid advance of artificial intelligence (AI) solutions to emergent challenges in drug discovery and global health narrows, there is an immediate need to harness big data, AI, and shared research infrastructure to optimize the performance of AI systems on small data sets in biomedicine. The coordination of such systems is a massive and complex engineering challenge that must safeguard researcher freedom while preventing misuse and enhancing collaboration. Effective governance of AI in the life sciences and biomedicine requires specifying requirements and mechanisms to monitor the evolution of each constructed system with respect to agreed-upon principles.

**Keywords:** Artificial Intelligence (AI), Big Data Analytics, Disease Research, Treatment Development, Precision Medicine, Predictive Modeling, Biomedical Data, Machine Learning, Genomic Analysis, Drug Discovery, Healthcare Innovation, Data-Driven Medicine, Clinical Decision Support, Translational Medicine, Real-World Evidence.

## 1. Introduction

In this age of rapid digitalization, data, both from traditional sources (omics, clinical notes, demographic data, etc.) and novel sources (wearables, sensors, etc.), are being collected at an unprecedented rate. Though first developed in the field of astrophysics, the term "big data" is frequently invoked regarding the medical domain. Big data is generally characterized by its volume (size), velocity (speed), variety (diversity), veracity (trustworthiness), and value (usefulness). Usually, only the first and last characteristics are discussed in biomedicine. The advent of big-data processing has accelerated the generation and information extraction of many kinds of data, particularly unstructured data that flesh out a rather incomplete but holistic view of health and disease. For example, in one study, the amplitude of 090813 on Hubble Space Telescope's wavelength spectrum 0.1-1.15 μm revealed large amounts of carbonaceous materials, the temperature of about 1300 K, and a very low geometric optical depth for a flat and a much smaller scale height. These conclusions apply to an entire class of objects.
Big data have been used to identify new diseases and define and standardize disease categories. One study analyzed the electronic health records of 37 million people to identify the top 200 disease categories.

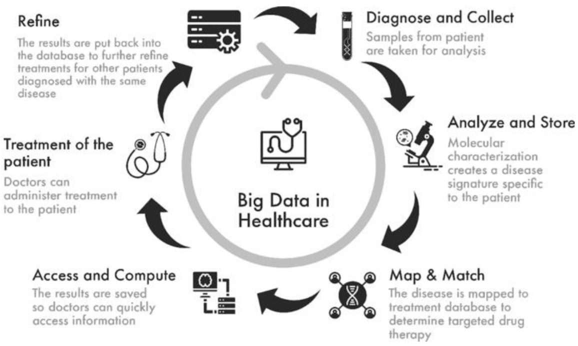


**Fig 1: Big Data Analytics in Accelerating Disease Research**

A combination of pathology-specific coding schemes was used to aggregate nearly 14 million pathology reports, and a natural language processing pipeline categorized and standardized the disease text in these reports to define a large collection of cancer status biomarkers. Other studies digitized and standardized records of drug use in 470,000 patients across multiple medical centers and defined over 15,000 new diseases to facilitate data sharing. AI tools are also being used to define and standardize digital pathology biomarkers. Non-COVID-19 infectious diseases have been detected based on patient records, drug use, and radiology reports, and surveillance networks have been established using data sharing from hospitals and drug prescription services. AI tools have suggested model alterations by discovering biomarker-clinical outcome associations in large gene expression and clinical outcome datasets on Glioblastoma patients.

### 1.1. Background And Significance

Artificial intelligence (AI) and big data applications hold great potential for advancing basic and translational biomedical research, in many respects, it is a currently emerging field. Data in translational and public health uses of AI may derive from electronic health records (EHR), clinical trials, academic medical centre bioinformatics, and population-level epidemiologic studies. Artificial intelligence (AI) – the ability to perform computations and analyses that would normally require human intelligence – will advance human knowledge faster than any other individual scientific discovery. Machine learning (ML) and deep learning (DL) are presently the most widely used AI techniques, where in ML, predictions are generated by data-driven mathematical models, and in DL, predictions are made with deep neural networks that automatically learn representations from data. The practical application of AI technologies is contingent on the availability of data, and big data encompasses all sorts of datasets that are capable of testing AI architectures.

**Equ 1 : Disease Insight Generation**

$$\mathrm{DI} = f(\mathrm{BD}, \mathrm{ML}, \mathrm{BioK})$$

- **DI**: Disease Insights
- **BD**: Big Data (e.g., genomic data, EHRs, imaging)
- **ML**: Machine Learning models
- **BioK**: Biological Knowledge (e.g., known pathways, ontologies)
- **f(...)**: Represents the complex integration of data, models, and expert knowledge

## 2. Understanding Artificial Intelligence in Healthcare

Over the next few years, a substantial boost in the efficiency of healthcare delivery worldwide is anticipated. This boost is expected to be driven by the rapid development of artificial intelligence (AI) techniques and big data applications, which enhance the quantitative and qualitative healthcare delivery. One particular area of application of big data analysis, which is prominent in the biomedical sciences, is disease research, where enormous amounts of information are generated and collected around a variety of experimental systems. This volume of information includes, but is not limited to, patient diagnostics, clinical outcome information, as well as molecular profiles of the biological material collected, in terms of sequencing, imaging data and high-dimensional multiparameter flow/cyto/metabolo-omics cytometry data.

However, the broad application of big data-based healthcare where patient and molecular information is routinely collected and then analyzed to better assess disease risks for efficient prevention, early detection and targeted therapy has not yet materialized. This is partly because medical care and treatment have traditionally focused on disciplines and organs. The understanding of disease occurrence and progression based on the abnormality of a single gene/protein, or mutation/deletion of a chromosomal region, or abnormality of one imaging marker on a certain type of digitized image is still the goal of most healthcare researchers and clinicians alike. Under this workflow of disease research and clinical practice, researchers are focused on the possible influence of current advancements in healthcare AI methods such as machine learning, deep learning and natural language processing on the future of medicine, especially targeting its diagnosis, prognosis and treatment.

In a survey of experienced medical doctors from various countries, only when disease-associated physical alterations of organs or images exist do they believe AI-based diagnostic systems can be beneficial aids to clinical care. They are skeptical of AI systems suggesting false positives in their clinical practice. Framing the task as disease etiology discovery, several recent and concomitant developments in healthcare big data generation and analysis methods that could potentially provide clues for interpreting the randomness of diseases at the systems level are introduced. The proposal is that the randomness of disease occurrence and progression describes the occurrence of some absent or aberrant biological elements or processes under an initialized condition. Thereby, the utilization of healthcare mechanisms poses new challenges in AI and applications.

**2.1. Definition and Scope**

Computer science has advanced to the point where computer vision systems based on artificial intelligence (AI) algorithms can be trained on annotated images to produce automatic predictions for unseen images. If certain conditions are satisfied, these algorithms have been shown to outperform human experts in image-based diagnostics of clinical significance. However, the percentage of digital pathology study groups employing deep neural networks (DNNs), a type of AI, is relatively small. To facilitate the scientific community's participation in this technological shift, a comprehensive overview of various DL frameworks of increasing complexity is provided, suggesting good practices for implementing digital pathology AI algorithms for robustness, generalizability, and reliability.

Biomedical research is increasingly characterized by rapid advances in digital imaging technology, largely attributable to the democratization of imaging platforms and imaging modalities. More sophisticated and affordable microimaging equipment has led to a surge in small animal imaging data. At the same time, rapid advancements in whole slide imaging (WSI) techniques are transforming spatially resolved histopathology data generation in preclinical and clinical investigations. Disparate sources of biological data have become accessible due to advances in bioinformatics pipelines used for genomics, transcriptomics, proteomics, and metabolomics profiling. Powerful image analysis and processing tools and methods for data mining, classification, clustering, and population genomics have been developed to tackle the new challenges posed by large imaging datasets and their heterogeneity.

In spite of the rapid generation of big data, the biomedical and clinical research fields are still shifting from small data to the big data era, as there are challenges in the acquisition and sharing of imaging data and omics data. Specifically, the acquisition of sufficient imaging and genomic data has been limited by costs, equipment availability, and institutional regulations. Even if there is sufficient data, concerns exist over compliance with ethical requirements for sharing medical data and protecting patient privacy, which has hindered access to data. Moreover, even for data generated in similar experiments, separate image datasets are generated due to heterogeneity in equipment and protocols. For instance, image scanners and criteria for tumor typing can differ among hospitals conducting the same preclinical study. Such a gap gives rise to a domain shift between source data and test data, leading to a decline in performance of most AI predictors trained in the source data.

**2.2. Historical Context**

Advancements in data and sample generation technologies within healthcare systems have grown exponentially. This begets an increasing volume of 'big data,' surfacing complex multi-omic patterns that yield integrative and novel biological insights yet are hidden in informatic noise. Many questions remain unfathomable until an innovative means to manipulate the large data sets arises. Conclusively, as big data accumulates, there is an increasing demand for robust methodologies to extract signal, model complexity, and develop informative prior knowledge of the biological systems of interest. While the promise of working with big data in chronic disease research appeals, the challenges are nontrivial. For instance, the relevance of big data biobank and AI-based drug repurposing theory has been identified, issues have arisen about the safety of reusing datasets to predict drug effects or adverse reactions on novel strains of microorganisms. Therefore, detailed consideration is warranted for appropriate cautionary measures and technical solutions based on sound computational platforms.

Such heterogeneous knowledge bases must be curated, integrated, and connected to yield insights that are more informative than those obtainable from any one source. Alternatively, complex models need to be developed that embrace their multiplicity and heterogeneity to better characterize, and ultimately understand, biological systems whose contributions span multiple levels and types of organization. By 'complex models' is meant statistical and mechanistic models that encodes known mechanistic principles. For example, much is known about the ways that molecular properties of chemical compounds explain their pharmacological action. Systems pharmacology is now employing such known principles to build quantitative models that yield interpretable and predictive knowledge about complicated pharmacological problems, such as drug repurposing for COVID-19. Many other knowledge domains offer similar opportunities for reductionist modelling using ontologies and databases containing established knowledge about biological, medical, and environmental resources and their interactions.

Many data-driven technologies in Big Data and Artificial Intelligence (AI) could lead to developments in these directions. Complex models and carefully curated knowledge bases arising from community efforts could be complementary assets. AI data-hungry data-driven technologies developed around artificial neural networks will not harness the complexity and heterogeneity of the knowledge posed out of reach of conventional analytical methodologies. Evidence-based mathematical models rely on known laws and principles. Populations of differential or difference equations represent one intrinsic complexity of biological signals. Agent-Based Modelling and Environmental Modelling are examples of how mechanistic representations of the underlying knowledge can connect complex systems spanning heterogeneous domains/modeling scales. AI has been explored as a complement to these methods.

## 3. Big Data Analytics in Medical Research

Healthcare institutions face considerable challenges in implementing medical big data innovations. How to manage and store large multimodal data is a crucial hurdle. Electronic Medical Records (EMR) collected during hospital visits can be considered a rich resource of highly heterogeneous data types, including documents, images, and structured data. Two main types of EMR big data management systems: 1) a centralized system that collects the EMR data in one platform, and 2) federated systems that maintain the EMR data locally where it is produced, yet share a common data query service. Therefore, there are great challenges in helping clinical professionals take full advantage of the EMR big data.

There are two key complications in analyzing EMR data. First, the EMR data are heterogeneous. Different data types require different data processing tools and analytics. Second, the EMR data are rarely structured. How to extract knowledge from the unstructured and semi-structured data types is one of the key complications in data analysis. For example, the unstructured data sources include clinical texts such as discharge notes and clinical notes; the SEMR data sources are stored in key-value pairs and short time series.

The main challenge of EMR data processing is in standardization and establishing frameworks. Various schema technologies have been proposed to enhance the only closed-format and closed-in clauses suffering from the idea of XML processing to handle heterogeneous EMR data. Existing technologies for heterogeneous EMR data processing can be classified into three categories: (i) Syntax process and data quality control (ii) Structure standardization. The missing data impacts patient treatment practice and the understanding of diseases.

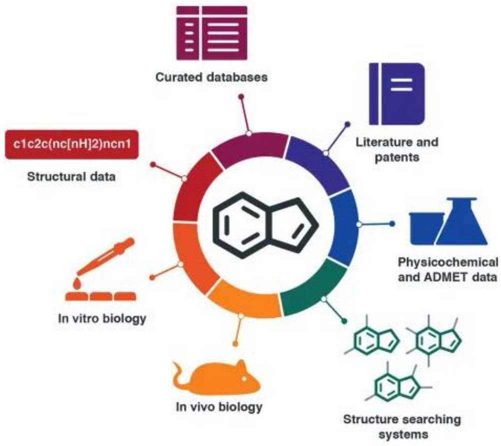


**Fig 2: Role of Big Data in Developing New Medicines**

### 3.1. Data Sources and Types

Applying AI Technology to Disease Research.

Leveraging AI technology has become an important aspect of disease research. However, with the rapid development of AI fields, the published findings may involve different AI algorithms and different use of AI algorithms, making systematic analysis difficult. Thus, subfields of disease research where AI components are applied are summarized, emphasising the setting of AI in these areas of research.

3.1. Data Sources and Types

The amount and quality of data greatly impact the performance and functionality of prediction algorithms and the generalization ability of AI technologies. In disease research, an increasing number of genomics or disease-related big data are being generated and are openly accessible. Among them, available datasets, structural variant revelation, somatic mutation, copy number variation, mutation data, gene expression data and histological image data. Data types of genome and transcriptome in cancer have been profiled. The utilization of multi-omics big data in cancer research has led to new findings regarding consensus, cross-talks, collective influence, diversity and molecular networks in oncogenic processes across cancer types. In heterogeneous mono-disease entities, large-scale multi-omics data types, new methods, AI technologies and their advances, key findings, bottlenecks and challenges regarding the application of multiple pairwise data types, as well as the current status and future perspective regarding the application of multi-omics in both basic and translational research.

Exploits of machine learning metrology, data sources and priors in predicting. Regarding data source, in addition to general databases, the packs of report data come from a provincial hospital. One data pack targets worn cardiovascular episodes and contains 285554. RNA-sequence data come from a database. Thirteen pre-disqualified and 20 and 33 serums are from 10 patients with head and neck cancer. Blood lipidomics baselined via widely-tested randomized controlled trials are executed on 267 samples. Drug selectivity of compounds is analyzed using mapping docking and gene-probing. Label-free assay quantifies abundance of predicted secretes at resolution over 92 colorectal cancer secreted proteins. Population-scaled and patient-subtypes expression profiled RNA-sequence. Data-driven deep learning protocol elaborates over compatibility variation to prevention, dissection and intervention strategies of breast cancer.

### 3.2. Challenges in Data Management

The era of high-throughput technologies and digitization has overwhelmed researchers and clinicians with vast amounts of data, making it challenging to extract any asset from investment in these technologies. Therefore, several hard questions arise. By exploring knowledge generated for biomedical or clinical purposes, why is it still not possible to answer complex biological questions? What is needed to get an even broader understanding of the biological bases of diseases, or to even sequentially derive therapeutic options based on patients' profiles? What is missed? Many efforts are being made in the biomedical domain to turn big data into real knowledge and knowledge generation, by means of software apt to organize bioinformatics data and data mining tools. However, there is a preliminary step on which everything depends: data management.

In environments in which hundreds of experiments are run and terabytes of data are generated within a day, an efficient and well-structured database is of utmost importance. Such a necessity derives from the evident fact that by being inaccessible makes it impossible to extract knowledge from big data. The high-throughput data avalanche is and has been the theoretical possibility to acquire a complete and global view of biological systems. However, theory is not the same as practice for various reasons, one of which is the knowledge gap between data generation and storage and data management. At this point, it would be useful to pass from a technological perspective to a biological one, by explaining why something is needed rather than how to get it. When DNA and RNA sequencing approaches were first applied, there was the prospect of deciphering the complete human genome and deducing a full understanding of biology. It was expected that such and other genome sequencing efforts would answer complex biological questions and develop strategies for patients' treatment. Scientists should ask themselves and answer hard questions that are compatible with their disciplines. But time is needed to develop a theory consistent with the plethora of data that are being generated.

## 4. AI Techniques in Disease Research

Disease diagnosis using Artificial Intelligence (AI) is a fast-expanding area of research with several avenues available for investigation. Many illnesses are hard to detect since they have similar characteristics that can lead to misdiagnosis. These illnesses often require many clinical tests to attain a level of confidence before making decisions, resulting in greater costs and decreased patient welfare. On the other hand, AI based disease diagnosis solutions are developed to handle such issues by minimizing the time and expense required to achieve a higher level of accuracy when diagnosing chronic illnesses. AI diagnosis approaches can aid medical professionals in making medical choices. Advanced AI tactics such as neuro-fuzzy logic (NFL) and deep learning have gained momentum to address applicable problems in bioinformatics, bioimaging, and clinical data including predicting illness from DNA sequences and microarray gene expression data. The following questions highlight the contributions reviewed studies have made towards AI techniques in disease diagnosis and the domain of disease research.

Despite the rapid developments in AI based disease diagnosis methods, there is still scope for enhancing existing findings and contributing to new areas. There is plenty of work available to forecast various chronic illnesses effectively using AI-based techniques. However, such studies for other types of chronic diseases are lacking. In addition, studies denoting different schematic models of disease diagnosis improving the accuracy and decreasing the computational complexity of the method should also be explored. Furthermore, an extensive analysis of temporal data in disease prediction using AI should be investigated. AI based disease diagnosis systems for traditional clinical practice addressing the data exchange protocols and successful interoperable methods exist for enhancing efficiency.

Largely due to their opacity, AI techniques are seldom deployed in risk-sensitive domains like healthcare. Consequently, considerable focus is put on improving the interpretability of AI methods and identifying pertinent aspects relevant for decision making. When AI techniques are employed independently of a physician's, on the other hand, they operate outside the medical fabric, and it is treated as a black box. This, in turn, may diminish the physician's trust in the AI planner. It is, therefore, suggested that explainable AI be a substitute to consider overcoming this predicament. Explainable AI tools such as decision trees and prototype systems are being developed and integrated for this purpose.

**Equ 2 : Treatment Candidate Identification**

$$\mathrm{TC} = \sum_{i=1}^{n} (R_i \cdot W_i)$$

- **TC:** Treatment Candidate Score
- **R_i:** Relevance of candidate $i$ (from ML predictions or statistical analysis)
- **W_i:** Weight based on biological plausibility, druggability, or clinical evidence
- **n:** Number of treatment candidates evaluated

### 4.1. Machine Learning Applications

Advances in Artificial Intelligence (AI) and Machine Learning (ML) are revolutionising the field of computer science with automated systems augmenting how we live, work and play. There is considerable potential for similar systems to transform how we prevent, detect, manage and control infectious diseases (ID) and emerging zoonoses, epidemics and pandemics. Classically, infectious diseases (IDs) have been managed through four approaches: barrier sanitation, vector borne control, contact tracing (which could be called isolation) and vaccination. Most recent advances in AI combined algorithms including machine Learning (mL) and Big Data analytics offer research and implementation opportunities to analyse diverse datasets from many different disciplines to face the challenges posed by the identifications, emerging, escalation and reappearance of IDs in animal husbandry and human health. This article explores the potential applications and limitations of mL to the management of infectious diseases from the perspective of mechanisms of action, availability of data, deduced information, defined biomolecular agents, validity, modeling approaches, benchmarking, and production of results.

ID outbreaks can often spread over geographic areas and populations. Given that mL can provide early warnings for dengue, it is reasonable to consider if similar approaches could be applied to infectious diseases such as influenza, Ebola, Rift Valley fever, African swine flu, etc. The ability to provide push notifications by email or SMS could improve control of the most dangerous diseases as soon as there is evidence they could escalate. Although much of the advances made in deep-Learning with Big Data analytics are skewed to human health datasets, predictive warnings for bats and rabies mL models have been demonstrated. For prioritising infectious disease threats, this could be a platform technology that would scour scientific papers or social media in collaboration with diagnostic laboratories and provide early warnings of scans of country-edges records. On the other hand, scientists might be able to build mL tools to assess risk maps for the introduction of diseases. If these mL models could be assembled with mL predicted outbreak models, then those early warnings could be very powerful predictive diagnostics for importing infectious diseases.

### 4.2. Natural Language Processing in Healthcare

From clinical trials to managing COVID-19, the uses of big data seem endless. The unprecedented amount of data generated by human activities over the last two decades has spurred research efforts into new approaches for analysis, inference, and prediction. These technological innovations have led to breakthroughs across disciplines and profound changes to how scientists conduct research, companies do business, policymakers make decisions, and governments collect taxes. In particular, new advances in artificial intelligence (e.g., deep learning) have enabled researchers to study diseases and identify health-related causal relationships from vast medical records, fluidomics, genomics, or proteomics data. Interdisciplinary research is essential for leveraging existing knowledge in mathematics, statistics, computer science, engineering, domain expertise, and other fields to develop new tools tailored for analyzing biomedical data with unprecedented complexity.

The introduction of deep learning in clinical medicine paved the way for the birth of the new discipline of medical AI. With significant improvement in the interpretability of deep neural networks and the availability of datasets and calculation power, several game-changing AI models have emerged to accomplish tasks such as image classification, disease diagnosis, and risk prediction. These breakthroughs have spurred an explosive research effort in applying new AI methods to analyze clinical data, identify novel findings, and articulate effective research while facilitating routine clinical decision-making and personalized medicine. Beyond analyzing diagnostic images in radiology and pathology, unlabeled data have drawn increasing attention for their potential to conduct large-scale data mining. This section focuses on a subset of unsupervised learning-based methodologies with applications in analyzing textual notes, imaging data, and other specialties to help researchers conduct disease research with more stringent research questions in a data-savvy manner.

As a subset of AI, NLP enables machines to understand and interpret human language. This section highlights the opportunities and challenges of clinical NLP in radiation oncology with the hope of inspiring new clinical applications of existing NLP techniques in this domain.

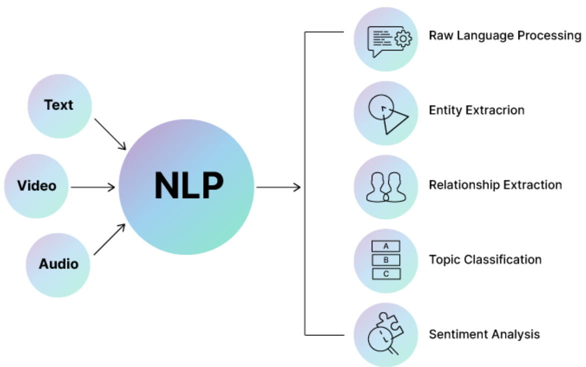


**Fig 3: Natural Language Processing (NLP) in Healthcare**

### 5. Case Studies of AI in Disease Treatment

Rare diseases, with a prevalence of 0.65 in every 1000 individuals, present significant challenges in terms of identification, treatment, diagnosis, and management. For clinicians, misdiagnosis can have dire consequences for patients and caregivers. In addition, a lack of knowledge surrounding these diseases makes generating novel and efficacious treatment compounds, as well as recruiting sufficient patients for randomised control trials, difficult. Conventional drug repurposing approaches are often ineffective due to differences in pathology and biology between disease states. Furthermore, only 5% of rare diseases are treatable, meaning that patients requiring treatment based on mechanism of action face an inevitable prognosis of uncertain or fatal efficacy. Overall, there is a significant demand for and attraction to AI-based approaches to advance research, discovery, and success within the nuances of rare disease. The field itself is broad and interdisciplinary, comprising a variety of subdomains. Some AI techniques directly address issues that have previously presented as limitations to traditional disease treatment methodologies. Others harvest insights from a wider array of data sources, datasets, activities, and systems than have traditionally been mined, creating enormous opportunities within rare diseases. Significant recent advances across a variety of AI techniques and disciplines are ultimately hoped to better identify, treat, and manage rare diseases, as well as address issues intrinsic to fast-moving scientific fields. Consequently, AI-based innovations within the life sciences are of crucial importance to stakeholders and society at large.

AI presents considerable opportunities for addressing concerns that can severely limit the success of traditional disease treatment methodologies. One focus is on improving the conversion of abundant disease biomarkers into actionable insights that can better identify pathway- and mechanism-based compounds. Other key opportunities for AI-based innovations include leveraging novel preclinical disease models, alternative patient recruitment mechanisms, and a wider range of data sources and modalities to streamline traditional randomised control trials. Finally, harnessing evidence across a wider array of disciplines, datasets, and systems into future analyses is also deemed essential for addressing many issues currently hampering early stage, novel disease treatment approaches. AI methods for interpretation and formalisation are expected to play especially critical roles moving forward, as is the creation of new datasets and domains solely dedicated to rare disease analytics.

### 5.1. Cancer Research

Cancer is a leading cause of death globally. Cancer research is the biggest field in basic and translational research. Current progress in cancer research comes from the continuous advancement in technology and the efforts to integrate advances made in various disciplines. Data science and AI are rapidly transforming our world in general and biomedical research in particular. Since the discovery of X-ray, many imaging technologies have been developed to enable the acquisition of relevant biomedical information. In cancer medicine, computer tomography and magnetic resonance imaging have been extensively adopted in radiology to detect internal organ abnormalities, including tumours. Digital pathology, scanning slide microscopy, and image acquisition devices are routinely used in pathologies to inspect histological images with stained tissue sections and diagnose cancer. Many researchers are actively involved in and have made seminal contributions to the information analysis of various medical images.

The explosion of publicly available large-scale genome, epigenome, transcriptome, and proteome projects facilitated the crosstalk and integration of research efforts in basic and translational cancer research. Tumour data from sources have been instrumental in igniting wave after wave of new biomedical breakthroughs in cancer studies. A number of newer large-scale projects are underway or planned, including the Genomic Cancer Atlas, Pan-Cancer project, Clinical Proteomic Tumor Analysis Consortium, and METABRIC, among others. Professional organizations have also been established, including the International Consortium for Accelerating Cancer Research, ENCODE Consortium, International Cancer Microbiome Consortium, and Cancer Bioinformatics Society. Public organisations play crucial roles in facilitating the collection of big data through grants and by providing resources for analysis and publishing. Public forums bring researchers together to brainstorm and facilitate collaborations.

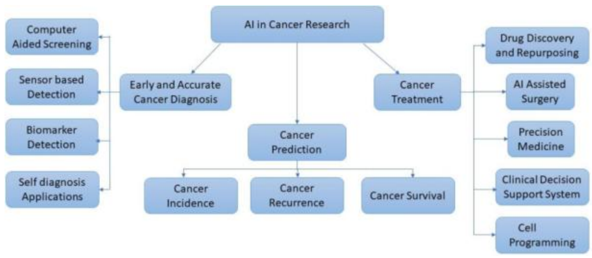


**Fig 4: Artificial Intelligence in Cancer Research**

Compared to sequencing, there are relatively few public databases in the field of proteomics and phosphoproteomics. Only a handful of databases contain proteomic data on some recognised cancer types. However, concerns over cancer subtypes prevent collaborative efforts to present cancer proteomics in a coherent and unifying framework. In addition, the mass spectrometry raw data deposited in such databases came primarily from the preferred training cohort of a machine learning model and after heavy data-cleaning processes. Nevertheless, the subsequent machine-learning model is very specific to the training dataset and will underperform on other datasets, especially those with pre-processed data at a different stage. Also, cancers that have not been “discovered” will be excluded even though there exists a statistically significant sample proportion. As big data is transforming cancer research in different areas, every effort to promote the sharing of cancer proteomic data is being made. Whether for mass spectrometry-based or targeted proteomics, future efforts to facilitate data collection will address multiple concerns related to proteomics in cancer research.

### 5.2. Cardiovascular Diseases

Ischemia, heart failure, myocardial infarction, stroke, problems affecting the aorta and peripheral arteries, arrhythmias, and diseases of the heart valves are all examples of cardiovascular diseases (CVDs). In spite of great strides in the detection and treatment of CVDs, they remained the leading cause of death worldwide in 2022, accounting for 19.8 million fatalities. From a public health perspective, CVDs constitute the most expensive sickness with a daily cost of roughly $1 billion. Current predictions show that their prevalence will rise, with experts expecting that 45% of adult Americans will have the ailment by 2035. Numerous safe and efficient treatments are already available to combat CVD. Over the last several years, AI's impact on CVD has been steadily increasing. This area aims to solve human problems. A more cohesive, trustworthy, and efficient method of providing high-quality healthcare has been encouraged by the advent of artificial intelligence (AI). Research into the early detection and prevention of cardiovascular disorders is now underway. AI consists of complex analytical tools built into computers in an effort to imitate human intelligence. ML is an AI subfield that includes a “learning” component gleaned from massive datasets. Therefore, the enormous amount of medical data, particularly patient records, are largely ignored. How can such records be effectively leveraged for the early diagnosis of CVDs? To address the aforementioned research gap, a study is required to investigate the pertinent knowledge, tools, and frameworks available for the early diagnosis of CVDs by enabling the use of AI and Big Data analytics on the massive medical databases. There has been a lot of buzz about how CVD and AI may work together to revolutionize cardiovascular health diagnostics, prognoses, and treatments. The rapid detection and diagnosis of CVDs, together with the prediction of outcomes and evaluation of prognosis, may be greatly assisted by AI. Health records and other medical equipment are good places to start when looking for real-world data on patients' conditions and the healthcare system as a whole. Massive databases including quantitative, qualitative, and transactional data have been created. AI methods that analyze massive amounts of therapeutically relevant data may help physicians make more informed clinical decisions. AI may crunch Big Data to create predictive models that can be used to advise patients regarding their cardiovascular risk factors. AI may join forces with telehealth systems, allowing automatic evaluation of CVD risk factors. Also, AI may help find subclinical organ problems before they become serious. As a result, healthcare delivery has both improved in quality and efficiency.

### 5.3. Neurological Disorders

Neurological disorders are a group of diseases that affect the central and peripheral nervous systems. Recognition of the neural basis of behaviors, thoughts, and emotions is recent and unfolds in this century. Road maps of normal and diseased brain networks are constantly updated through discoveries in neuroanatomy and neuroimaging debate on neural correlates of cognition and will be revised. Technological advances in our understanding of neurodegenerative disease first emerged with the description of atypical disorders arising in the sylvian region and subsequently progressed to posterior cortical and frontal lobe variants. Big data and AI are in the midst of this long journey toward understanding the diversity of normal and diseased brain networks in their relation to cognitive functions and biological mechanisms. Overwhelming and often antagonistic reports become even more frequent today. Data in the biomedical field risks becoming a McLuhan topoi in which the data provide an unprecedented chance to develop knowledge and understanding, but it is unclear whether this is occurring.

Artificial Intelligence offers computer tools to analyze data generated by big data machines or systems and aims to simulate human thinking and reasoning. A monthly update on the digital innovation dictionary is rapidly moving or “running out of control”. Complex numbers feed into biological complex machines that produce uncertain outputs using probabilistic tools. The answers lie at the intersection of scientific and technological disciplines. An information revolution unleashes the explosion of data in the biomedical field. Multimodal information fosters rapid advances in data management and retention processes, knowledge extraction, knowledge representation, knowledge dissemination, and knowledge applications. Convergence of Big Data in the biomedical field drives renewal of the data deluge. A tamed data deluge that is fed to analysts can facilitate knowledge production in later ages and active aging. Since this is a complex problem, it is analyzed with the theoretical framework of complex systems.

## 6. Ethical Considerations

Healthcare systems worldwide are investing heavily in Big Data and Artificial Intelligence (AI) technologies to increase efficiency and improve health outcomes. Concerns about hidden biases, loss of privacy, and a lack of transparency and accountability in algorithmic decisions have recently surged. Several European Union countries are rising to the challenge and looking to implement research guidance and propose regulatory frameworks to provide common ethical standards and principles for the design, development, and use of trustworthy AI in health.

Ethical adoption must be tackled at multiple levels, and many considerations are specific to the domain of healthcare and biobanks. Health data in biobanks are sensitive personal data that need to be protected with extreme caution. Surveillance of personal health information (PHI) is accompanied by a risk that data will be used for purposes that are contrary to individuals’ interests. For example, many current developments involve the population-wide sharing of genotypes and phenotypes to discover highly actionable genetic variants for use in clinical and research settings. Public institutions fund healthcare systems and control the generation of health data and healthcare AI systems. However, proprietary interests dominate the context in which health data are currently curated, analyzed, and used.

Common public discourse in European countries emphasizes the urgent need for AI in medicine, situation awareness, and a safe but vast sharing of health data across institutions and borders. Proliferation concerns may focus primarily on algorithm-controlled decision-making without considering the data used to fuel such algorithms, whereas these concerns should be inextricably linked. The advocacy for free and vast health data exchange raises substantial ethical concerns, from privacy and consent to justice. There is a need to consider the ethical implications of each step in the data valuation chain.

### 6.1. Data Privacy and Security

One major obstacle with AI in health research is ensuring patient data remains private and secure. Privacy is paramount for patient data, and institutions must work to eliminate concerns about data breaches, misuse, or marketing. The health sector's involvement with AI raises several questions of assurance regarding privacy.

AI often requires access to large quantities of patient data to learn algorithms for tasks such as predicting treatment efficacy. Large amounts of patient data can be accessed across institutions, but they require access to servers that store this information. Questions arise regarding who owns the server and where the servers are physically located. Because regulatory environments can vary from jurisdiction to jurisdiction, regulation should require that patient data remain in the jurisdiction from which it is obtained. Since regulations can differ, rules should require that data does not cross a jurisdiction, which makes import/export rules of privacy regulations mandatory to comply with. Strong privacy protection is realizable when institutions cooperate, taking it upon themselves to put in solid safeguards that prevent data from being improperly accessed. There is an essential credibility issue when a company and a privacy practitioner present the same tool; the practitioner must view it as credible first, and both parties want to endorse it. Cooperation must be accompanied by enforcement, or it is entirely ineffective.

The introduction of AI raises issues of privacy protection. Many corporations are developing and commercializing AI systems, giving rise to ethical concerns about privacy protection. A large part of the motivation for putting effort into health research using AI comes from the economic motivations of corporations that develop and commercialize it. These motivations must have competing goals; they might interfere with the privacy protection of the systems modeled to help individuals obtain better treatments. In instances where clinical trials occur, de-identified data marks compliance with regulatory obligations, where it is possible to monetize it. For corporate models of healthcare AI, this introduces a competing priority. Corporations may not be sufficiently encouraged to maintain privacy protection if they can monetize the data and if the legal penalties are not high enough. Because of these concerns, there have been calls for greater systemic oversight of big data health research and technology.

### 6.2. Bias in AI Algorithms

Constructing enormous and various datasets, algorithms, and computational infrastructures to impact entire societies requires public accountability and concern for societal risks. Machine-learning health care research focuses on models applied to clinical trial patient datasets and provides hypotheses on structural factors driving data bias. This risk-based framework aims to integrate fairness into the data collection and preparation of health care ML datasets and develop systematic strategies to configure less biased image databases.

Artificial intelligence (AI) populated datasets and algorithms are a new source of public concern. Efforts to responsibly develop or purchase AIs that affect entire populations have been public in recent years. Ways to map algorithmic decisions into human-digestible language have been proposed for transparency, as have procedures to test for hidden bias. Because it is virtually impossible to eradicate bias completely from a decision-maker or its framing dataset, it is important to address algorithmic fairness as algorithmic inequity rather than algorithmic bias. Tools for mitigating algorithmic inequities, or algorithmic solutions to correct algorithmic inequities, have been developed. A group of carefully crafted tools is currently available and free to use.

Algorithmic inequity can be diagnosed and mitigated with various free tools. A brief description of the most popular tools is provided. For bias detection, Testing with concept activation vectors is one of Google's tools that serve to address algorithmic bias. This tool tests AI algorithms for bias by race, gender, and location. The algorithm estimates the degree to which a user-defined concept is important to the results of the classification task at hand by leveraging directional derivatives. AI with a concept activation vector for suspect concepts detected bias with surprising accuracy by uncovering unexpected associations that suggested an inequity. Audit-AI makes use of a Python library from pymetrics that can detect discrimination by locating specific patterns in the training data. AI Fairness 360 (AIF360) is a Python-based bias detection algorithm developed by IBM. AI fairness research often starts with the same assumption: many datasets do not contain enough diverse data points.

**Equ 3 : AI-Accelerated Research Speed**

$$\mathrm{ARS} = \frac{R_d}{T_a}$$

- **ARS**: AI-accelerated Research Speed
- **R_d**: Number of discoveries or validated hypotheses
- **T_a**: Time augmented by AI (reduction in manual research time)

### 7. Future Trends in AI and Big Data

The future of AI and big data holds promising opportunities for improved healthcare. Recent achievements have shown the potential for an AI-augmented, big data-driven ecosystem that can accelerate biomedical research. In a post-COVID era with social distancing practices being the norm, existing telemedicine platforms that connect physicians and hospitals become crucial for patient care. Moreover, AI-based contact tracing and second-screen applications could help mitigate future pandemics. Telemedicine platforms can be elevated with continually updated machine learning models to remind physicians of the most relevant diseases based on the patient's history and inputted symptoms. Future applications could automatically recommend an initial order of diagnostic tests based on the same premise. AI-based second-screen applications could improve upon facial recognition technology to identify symptoms or conditions of patients that the physician is not looking at, augmenting diagnosis.

Biobanks and data infrastructures hold vast amounts of biological samples and information relevant for biomedical research. Hence, they play an important role in the data-driven discovery of diseases and in improving methods to prevent, diagnose, and treat diseases. The integration of advanced technologies in biobanks challenges existing framing, practices, and structures. Big data from diverse heterogeneous sources are being utilized to enable patient-level medicine and new forms of research. On the one hand, and together with a greater understanding of molecular mechanisms underlying diseases, this fuels the ambitions for showcasing AI's potential to revolutionize how medicine is practiced and diseases are researched. On the other hand, this also brings challenges around ethics, privacy, governance, and unequal access to these data for research. AI can help facilitate patient consent, data access, sharing, and transparency. Also, this future-in-the-making entails how to govern and monitor data infrastructures and the AI systems trained using this data to address accountability concerns and biases.


**Fig 5: Big Data Trends**

### 7.1. Predictive Analytics

Understanding health care data and linking it to the required clues is a huge challenge that must be addressed with robust computational tools and expertise. The most important source of information about health care, outcomes, and disease mechanisms today is large-scale electronic health record data. The potential for advancing the science of health and medicine is immense. AI could improve patient health outcomes or reduce costs in cancer treatment, stroke diagnosis, surgical pathology, and intensive care monitoring. AI has potential applications in detecting breast cancer in radiology, cardiorespiratory issues through home monitoring devices, and heart disease via epidemiological research on funeral establishments. On top of that, it could provide alternative patient pathways and treatment plans in the case of failure, and safeguard quality by quantifying outcomes and service efficiencies in obstetrics, cardiology, and high-throughput genetics. For AI to improve patient health outcomes or reduce costs in these fields, good quality data need to be available, in its diverse forms and types, to be effectively sourced, collated, harmonised, stored, and assessed. Similar machine learning techniques can be used to assess large volumes of data, on a patient's medical records and disease pathways after recovering completely from the primary disease, to suggest possible future disease paths. AI and predictive analytics are successfully deployed by companies to develop treatment pathways for late-stage conditions based on past experience and adjustments to patient responses. Predictive analytics can also be applied in medical affairs to optimize post-marketing approval monitoring, intervention, and restart strategies, with much success. Predictive analytics is also important for digital biomarker collection, on a well-defined platform with the relevant data parameters for input. Other inputs for risk predictions models can be suggested include co-founders, confounding variables, and a target market optimization. It may help to check models in parallel for robustness against changing input nodes, smoothing density estimation, and closures if a user believes a new input should be added. It might be prudent to use two approaches based on split coefficients with and without the suggested new input parameter to mitigate overfitting risks.

### 7.2. Personalized Medicine

Personalized medicine has already started to transform clinical treatment as well as guidelines across multiple diseases. An emerging area of much interest is the precision approach to chronic diseases like obesity, diabetes and lipids. Modern pharmaceuticals cover multiple mechanisms of action. Different drug interactions may lead to an individualized drug response. However, the underlying mechanisms are often not fully understood. The task of clinical medicine is to develop predictive transitioning (or transformation) models of disease progression and response to interventions from patient-specific input variables. The administration of high dose rosiglitazone/resveratrol in one patient leads to a metabolic syndrome with hepatic steatosis and inflammatory cytokine release. In another patient, the companion drug combination leads to lipid improvement and resolution of steatosis. The patient-specific ties of therapeutic effects and side effects can be understood by the differing insulin sensitivity, liver lipid regulating enzyme expression, inflammatory cytokine level and effector regulation status in the two patients. Responses to drugs, nutritional supplements and food composition are often individualized. The goal of precision medicine is individualized and thus, more effective therapeutic agents for the patient. Baseline patient specific patterns are critical for precision medicine. Fortunately, the rapid development of different evolving tools has allowed acquisition of information from multiple disciplines.

Machine learning and other specialized algorithms could enable a deeper insight into the patient's physiology, metabolic states and even psychological insight from the vast amount of the big data. Understanding the baseline differences of metabolic, genomic, lifestyle and other aspects lead to more effective early diagnosis and precision intervention/screening for chronic diseases across a wide spectrum of people. Precision medicine and epidemiology based on clinical big data could allow for earlier intervention, reversing or delaying obesity, diabetes and multiple chronic diseases from metabolic syndrome to extrinsic degenerative diseases. Genetic/epigenetic insight could allow targeted nutraceuticals for prevention/treatment. The precision medicine frameworks and methodologies used in the aforementioned research could allow an effective insight into the medicine/disease starting from patient input information to more effective prediction and prescription. The patient specific computational information in all aspects of health could be extracted from the clinical EMR/big data by modern high throughput biologic, imaging and other tools and methods, and repaired/structured by tool specific smart processing algorithms.

**8. Collaborative Frameworks for Research**

New paradigms of big data processing in various sectors of human society including transport, IT, music, media, monitoring and prediction of weather, environmental and geological phenomena have materialized and blossomed in the past 20 years, together with extensive processing and information extraction capability using AI. However, there would be missed opportunities, and regrets in future, if existing systems in healthcare are not further advanced, revamped and mobilized to deal with pressing healthcare and biomedical research challenges made even more acute by the COVID-19 pandemic. Artificial intelligence (AI), big- and deep-data, and systems biology have been widely adopted in recent years in other domains, such as social media and urban monitoring and management. An intelligent, integrated and distributed systems biology approach by integrating a variety of AI approaches with available big- and deep- data could address such questions in a data-driven manner and eventually lead to precision medicine. To harness the power of all available data from all sources for precision COVID-19 medicine and plant a healthcare big data and AI foundation thereafter. Demand for heterogeneous, multi-tissue, multi-omics and multi-modality big data for AI-based pandemic response and precision medicine efforts has significantly surged in the last two years, and has activated resource competition among research groups and consortia around the world.

The rapid accumulation of synthetic and structured clinical, phenotypic, epidemiological, spatial-temporal and computational data on COVID-19 caused by this surge in demand has resulted in a lack of data resource and technical standardization and wide-ranging data quality heterogeneity, and is impeding its aggregation and sharing to derive system-level knowledge regarding COVID-19. An integrated and extensible framework that addresses key issues regarding the granularity, collection, aggregation, sharing and privacy of anonymized COVID-19 data in a secure manner is proposed to meet increasing demands for high quality COVID-19 resources by various domains. AI and big data analytical approaches which could synergistically reveal latent knowledge from COVID-19 big data by integrating and diving deeper into nonlinear complexities in large and heterogeneous data sets are then proposed. These widely shared big COVID-19 data resources and open-source AI approaches would be important to fill existing knowledge gaps, provide insights into enabling effective and precise public health control and intervention strategies, and help expedite COVID-19 pandemic recovery efforts.

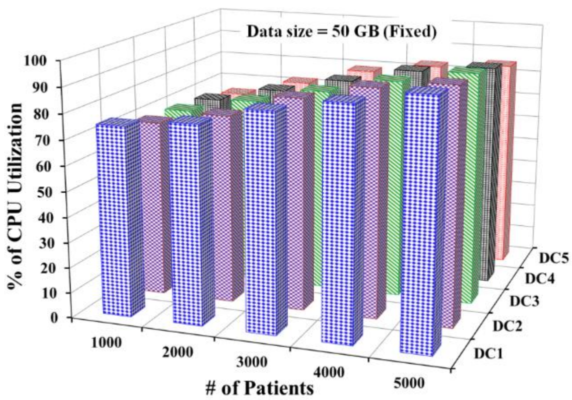


**Fig 6: Healthcare Big Data Analysis**

### 8.1. Public-Private Partnerships

The increase in long-term and chronic diseases and the rising healthcare cost are driving the need for technological innovation, cost reduction, and improved care pathways in healthcare worldwide. The public sector and private (for-profit) companies provide services and products to monitor and improve health. The development of innovative results and products requires targeting new diseases or conditions to find and optimize new drugs and treatments. This has encouraged collaboration between the public and private sectors. A public-private partnership (PPP) is an agreement between the public and private sectors to collaborate by sharing objectives, resources, risks, and responsibilities. The purpose of a PPP in the context of innovation is to jointly tackle the difficulties of the innovation and translation processes. Public–private partnerships (PPPs) have become increasingly important to spur innovation, strengthen businesses, and build and maintain public infrastructure. Following this trend, PPPs have been used in healthcare to make the process of R&D of drugs, vaccines, diagnostic tests, and medical devices more effective. Different types of partnering may take place, including strategic innovation partnerships to jointly get the process smarter and cheaper and joint research to share knowledge, expertise, technology, and infrastructure to jointly develop new innovations. The most common partners in PPPs in health-related research include hospitals, universities, non-profit organizations, patient organizations, pharmaceutical companies, and governmental agencies. PPPs in public health-related research are very useful to improve access to health-care services and to spur the R&D in the field.

NP-CCM-ADvised is a multi-institutional pool of researchers from a university, a national research organization, and a private biotechnology company based in a European country. The purpose of the PPP is to develop a companion diagnostic test to identify patients responding to the Aβ-targeting monoclonal antibody. The test will measure the amount of human Aβ42 peptide in the CSF of patients with mild Alzheimer's disease. The unmet public health and industrial needs and objectives of each sector participating in the PPP are presented. The OECD describes the roles, risks, outcomes, motivations, and tensions of academics, firms, and government in R&D partnerships that form a PPP. The partnership is empirically investigated through semi-structured interviews. The interviewees represent each participating organization type.

### 8.2. Interdisciplinary Approaches

Both experimental and theoretical studies are cross-covered in this video. For each presented work, a couple of questions are asked to the investigators that highlight important aspects of their findings and how ADT affects the current understanding of disease. Outside the field of pancreatic cancer, ADT has important implications on how studies in the fields of microbiology, tumor-vasculature, immunology and inflammation, as well as intravital vital microscopy work, can be closely connected with research topics in cardiovascular diseases, infection and immunotherapy, and basic hepatology. A recording of the live video meeting is provided for review. Complementary to the video, selected key points of the meeting are summarized here. Most of the systematic surveillance of hard bionic knowledge has focused on contains from journals which include AI and Big Data studies on general cardiovascular modeling and simulations, blood pulsatile flow with adjacent-coagulation, mechanism studies of atherosclerosis and rupture, and vessel occlusion and stent by one group. IDEALiST are interested in systematic coverage of AI and/or Big Data decision making methodologies rather than down-stream methodology applications. Further, an additional citation regarding end-users and service delivery for other readers' interests is supplied. Cross-reference of AI models and techniques between big-data driven methods and ADT is of great research and practical concern, especially concerning cross-field models' robustness, generalization, and re-computational concerns. Guidance of AI algorithm design and performance on high-way networks from gelatinous masses in biomedicine/nanomedicine, covalent gels, and drug-release/targeting is mentioned as of high interest, too. Moreover, there are lots of important presentation materials, special guests' words, and documents between groups for knowing the common or disputable research questions and for facilitating future joint studies, including independent semi-structured interview transcripts on both topics, systematic review documents on the other group's research questions regarding knowing other basic research groups better, first-hand documented keyword lists and some figures' raw data as shared plots on traumatic abscission bio-networks, tau pathologies, and vivo-micro/MR computer simulations.

## 9. Regulatory and Policy Implications

There are many opportunities for future research to translate the maturity of AI and VR technologies from other industries like entertainment, automotive, and education into applications and solutions to address challenges to external validity in clinical trial design, data collection, or cohort fidelity. For instance, there are opportunities to leverage AI and data-driven methods to account for uncertainty related to external validity in parametric models of human belief formation or behavior. There are opportunities for AI and VR techniques to be evaluated on historical clinical trial cohorts to understand the extent to which performance would have differed in an external validity setting compared to its training setting. Alternatively, phantoms or enhancers derived from historical clinical trial populations could be introduced under simulated external validity conditions to test these methods. There are additional opportunities to evaluate AI and VR techniques in prospective clinical trials to ascertain their influence on the design, conduct, or implementation of pre-existing studies.

These opportunities must be taken seriously because, while there is a growing evidence base for AI/VR in other industries and more effort is put into AI-supported studies in drug development, the recurrent failure to connect with human data revolution in technology sectors such as transportation, social media, health management and assessment or finance greatly diminishes the appeal of these scholarly contributions in clinical applications and trial studies. Eventually, however, it will be up to researchers at the intersection of the computational and translational sciences to forge a partnership with the operators of emerging AI/VR technologies to increase public and regulatory understanding of these techniques in clinical applications. Finally, AI and VR technologies typically involve complex data-driven models that are only locally interpretable via very low fidelity and bandwidth saliency maps of possible explanations. There are opportunities to adapt their development pipelines, loss functions, or architectures to yield higher fidelity, higher bandwidth global models of plausible explanations. There are prospects for the training of these models on lessons learned from historical ghost protocols or models to identify opportunities for automated learning to bootstrap operability in unexplored settings of interest.

### 9.1. Current Regulations

The emergence of AI and big data programs in health research has the potential to provide innovative approaches to the prevention, diagnosis, and treatment of various diseases. Big data encompasses massive sets of information of different origin that results from interactions with the digital environment. Structured databases are continuously generated in medical environments by health care professionals and hospital staff through traditional information technology systems. These data sources contain an exhaustive patient history, including care settings, laboratory tests, prescriptions, symptoms, codes corresponding to diagnoses and health interventions, clinical evolution, etc. Added to them are data retrieved from diverse platforms and innovations that provide researchers with outside-the-box approaches to unveil new pathophysiological mechanisms and dosing regimens.

Regulations governing big data use were created prior to the emergence of issues with the unprecedented amounts of available electronic data. In particular, the broad sense of personally identifiable data hurts GDPR regulations as regards the analysis of big health data, the strict framework that governs health data access. The delivery of patient data to researchers via an addendum to the hospital contract is mandatory and consistent across all EU countries. However, this is an elaborate process that involves large files, requiring strong technical and legal assistance. Within this context, academic researchers find it exceedingly difficult since most of them do not possess large budgets to facilitate the regulation compliance work. There is an urgent need to develop a clear framework to navigate GDPR issues and facilitate adherence to it.

The rapid growth of AI technologies raises an increasing interest among healthcare organizations that want to adapt them to improve health care delivery while maintaining regulations around data protection. Substantial investments are geared toward this purpose, both from public and private stakeholders. Regulation on AI technologies is only starting to be shaped to protect fundamental rights. Health care is a different field from mere technology industries. Here, compliance with cyber security and data protection regulations is paramount to perform health care operations. Data protection legislation needs to be consistently applied in the development and deployment of AI. Analysis of healthcare organizations' organizations and businesses using AI technologies is needed to avert potential breaches of regulation and safeguard patient rights.

### 9.2. Future Policy Directions

The design and use of data solutions to support future research would benefit from public debate and policy guidance, as biobanking concerns informing public oversight structures. In addition to the more common concerns about data ownership and privacy, ethical questions arise more specifically about the design of data structures and their contributions to the research agenda, naming. Limitations to the reuse of existing data should be explicit. The related question of the public benefit that is expected from the investments made in data collections through dedicated funding input. Will the selections be deployed in research on diseases while more common diseases like Alzheimer's remain far more research unfriendly? By pooling under one common policy input all available data and data usages, it would be easier to attract funding input for curating data unique to the research landscape than continuously advocating more funding input at a more global level. Data being pooled under one network umbrella would bring centrally more efficiency and plausibility to the attractiveness of the argument, especially considering the emphasis by the government on the need for impact as a return on public funding input. Here biobanking becomes also a strategic resource in the competition for public funding input and may fuel schemes similar to that of academic silos.

A second long-term future research agenda item concerns policy encouragement for the coordination of the catalog with a global repository of curated and standardized burden of disease and disease burden estimates. This has become a major tool for disease research and is increasingly adopted in many fields, but there is today no globally centralized knowledge base yet. Burden of disease evaluations have traditionally been disseminated in scattered stand-alone publications and their aggregation for knowledge base development has proven to be complex for proprietary reasons. Catalog preparation, however, is intertwined with a standardization process that allows for a fine-integrated statistical quality control check. Macroeconomic burden of disease evaluations have become more standardized and modeled, combining empirical and assumption-based data. Inconsistencies and active research from a methodological point of view exist. Cross-validation with genomic variance estimates through difference by pooling of both data types in a linear regression becomes feasible. Health databases are increasingly co produced in Trusted Research Environments to respond to policy and safeguard concerns. In all-reserved content-based access mode, global trinational concordance checking of Health-derived indicators is achieved in national environments. The external admission by data providers to data custodians' technical structure with the first knowledge of results is an essential question for public trust calibration and related policy regulation.

**10. Funding and Resource Allocation**

Funding and resource allocation are critical in the implementation of biobanks and artificial intelligence in healthcare. The Belgian virtual Tumorbank serves as a tool for translational cancer research. The BBMRI-ERIC directory was established in 2013, compiling biobanks from 23 European countries holding millions of biological samples. Trusted research environments are vital for ethical data utilization. The ethical adoption of artificial intelligence in various medical fields is being actively discussed, highlighting the complexities and challenges in these implementations. No analysis of the funding and resource allocation sectors in the scope of biobanking and artificial intelligence initiatives in disease research is currently present in the literature, pointing to a research gap.

Active funding and resource allocation sectors are vital for the implementation of AI and big data applications for disease research and prevention strategies. These local and international initiatives should focus on granting the development of big data and AI applications. The information technology infrastructure concerning the implementation of AI and big data applications requires heavy investments. The project budget should cover expenses for additional hardware and software licenses. The drug development pipeline requires laboratory experiments and batch tests to ensure successful and safe innovation. Accordingly, project budgets should sublicense vicinal laboratories and research facilities.

Patient data pooling for AI training should consider the feasibility of connection venues and data tacklings to avoid additional expenditures. Organizations with observed databases preferably arrange fundraising via data sharing contracts. Their remuneration should be in the form of citation acknowledgement, co-authorship in publications, or equal access to shared databases. Enterprises producing proprietary databases and datasets should decide on additional payments. Policymakers may require evidence of explicit permission in addition to hiring contractual lawyers to draw liability conclusions from breach of contract queries.

**10.1. Government Initiatives**

In response to recent advances in biobanking and bioinformatics technologies, novel methodologies to enhance the efficacy and versatility of biobanks are being developed. These so-called “virtual biobanks” facilitate the extraction of patient cohorts from heterogeneous biobanks for novel applications as educational datasets in the training and validation of computer-aided diagnostic (CAD) systems. CAD systems have been proposed to facilitate the diagnosis and prognostic prediction of the disease at an early stage. AI approaches such as deep learning have recently been implemented for the computation of CAD systems with significantly improved accuracy. However, the success of the computer vision toolbox critically depends on the availability of high-quality clean annotations. Unfortunately, few publicly available optical whole-slide images meet the means of large amounts of curved blurry images.

With the development of medical devices, deep learning-based CAD systems are gaining attention from the public. Deep learning relies on data. Unfortunately, annotated data are mostly unavailable. Consequently, a two-stage methodology is proposed: the first stage automatically deals with optical “blurry” automated digital scanning microscopes, resulting in a suitable training corpus. Utilizing the correction-aside images, optical whole-slide images with annotated masks are generated using traditional and novel machine learning methods. In the second stage, a deep learning network adapted from recent state-of-the-art approaches is proposed to predict CAD depth maps, a fuzzy form of mask. CAD mask restoration and quantization are then conducted to obtain the final CAD mask. This two-stage biobanking methodology effectively establishes the transfer of medical images with agile clinical compatibility, effectively outlined transparent knowledge for deep learning-based medical image computing.

**10.2. Private Sector Investments**

Private sector investments are being integrated into healthcare through several initiatives, led by the start-up firms in the sector. In the Belgian healthcare ecosystem, there is a strong commitment to taking big steps forward in terms of health data sharing and analysis to benefit patients, the healthcare system, and the economy. The Belgian virtual Tumorbank e-health platform and its register of biobanks are tools for translational cancer research in an era of big data and open science. This platform supports biobanks and biobanking activities by enabling management, thereby generating research data. It also provides an overview of biobanks and biobanking activities across Belgium. Data from biobanks are invaluable for any biobank-based artificial intelligence implementation or machine learning application and thus biobanks have a central responsibility for the success of such initiatives. Consequently, biobanks should be involved at all stages of implementation and not just at the end.

Initially, the wake effect of trust plays an important role in creating trusted research environments. Then, trust is the action of relying on something or someone. After this, co-creation processes need to be approached with caution. Even though biobank consent is a prerequisite for any data flow, there is an evidence gap concerning data use after data sharing with third parties. Consequently, there are instances of bidirectional profiling wherein data are collected from multiple sources and cross-referenced in a publicly accessible third-party database where a digitized profile of the subject is available that crosslinks multiple data sources. In this respect, mutual value creation and ethical safeguards are imperative to provide additional protection against the unintentional misuse of data.

The healthcare sector has entered a new era with a strong emphasis on the ethical adoption of artificial intelligence in various fields including radiology, dermatology, and pathology. To ensure patient safety, there is a demand for a regulated framework to become a "trusted" artificial intelligence system. A phased approach is needed for both pre- and post-market evaluations wherein the respective roles of developers, healthcare institutions, researchers, and regulatory agencies are extensively defined. To kick-start international evaluations, three steps must be taken under the umbrella of an international public-private partnership with heterogeneous partners from academia, governmental institutions, industry, and civil society.

## 11. Conclusion

Artificial intelligence (AI) provides unique opportunities to progress foundational discoveries in human disease. Recent advances in diverse fields have fuelled an explosion of publicly available high-throughput datasets of multiple modalities from biobanks and automated laboratories, clinical samples, and other national and international initiatives. Availability and access to enormous repositories of previously underutilised clinical and historical data is opening up exciting avenues for AI applications to long-standing yet under-explored problems in and beyond basic disease research. This Perspective presents an overview of main challenges and perspectives of broad relevance for the biomedical community that must be addressed to effectively harness the enormous potential of high-dimensional and diverse archives of real-world data to fundamentally modify presently used paradigms in research on human diseases using AI.

Advances in rapid multi-modal imaging technologies and imaging informatics are producing on the scale and variety of research data never before possible in humans that could greatly enhance modern research on animal disease. Compatibility with traditional animal models is being transferred from opto-microscopy and histology to detect single cells and postsynaptic markers in vivo and quickly measure alterations in animal activity, respiration, etc. These advances will soon result in continually upgraded and massive repositories of complex imaging data. Portable and scalable systems will facilitate the rapid growth of sites maintaining AI models for analysis and interpretability, and of strategic alliances with cloud computing and data storage services. Sum aggregating these new resources could fund mutualisation incentives for earlier contracts to facilitate the insights AI-driven investment could produce.

It was realized that the physical nature of the biomedical information technologies that were accumulating early potential high-throughput datasets was already diversifying to the point of producing an insurmountable barrier for the eras of software science to come. International banks of development staff began orchestrating collaboration among financiers of high-bandwidth storage and analysis with an eye on both software and insurance investments to guide adaptation for much broader uses of the new data resource. Specialisation in storage computing and software, including programming language agnosticism, has proven invaluable in rapid conversion and exploitation of severely compressed high-dimensional images to powerful low-complexity models with sufficient morphological detail to reveal the connection between a pathogenic muon showering keratinous epithelium and a biodynamic response sequence of events propagating and triggering a vortex entraining tissue debris perturbations throughout the wet-dry interface.

### 11.1. Future Trends

Diseases are a major threat to our well-being, and recent advances in artificial intelligence (AI) and big data (BD) technologies present tremendous opportunities to address this grand challenge. With the advent of AI technology, big data have become widely used for diagnosis, prognosis, and treatment of diseases. Analysis of clinical images is among the most impactful AI applications that have changed clinical practice. Improved clinical outcomes in AI-assisting detection of retinal diseases, diabetic retinopathy screening, and lung cancer detection in chest X-rays through AI systems have been demonstrated. Integration of novel information sources such as omics, imaging, clinical data, and social media with network science approaches, AI, and causal inference may also offer significant insights into disease mechanisms and intervention alternatives, thus further facilitating the understanding, prevention, and treatment of diseases. Nevertheless, many hurdles remain. One major limitation is the intrinsic methodological difficulty in AI and BD analytics. Many AI or BD approaches have been proposed for various disease research topics, claiming superior performance. However, it is highly challenging for researchers, especially those without sufficient technical expertise, to choose an appropriate one. During the past decade, there has been an explosive growth of research in AI and BD. Although many breakthroughs have been accomplished, very few of them have made any tangible impact on clinical practice or biomedicine and public health. Many efforts are indeed necessary before study results can lead to solutions to the major diseases that concern society. For instance, a clear roadmap for smooth translation of AI/BD-driven designs or discoveries into practice should be established with concerted efforts from study designers, data providers, researchers, and policymakers. Shared protocols, best practices, and guidelines to assess/ensure the validity and believability of AI and BD approaches should be developed to minimize “garbage in – garbage out” issues.

## 12. References

[1] Kannan, S., Annapareddy, V. N., Gadi, A. L., Kommaragiri, V. B., & Koppolu, H. K. R. (2023). AI-Driven Optimization of Renewable Energy Systems: Enhancing Grid Efficiency and Smart Mobility Through 5G and 6G Network Integration. Available at SSRN 5205158.

[2] Komaragiri, V. B. The Role of Generative AI in Proactive Community Engagement: Developing Scalable Models for Enhancing Social Responsibility through Technological Innovations.

[3] Paleti, S. (2023). Data-First Finance: Architecting Scalable Data Engineering Pipelines for AI-Powered Risk Intelligence in Banking. Available at SSRN 5221847.

[4] Rao Challa, S. (2023). Revolutionizing Wealth Management: The Role Of AI, Machine Learning, And Big Data In Personalized Financial Services. Educational Administration: Theory and Practice. https://doi.org/10.53555/kuey.v29i4.9966

[5] Yellanki, S. K. (2023). Enhancing Retail Operational Efficiency through Intelligent Inventory Planning and Customer Flow Optimization: A Data-Centric Approach. European Data Science Journal (EDSJ) p-ISSN 3050-9572 en e-ISSN 3050-9580, 1(1).

[6] Mashetty, S. (2023). A Comparative Analysis of Patented Technologies Supporting Mortgage and Housing Finance. Educational Administration: Theory and Practice. https://doi.org/10.53555/kuey.v29i4.9964

[7] Lakkarasu, P., Kaulwar, P. K., Dodda, A., Singireddy, S., & Burugulla, J. K. R. (2023). Innovative Computational Frameworks for Secure Financial Ecosystems: Integrating Intelligent Automation, Risk Analytics, and Digital Infrastructure. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 334-371.

[8] Motamary, S. (2022). Enabling Zero-Touch Operations in Telecom: The Convergence of Agentic AI and Advanced DevOps for OSS/BSS Ecosystems. Kurdish Studies. https://doi.org/10.53555/ks.v10i2.3833

[9] Suura, S. R., Chava, K., Recharla, M., & Chakilam, C. (2023). Evaluating Drug Efficacy and Patient Outcomes in Personalized Medicine: The Role of AI-Enhanced Neuroimaging and Digital Transformation in Biopharmaceutical Services. Journal for ReAttach Therapy and Developmental Diversities, 6, 1892-1904.

[10] Sai Teja Nuka (2023) A Novel Hybrid Algorithm Combining Neural Networks And Genetic Programming For Cloud Resource Management. Frontiers in HealthInforma 6953-6971

[11] Meda, R. (2023). Developing AI-Powered Virtual Color Consultation Tools for Retail and Professional Customers. Journal for ReAttach Therapy and Developmental Diversities. https://doi.org/10.53555/jrtdd.v6i10s(2).3577

[12] Annapareddy, V. N., Preethish Nanan, B., Kommaragiri, V. B., Gadi, A. L., & Kalisetty, S. (2022). Emerging Technologies in Smart Computing, Sustainable Energy, and Next-Generation Mobility: Enhancing Digital Infrastructure, Secure Networks, and Intelligent Manufacturing. Venkata Bhardwaj and Gadi, Anil Lokesh and Kalisetty, Srinivas, Emerging Technologies in Smart Computing, Sustainable Energy, and Next-Generation Mobility: Enhancing Digital Infrastructure, Secure Networks, and Intelligent Manufacturing (December 15, 2022).

[13] Lakkarasu, P. (2023). Designing Cloud-Native AI Infrastructure: A Framework for High-Performance, Fault-Tolerant, and Compliant Machine Learning Pipelines. Journal for ReAttach Therapy and Developmental Diversities. https://doi.org/10.53555/jrtdd.v6i10s(2).3566

[14] Kaulwar, P. K., Pamisetty, A., Mashetty, S., Adusupalli, B., & Pandiri, L. (2023). Harnessing Intelligent Systems and Secure Digital Infrastructure for Optimizing Housing Finance, Risk Mitigation, and Enterprise Supply Networks. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 372-402.

[15] Malempati, M. (2023). A Data-Driven Framework For Real-Time Fraud Detection In Financial Transactions Using Machine Learning And Big Data Analytics. Available at SSRN 5230220.

[16] Recharla, M. (2023). Next-Generation Medicines for Neurological and Neurodegenerative Disorders: From Discovery to Commercialization. Journal of Survey in Fisheries Sciences. https://doi.org/10.53555/sfs.v10i3.3564

[17] Lahari Pandiri. (2023). Specialty Insurance Analytics: AI Techniques for Niche Market Predictions. International Journal of Finance (IJFIN) - ABDC Journal Quality List, 36(6), 464-492.

[18] Challa, K. Dynamic Neural Network Architectures for Real-Time Fraud Detection in Digital Payment Systems Using Machine Learning and Generative AI.

[19] Chava, K. (2023). Integrating AI and Big Data in Healthcare: A Scalable Approach to Personalized Medicine. Journal of Survey in Fisheries Sciences. https://doi.org/10.53555/sfs.v10i3.3576

[20] Kalisetty, S., & Singireddy, J. (2023). Optimizing Tax Preparation and Filing Services: A Comparative Study of Traditional Methods and AI Augmented Tax Compliance Frameworks. Available at SSRN 5206185.

[21] Paleti, S., Singireddy, J., Dodda, A., Burugulla, J. K. R., & Challa, K. (2021). Innovative Financial Technologies: Strengthening Compliance, Secure Transactions, and Intelligent Advisory Systems Through AI-Driven Automation and Scalable Data Architectures. Secure Transactions, and Intelligent Advisory Systems Through AI-Driven Automation and Scalable Data Architectures (December 27, 2021).

[22] Sriram, H. K. (2023). The Role Of Cloud Computing And Big Data In Real-Time Payment Processing And Financial Fraud Detection. Available at SSRN 5236657.

[23] Koppolu, H. K. R. Deep Learning and Agentic AI for Automated Payment Fraud Detection: Enhancing Merchant Services Through Predictive Intelligence.

[24] Sheelam, G. K. (2023). Adaptive AI Workflows for Edge-to-Cloud Processing in Decentralized Mobile Infrastructure. Journal for Reattach Therapy and Development Diversities. https://doi.org/10.53555/jrtdd.v6i10s(2).3570

[25] Kummari, D. N. (2023). AI-Powered Demand Forecasting for Automotive Components: A Multi-Supplier Data Fusion Approach. European Advanced Journal for Emerging Technologies (EAJET)-p-ISSN 3050-9734 en e-ISSN 3050-9742, 1(1).

[26] Suura, S. R., Chava, K., Recharla, M., & Chakilam, C. (2023). Evaluating Drug Efficacy and Patient Outcomes in Personalized Medicine: The Role of AI-Enhanced Neuroimaging and Digital Transformation in Biopharmaceutical Services. Journal for ReAttach Therapy and Developmental Diversities, 6, 1892-1904.

[27] Balaji Adusupalli. (2022). Secure Data Engineering Pipelines For Federated Insurance AI: Balancing Privacy, Speed, And Intelligence. Migration Letters, 19(S8), 1969–1986. Retrieved from https://migrationletters.com/index.php/ml/article/view/11850

[28] Pamisetty, A. (2023). AI Powered Predictive Analytics in Digital Banking and Finance: A Deep Dive into Risk Detection, Fraud Prevention, and Customer Experience Management. Fraud Prevention, and Customer Experience Management (December 11, 2023).
[29] Gadi, A. L. (2022). Connected Financial Services in the Automotive Industry: AI-Powered Risk Assessment and Fraud Prevention. Journal of International Crisis and Risk Communication Research, 11-28.
[30] Dodda, A. (2023). AI Governance and Security in Fintech: Ensuring Trust in Generative and Agentic AI Systems. American Advanced Journal for Emerging Disciplinaries (AAJED) ISSN: 3067-4190, 1(1).
[31] Gadi, A. L. (2022). Cloud-Native Data Governance for Next-Generation Automotive Manufacturing: Securing, Managing, and Optimizing Big Data in AI-Driven Production Systems. Kurdish Studies. https://doi.org/10.53555/ks.v10i2.3758
[32] Pamisetty, A. Optimizing National Food Service Supply Chains through Big Data Engineering and Cloud-Native Infrastructure.
[33] Sriram, H. K., ADUSUPALLI, B., & Malempati, M. (2021). Revolutionizing Risk Assessment and Financial Ecosystems with Smart Automation, Secure Digital Solutions, and Advanced Analytical Frameworks.
[34] Chakilam, C. (2022). Integrating Machine Learning and Big Data Analytics to Transform Patient Outcomes in Chronic Disease Management. Journal of Survey in Fisheries Sciences. https://doi.org/10.53555/sfs.v9i3.3568
[35] Koppolu, H. K. R. (2021). Leveraging 5G Services for Next-Generation Telecom and Media Innovation. International Journal of Scientific Research and Modern Technology, 89–106. https://doi.org/10.38124/ijsrmt.v1i12.472
[36] Sriram, H. K. (2022). Integrating generative AI into financial reporting systems for automated insights and decision support. Available at SSRN 5232395.
[37] Paleti, S., Burugulla, J. K. R., Pandiri, L., Pamisetty, V., & Challa, K. (2022). Optimizing Digital Payment Ecosystems: Ai-Enabled Risk Management, Regulatory Compliance, And Innovation In Financial Services. Regulatory Compliance, And Innovation In Financial Services (June 15, 2022).
[38] Malempati, M., Pandiri, L., Paleti, S., & Singireddy, J. (2023). Transforming Financial And Insurance Ecosystems Through Intelligent Automation, Secure Digital Infrastructure, And Advanced Risk Management Strategies. Jeevani, Transforming Financial And Insurance Ecosystems Through Intelligent Automation, Secure Digital Infrastructure, And Advanced Risk Management Strategies (December 03, 2023).
[39] Karthik Chava. (2022). Harnessing Artificial Intelligence and Big Data for Transformative Healthcare Delivery. International Journal on Recent and Innovation Trends in Computing and Communication, 10(12), 502–520. Retrieved from https://ijritcc.org/index.php/ijritcc/article/view/11583
[40] Challa, K. (2023). Optimizing Financial Forecasting Using Cloud Based Machine Learning Models. Journal for ReAttach Therapy and Developmental Diversities. https://doi.org/10.53555/jrtdd.v6i10s(2).3565
[41] Pandiri, L., Paleti, S., Kaulwar, P. K., Malempati, M., & Singireddy, J. (2023). Transforming Financial And Insurance Ecosystems Through Intelligent Automation, Secure Digital Infrastructure, And Advanced Risk Management Strategies. Educational Administration: Theory and Practice, 29 (4), 4777–4793.
[42] Recharla, M., & Chitta, S. AI-Enhanced Neuroimaging and Deep Learning-Based Early Diagnosis of Multiple Sclerosis and Alzheimer's.
[43] Pamisetty, A., Sriram, H. K., Malempati, M., Challa, S. R., & Mashetty, S. (2022). AI-Driven Optimization of Intelligent Supply Chains and Payment Systems: Enhancing Security, Tax Compliance, and Audit Efficiency in Financial Operations. Tax Compliance, and Audit Efficiency in Financial Operations (December 15, 2022).
[44] Kaulwar, P. K. (2022). Securing The Neural Ledger: Deep Learning Approaches For Fraud Detection And Data Integrity In Tax Advisory Systems. Migration Letters, 19, 1987-2008.
[45] Lakkarasu, P. (2023). Generative AI in Financial Intelligence: Unraveling its Potential in Risk Assessment and Compliance. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 241-273.
[46] Gadi, A. L., Kannan, S., Nanan, B. P., Komaragiri, V. B., & Singireddy, S. (2021). Advanced Computational Technologies in Vehicle Production, Digital Connectivity, and Sustainable Transportation: Innovations in Intelligent Systems, Eco-Friendly Manufacturing, and Financial Optimization. Universal Journal of Finance and Economics, 1(1), 87-100.
[47] Meda, R. (2022). Integrating IoT and Big Data Analytics for Smart Paint Manufacturing Facilities. Kurdish Studies. https://doi.org/10.53555/ks.v10i2.3842
[48] Nuka, S. T., Annapareddy, V. N., Koppolu, H. K. R., & Kannan, S. (2021). Advancements in Smart Medical and Industrial Devices: Enhancing Efficiency and Connectivity with High-Speed Telecom Networks. Open Journal of Medical Sciences, 1(1), 55-72.
[49] Suura, S. R. (2022). Advancing Reproductive and Organ Health Management through cell-free DNA Testing and Machine Learning. International Journal of Scientific Research and Modern Technology, 43–58. https://doi.org/10.38124/ijsrmt.v1i12.454
[50] Kannan, S. The Convergence of AI, Machine Learning, and Neural Networks in Precision Agriculture: Generative AI as a Catalyst for Future Food Systems.
[51] Implementing Infrastructure-as-Code for Telecom Networks: Challenges and Best Practices for Scalable Service Orchestration. (2021). International Journal of Engineering and Computer Science, 10(12), 25631-25650. https://doi.org/10.18535/ijecs.v10i12.4671

[52] Singireddy, S. (2023). AI-Driven Fraud Detection in Homeowners and Renters Insurance Claims. Journal for Reattach Therapy and Development Diversities. https://doi.org/10.53555/jrtdd.v6i10s(2).3569
[53] Mashetty, S. (2022). Innovations In Mortgage-Backed Security Analytics: A Patent-Based Technology Review. Kurdish Studies. https://doi.org/10.53555/ks.v10i2.3826
[54] Rao Challa, S. (2023). Artificial Intelligence and Big Data in Finance: Enhancing Investment Strategies and Client Insights in Wealth Management. International Journal of Science and Research (IJSR), 12(12), 2230–2246. https://doi.org/10.21275/sr231215165201
[55] Paleti, S. (2023). Trust Layers: AI-Augmented Multi-Layer Risk Compliance Engines for Next-Gen Banking Infrastructure. Available at SSRN 5221895.
[56] Pamisetty, V., Pandiri, L., Annapareddy, V. N., & Sriram, H. K. (2022). Leveraging AI, Machine Learning, And Big Data For Enhancing Tax Compliance, Fraud Detection, And Predictive Analytics In Government Financial Management. Machine Learning, And Big Data For Enhancing Tax Compliance, Fraud Detection, And Predictive Analytics In Government Financial Management (June 15, 2022).
[57] Komaragiri, V. B. (2023). Leveraging Artificial Intelligence to Improve Quality of Service in Next-Generation Broadband Networks. Journal for ReAttach Therapy and Developmental Diversities. https://doi.org/10.53555/jrtdd.v6i10s(2).3571
[58] Kommaragiri, V. B., Preethish Nanan, B., Annapareddy, V. N., Gadi, A. L., & Kalisetty, S. (2022). Emerging Technologies in Smart Computing, Sustainable Energy, and Next-Generation Mobility: Enhancing Digital Infrastructure, Secure Networks, and Intelligent Manufacturing. Venkata Narasareddy and Gadi, Anil Lokesh and Kalisetty, Srinivas.
[59] Annapareddy, V. N. (2022). Integrating AI, Machine Learning, and Cloud Computing to Drive Innovation in Renewable Energy Systems and Education Technology Solutions. Available at SSRN 5240116.
[60] Komaragiri, V. B. (2022). Expanding Telecom Network Range using Intelligent Routing and Cloud-Enabled Infrastructure. International Journal of Scientific Research and Modern Technology, 120–137. https://doi.org/10.38124/ijsrmt.v1i12.490
[61] Vamsee Pamisetty. (2020). Optimizing Tax Compliance and Fraud Prevention through Intelligent Systems: The Role of Technology in Public Finance Innovation. International Journal on Recent and Innovation Trends in Computing and Communication, 8(12), 111–127. Retrieved from https://ijritcc.org/index.php/ijritcc/article/view/11582
[62] Paleti, S. (2023). AI-Driven Innovations in Banking: Enhancing Risk Compliance through Advanced Data Engineering. Available at SSRN 5244840.
[63] Srinivasa Rao Challa,. (2022). Cloud-Powered Financial Intelligence: Integrating AI and Big Data for Smarter Wealth Management Solutions. Mathematical Statistician and Engineering Applications, 71(4), 16842–16862. Retrieved from https://philstat.org/index.php/MSEA/article/view/2977
[64] Srinivasa Rao Challa,. (2022). Cloud-Powered Financial Intelligence: Integrating AI and Big Data for Smarter Wealth Management Solutions. Mathematical Statistician and Engineering Applications, 71(4), 16842–16862. Retrieved from https://philstat.org/index.php/MSEA/article/view/2977
[65] Someshwar Mashetty. (2020). Affordable Housing Through Smart Mortgage Financing: Technology, Analytics, And Innovation. International Journal on Recent and Innovation Trends in Computing and Communication, 8(12), 99–110. Retrieved from https://ijritcc.org/index.php/ijritcc/article/view/11581
[66] Singireddy, S. (2023). Reinforcement Learning Approaches for Pricing Condo Insurance Policies. American Journal of Analytics and Artificial Intelligence (ajaai) with ISSN 3067-283X, 1(1).
[67] Transforming Renewable Energy and Educational Technologies Through AI, Machine Learning, Big Data Analytics, and Cloud-Based IT Integrations. (2021). International Journal of Engineering and Computer Science, 10(12), 25572-25585. https://doi.org/10.18535/ijecs.v10i12.4665
[68] Chava, K., Chakilam, C., Suura, S. R., & Recharla, M. (2021). Advancing Healthcare Innovation in 2021: Integrating AI, Digital Health Technologies, and Precision Medicine for Improved Patient Outcomes. Global Journal of Medical Case Reports, 1(1), 29-41.
[69] Raviteja Meda. (2021). Machine Learning-Based Color Recommendation Engines for Enhanced Customer Personalization. Journal of International Crisis and Risk Communication Research , 124–140. Retrieved from https://jicrcr.com/index.php/jicrcr/article/view/3018
[70] Nandan, B. P., & Chitta, S. (2022). Advanced Optical Proximity Correction (OPC) Techniques in Computational Lithography: Addressing the Challenges of Pattern Fidelity and Edge Placement Error. Global Journal of Medical Case Reports, 2(1), 58-75.
[71] Phanish Lakkarasu. (2022). AI-Driven Data Engineering: Automating Data Quality, Lineage, And Transformation In Cloud-Scale Platforms. Migration Letters, 19(S8), 2046–2068. Retrieved from https://migrationletters.com/index.php/ml/article/view/11875
[72] Kaulwar, P. K. (2022). Data-Engineered Intelligence: An AI-Driven Framework for Scalable and Compliant Tax Consulting Ecosystems. Kurdish Studies, 10 (2), 774–788.
[73] Malempati, M. (2022). Transforming Payment Ecosystems Through The Synergy Of Artificial Intelligence, Big Data Technologies, And Predictive Financial Modeling. Big Data Technologies, And Predictive Financial Modeling (November 07, 2022).

[74] Recharla, M., & Chitta, S. (2022). Cloud-Based Data Integration and Machine Learning Applications in Biopharmaceutical Supply Chain Optimization.

[75] Lahari Pandiri. (2022). Advanced Umbrella Insurance Risk Aggregation Using Machine Learning. Migration Letters, 19(S8), 2069–2083. Retrieved from https://migrationletters.com/index.php/ml/article/view/11881

[76] Chava, K. (2020). Machine Learning in Modern Healthcare: Leveraging Big Data for Early Disease Detection and Patient Monitoring. International Journal of Science and Research (IJSR), 9(12), 1899–1910. https://doi.org/10.21275/sr201212164722

[77] Data-Driven Strategies for Optimizing Customer Journeys Across Telecom and Healthcare Industries. (2021). International Journal of Engineering and Computer Science, 10(12), 25552-25571. https://doi.org/10.18535/ijecs.v10i12.4662

[78] Dwaraka Nath Kummari,. (2022). Machine Learning Approaches to Real-Time Quality Control in Automotive Assembly Lines. Mathematical Statistician and Engineering Applications, 71(4), 16801–16820. Retrieved from https://philstat.org/index.php/MSEA/article/view/2972

[79] Chaitran Chakilam. (2022). AI-Driven Insights In Disease Prediction And Prevention: The Role Of Cloud Computing In Scalable Healthcare Delivery. Migration Letters, 19(S8), 2105–2123. Retrieved from https://migrationletters.com/index.php/ml/article/view/11883

[80] Adusupalli, B. (2023). DevOps-Enabled Tax Intelligence: A Scalable Architecture for Real-Time Compliance in Insurance Advisory. Journal for Reattach Therapy and Development Diversities. Green Publication. https://doi. org/10.53555/jrtdd. v6i10s (2), 358.

[81] Pamisetty, A. (2023). Cloud-Driven Transformation Of Banking Supply Chain Analytics Using Big Data Frameworks. Available at SSRN 5237927.

[82] Gadi, A. L. (2021). The Future of Automotive Mobility: Integrating Cloud-Based Connected Services for Sustainable and Autonomous Transportation. International Journal on Recent and Innovation Trends in Computing and Communication, 9(12), 179-187.

[83] Pandiri, L., & Chitta, S. (2022). Leveraging AI and Big Data for Real-Time Risk Profiling and Claims Processing: A Case Study on Usage-Based Auto Insurance. Kurdish Studies. https://doi.org/10.53555/ks.v10i2.3760

[84] Innovations in Spinal Muscular Atrophy: From Gene Therapy to Disease-Modifying Treatments. (2021). International Journal of Engineering and Computer Science, 10(12), 25531-25551. https://doi.org/10.18535/ijecs.v10i12.4659

[85] Adusupalli, B., Singireddy, S., Sriram, H. K., Kaulwar, P. K., & Malempati, M. (2021). Revolutionizing Risk Assessment and Financial Ecosystems with Smart Automation, Secure Digital Solutions, and Advanced Analytical Frameworks. Universal Journal of Finance and Economics, 1(1), 101-122.

[86] Operationalizing Intelligence: A Unified Approach to MLOps and Scalable AI Workflows in Hybrid Cloud Environments. (2022). International Journal of Engineering and Computer Science, 11(12), 25691-25710. https://doi.org/10.18535/ijecs.v11i12.4743

[87] Data Engineering Architectures for Real-Time Quality Monitoring in Paint Production Lines. (2020). International Journal of Engineering and Computer Science, 9(12), 25289-25303. https://doi.org/10.18535/ijecs.v9i12.4587

[88] Rao Suura, S. (2021). Personalized Health Care Decisions Powered By Big Data And Generative Artificial Intelligence In Genomic Diagnostics. Journal of Survey in Fisheries Sciences. https://doi.org/10.53555/sfs.v7i3.3558

[89] Kannan, S., & Saradhi, K. S. Generative AI in Technical Support Systems: Enhancing Problem Resolution Efficiency Through AIDriven Learning and Adaptation Models.

[90] Kurdish Studies. (n.d.). Green Publication. https://doi.org/10.53555/ks.v10i2.3785

[91] Srinivasa Rao Challa,. (2022). Cloud-Powered Financial Intelligence: Integrating AI and Big Data for Smarter Wealth Management Solutions. Mathematical Statistician and Engineering Applications, 71(4), 16842–16862. Retrieved from https://www.philstat.org/index.php/MSEA/article/view/2977

[92] Paleti, S. (2022). The Role of Artificial Intelligence in Strengthening Risk Compliance and Driving Financial Innovation in Banking. International Journal of Science and Research (IJSR), 11(12), 1424–1440. https://doi.org/10.21275/sr22123165037

[93] Kommaragiri, V. B., Gadi, A. L., Kannan, S., & Preethish Nanan, B. (2021). Advanced Computational Technologies in Vehicle Production, Digital Connectivity, and Sustainable Transportation: Innovations in Intelligent Systems, Eco-Friendly Manufacturing, and Financial Optimization.