International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2020): 7.803

Big Data-Driven Personalized Medicine: Integrating Genomics and Clinical Records Using ML

Sai Teja Nuka

Mechanical Engineer
Email: *saitejaanuka[at]gmail.com*
ORCID ID: 0009 - 0006 - 6549 – 9780

Abstract: ***The advent of next - generation sequencing technology has led to an exponential growth of genomic data. At the same time, the development of high - throughput methods for patient characterization, recording environmental factors around patients, and collating clinical data has led to an immense increase in clinical records. Analysis of large amounts of such diverse data can provide deeper biological insights about diseases and identify personalized treatments. Integrating and analyzing genomic data and clinical records from heterogeneous sources in a scalable and rapid manner pose significant technical challenges. In contrast to batch learning methods that require data storage and centralized processing, which are often inapplicable to medical data, online learning can process streaming data. Integrating them through online - learning - based systems can help analyze diverse data in a rapid and responsive manner without storage requirements. This is particularly reliable in a big data context, where online learning methods can scale even beyond data diversity to large data volume and unmatched data levels. It is therefore better suited for analyzing genomic information along with clinical and imaging data and for integrating information from diverse and concurrent data sources. In a data war, multiple health organizations develop ML - driven solutions to gain competitive advantages. A robust analysis and matching ML framework that guarantees data privacy allows organizations to compete with each other without the risk of data leaks. This guarantees data usage fairness and rewards the protection of patients' privacy. It also maintains the integrity of the institutional review board (IRB) process, which states that patient data will not leave the premises of institutions. Furthermore, it significantly reduces the risk of mass data breaches, which can have devastating repercussions on both patients and health organizations. Such reputations may include the loss of trust, reduced patient willingness to share data, and legal prosecution. In a similar vein, personalized medicine requires more diversity in treatment and diagnosis as opposed to traditional one - drug - fits - all strategies. Where there are millions of candidate models and parameters that can ingeniously predict treatment effects on prognosis but inadvertently introduce more complexities and uncertainties, a powerful framework that efficiently identifies reasonable explanations remains elusive.***

Keywords: Personalized medicine, Genomic data integration, Clinical data mining, Big data healthcare, Machine learning in genomics, Predictive analytics in medicine, Multi - omics data analysis, Precision medicine algorithms, Genomic clinical data fusion, AI in personalized treatment, Data - driven patient stratification, Health informatics ML, Biomarker discovery ML, Electronic health record (EHR) analytics, ML models for disease prediction.

1. Introduction

Personalized medicine is in a transition phase. It used to develop drugs with a targeted action on a specific target, such as an infected tissue, a cancerous cell, or a pathway involved in pathophysiology. The target is generally tightly connected to the formulation of the drug and is the basis for its pre - clinical and clinical development. However, at the same time, it is crucial to consider the opposite process. Analysis approaches hold the potential for the characterization of metabolic or genomic patterns responsible for disease propagation or work - up of therapeutic treatments. Artificial intelligence currently does not integrate with analysis in the development and validation of new drugs. It is crucial to widen the analysis capabilities currently available to academic and industrial labs to exploit this novel potential for the development of driven personalized medicine. A clinical trial, aiming at validating a personalized adaptive treatment with respect to a standard treatment in patients affected by biomarker positive mutations frequently found in advanced non small cell lung cancer, is considered. The static evaluation of responses collected within the framework of clinical trials raised the question on the feasibility of capturing, decoding, and characterizing single patient mutation - and - response patterns. The initial pharmacological agents alter either protein structures of kinases or intracellular pathways targeted by these kinases. The goal is to identify biomarkers predictive of therapeutic efficacy, in particular, for treatment with inhibitors. Specifically, to characterize responses of patients treated with an inhibitor and analyze the variations of the oncogenic mutation according to therapeutic outcome. Data on genomic signatures and therapeutic responses as part of a multi - centered validated trial are considered for analysis. Clinical records prospectively collected in a clinical trial have been used as training data to investigate deeper insights on patients' responsiveness. The challenge is to derive a functional evaluation of treatment effects from sparse clinical responses. In particular, to characterize both personalized responses revealed by data and predicted outcomes missed in clinical data. As shown in these early analyses, all machine learning methods successfully captured complex input/outcome maps on the training set, resulting in an average validation.

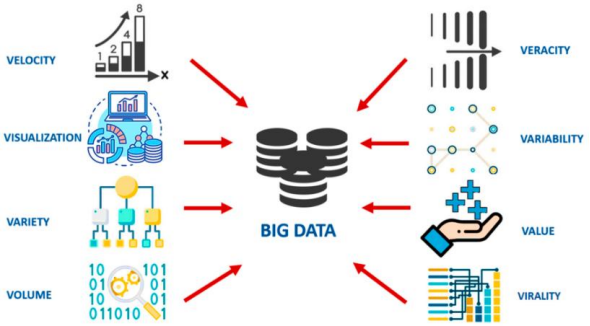

Figure 1: Genomics and Big Data Analytics for Personalized Medicine and Health Care.

1.1 Background and Significance

Personalized medicine is regarded as the most promising frontier of biomedical research and healthcare industries with sufficient temporal and spatial multi - omics data collection. Government and private organizations are harvesting multi - omics and clinical records at an unprecedented scale and rate. The treatment of diseases can be tailored based on the genetic background, lifestyle, and environmental exposures of patients. However, the success of precision medicine relies on efficient integration and model creation of the collection data. Such complexity arises from high dimensionality, missing values, and heterogeneous nature. The efficient integration and modeling of multimodal data is currently one of the most challenging grand challenges of artificial intelligence in the bio - health field. Here, representative studies and achievements, tools, and algorithms are reviewed to address the need for a systematic understanding of the state of the art. After the Human Genome Project trained researchers in the country to analyze single - omics, systematic studies on multi - omics integration will enable them to study the genomic landscape of diseases better. Nowadays, multiple diseases, for example, autism, cancer, and diabetes, can be modeled from a multi - omics perspective, which holds great promise for greater understanding and improved diagnosis and treatment of diseases. However, multi - omics integration is a challenging task due to the difficulties in accommodating the diverse nature of such expanded modalities.

Unbalanced features and missing values are common difficulties in the study design and scientific conduction. Such complexity requires advanced approaches with the ability to deal with multimodal integration. The integration is also hindered by the fast growth of the available data and omics, as acquiring new ones does not seem to end. A systematic review of the advantages and limitations of current statistical tools and computational platforms would help researchers navigate through the various choices of the field.

2. Understanding Big Data in Healthcare

Recent advances in biosensors and digital imaging techniques have led to an explosion in the generation of clinical data for biomedical research as well as healthcare delivery. All these developments have brought enormous advances in disease diagnosis and treatments, but have also introduced new challenges as large - scale information becomes increasingly difficult to store, analyze, and interpret. In recent years, medical record - keeping institutions, pharmaceutical companies, and biotechnology start - ups have been collecting vast amounts of clinical data, treatment outcome data, as well as data from genomics, transcriptomics, proteomics, micrometabolomics, and other high - dimensional data spaces. This problem has given way to a new era of “Big” Data in which scientists are exploring new ways to understand the large amounts of unstructured and unlinked data generated by modern technologies, and leveraging it to discover new knowledge. The ultimate goal is to convert all this data into information, and ultimately knowledge, hypotheses and predictions that can be tested, verified, and acted upon in the laboratory or the clinic.

Yet, despite these advances, for all the promises that Big Data holds for revolutionizing healthcare and disease treatment, we see few examples of Big Data being leveraged in healthcare despite the plethora of data that is available and the enormous socio - economic benefits that could be reaped. In fact, countries that invest massively in biomedical data generation technologies and pharmaceuticals could gain huge economic returns on the basic research they fund. It is hard to think of a field with more healthcare data than cancer and, therefore, one with more opportunities for Big Data discovery. Nevertheless, an analysis of the field of Oncology indicates that despite billions of dollars being invested in genomics data collection, TCE data collection, and possibly two competing drug - treatment databases (one for treatment regimens, the other for drug - drug interactions), there are little indications still that the whole is greater than the sum of these parts. Analysis of this fragmented landscape reveals large swathes of unmonetized data - generating healthcare institutions and small start - ups, and they are all doing their own mutually incomprehensible and isolated things.

2.1 Research Design

Given the rise in biobanks, rapidly decreasing costs of omics data generation, and new developments in core IT infrastructure, big datasets will soon be more readily available for biomedical research. New tools are required to navigate the information deluge in the precise medicine era. Nonetheless, the growth of new data types also comes with serious challenges. Biomedical research is often siloed, with genomic and clinical data belonging to different fields. Interdisciplinary research is rare, leading to inefficiencies and missed opportunities. Genetic variants and other biomarkers have been hypothesized as crucial components of outcomes - related treatment decisions, preventing adverse drug reactions and improving health care costs. Competitive methods based on traditional statistics and bioinformatics are thought to be reaching their limits, leading the field towards wide adoption of machine learning (ML) methods, which have made significant strides in traditional statistical and computational bioinformatics.

It has previously been demonstrated that ML can enhance the identification of actionable biomarkers beyond current efforts by focusing on the different characteristics of the new data types. This demonstrates the feasibility of genomics - driven and CSC - informed patient stratification, but only limited conclusions can be drawn for real - world application due to the relatively simplistic model systems employed. The integration of often much larger clinical records in ML rigorously trained on genome - scale data may be difficult in practice. As an effort to take a step towards this integration of big data sources, developments in genomic bioinformatics, chemoinformatics, and clinical records processing were translated into generalizable workflows.

The approach provided can incorporate any kind of static or temporal data on either biological systems or item characteristics, which can be amended to incorporate other biological data or to use different representation schemes for the input features. Nevertheless, some of the data transformation steps are specific to the particular data types considered. Multiple linear regression and similarity

ensemble clustering were chosen as the first models to be considered in predictive screening (PS). The proposed methods could also easily be replaced with other ML models or modules from more extensive native implementations in ML analysis environments.

3. Machine Learning Fundamentals

ML refers to a sub - field of artificial intelligence that uses computational techniques to learn patterns of relevant features from data to predict outcomes of unseen observations. Because big data in biomedicine is often collected as complex n - dimensional arrays, it shall be referred to here generically as 'data'. Genomic data are collected in form of n×5–200k matrices, where n is the number of subjects and the number of columns vary by a factor of 10 between different data types, e. g., n∼105 for expression data and n∼107 for methylation data. Clinical data can be treated as n×p tables, where p usually range from 6 for less enriched clinical records (i. e., only diagnosis and medications) to hundreds for some extensively characterized populations. ML approaches are fundamentally classified into two categories: supervised and unsupervised. Supervised ML learns a classifier from labeled inputs and corresponding outputs provided during training. Unsupervised ML acts on unlabeled input data, inferring properties about the discovered clusters and providing a basis for future classification. Recently proposed semi - supervised learning approaches build classifiers with the assumption that the unlabeled data can be grouped into multiple categories is a balancing hybrid between supervised and unsupervised learning.

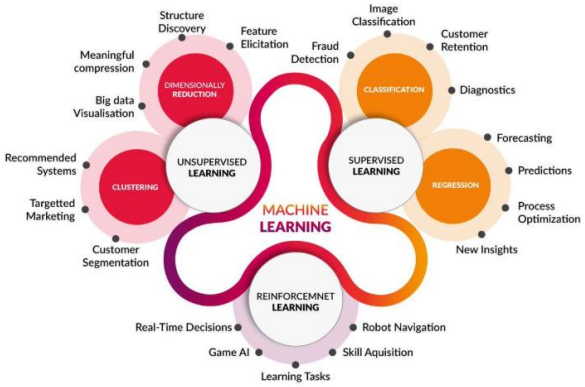

Figure 2: Machine Learning Fundamentals

Stockwell Transform (ST), a time - frequency analysis method that offers flexibility and time - frequency resolution is used to extract Time - Frequency Domain Features (TFDF) from one - dimensional signal data. Genomic, imaging, and multifactorial data/types converge to large multi - dimensional data stored in huge databases in the era of big data and Knowledge Discovery from Databases has drawn wide concern in recent years. Different from traditional Bioinformatics and Data Mining, 'deep learning' does not rely on domain knowledge, classifying raw data automatically from zero and can be seen as the best explorer instead of developers in learning representations of worldly objects. However, there are limitations in deep learning models for clinical data integration with electronic health records composed of heterogeneous data types, while complex learning and early stout might not complement with christening biomedical discoveries. The conclusion is also drawn that ML provides promising strategies to process different types of biomedical data and integrate diverse clinical features for precision medicine.

3.1 Types of Machine Learning

Understanding Machine Learning and Its Various Types

Technology, specifically computational technology, has advanced tremendously and continues to advance. The development of computer technology has also led to the introduction of this technology into many fields, including health care, education, marketing, and so on. These computer systems are now capable enough to digitally record information and process it to give results tailored to the requirements of a user, thereby automating the processes required. Medical records of patients are recorded in a physical format and, for the past few decades, in an electronic format. Development in information technology has enhanced the productivity of the medical field in managing electronic health records (EHRs).

Storing and managing such a high number of records has become a challenging task for health organizations. Many options are available today for health organizations to store and manage EHRs. The medical data of patients often include their personal records, medications taken, clinical notes, laboratory investigations that have been carried out, radiators and imaging information, allergies, immunizations, etc. This variety of data has become a source of information that, if properly processed, can help predict many diseases of personal concerns, with specific records of data resulting in a tailored health diagnostic tool, estimation of failures in health of the population, analysis of population health statistics in service to the society, and so on. Moreover, the analysis of this data with proper mathematical techniques can also help benefit medical institutions, health organizations, and laboratories in developing more sophisticated equipment for diagnosis.

Machine learning (ML) is a subset of artificial intelligence (AI) that enables systems to learn from data to identify patterns and make decisions with minimal human intervention. The main goal of machine learning is to construct a model that can accurately predict outcomes or classify data points. This model is a mathematical representation of the data learned through the underlying patterns. The growth in technology and its diffusion into society has led to the generation of huge amounts of data, which often cannot be dealt with manually. Consequently, the demand for tools and techniques to extract relevant information from data and build systems that could mimic human intelligence increased.

3.2. Key Algorithms Used in Healthcare

Machine learning (ML) refers to a set of technologies and methods used to outperform traditional algorithms. Machine learning models improve by learning from existing evidence and new data, making increasingly accurate predictions. One group of methods, supervised learning, builds prediction models from already collected evidence. A different class of algorithms, called unsupervised learning, generates models

without having pre - collected reliable prediction results. It uses only raw input variables to extract structure and knowledge from information. An example of unsupervised models is clustering techniques, which group entities based on similarity. Using supervised methods solely on the analysis of electronic health records data, deep - learning - based approaches straightforwardly integrate structured and unstructured data. Both structured and unstructured clinical data can be used to predict clinical events, nurture recommendations, and classify patient cohorts.

Interpretable systems, which generate a decision rule that is easily conveyed to the users and allow human understanding, are crucial in the medical environment. Health being a sensitive field with high stakes, systems that are not interpretable could fail to be trusted by practitioners. This occurs despite superior predictive performance and could lead to no adoption of their results for the benefit of patients. Various methods exist for interpreting supervised learning prediction models. These include rule - based systems and global surrogate models. Shapley Additive Explanations (SHAP) is a general methodology that generates explanations in the fu of identifying features that contribute most to a prediction.

In the healthcare ecosystem, predictions on patients from machine learning approaches, most of which have a black - box nature, can be troublesome. A lack of explanations of why a patient is predicted to be at risk leads to helpless practitioners not being trusted in their decisions. This ultimately defeats the purpose of improving the patient clinical pathway. ML is important in clinical pharmacology to enable better drug prescriptions. The amount of drug exposure depends on the patient's pharmacokinetic profile and relies on their parameters, genetics, clinical context, and administered drug. However, since these parameters can exhibit high variability and due to the lack of extensive knowledge of the model relationships, standard population models are often used. A standard regimen plan is selected for patients considered average with the extreme assumption that one size fits all. This causes variable therapeutic efficacy and side effects.

4. Data Integration Techniques

Although clinical data and OMICS data serve different purposes in patient care, recent studies show that clinical data possess strong predictive power while multi - OMICS data provide very descriptive demographics. On one hand, clinical data types are heterogeneous with various sizes and structures, e. g. text - based clinical records, numerical laboratory measurements, and coded diagnoses. On the other hand, multi - OMICS data types are homogeneous, displaying consistent numbers of patients and standardized matrix structures. The goal is to introduce data integration techniques that can bridge the heterogeneity of clinical data and high dimensionality of OMICS data, thereby facilitating the analysis of clinical data with embedding knowledge from genetics.

In detail, the i2b2 architecture itself is a three - layer architecture that consists of a central data layer that includes multiple and heterogeneous clinical databases, a server layer that includes i2b2 - related biomedical web services developed in Java to access the data layer, and representation layer that includes web clients for clinical output visualization. The focus was mainly on the data layer of this architecture to integrate different clinical databases into one i2b2 easily.

Given the diversity of clinical databases and demand for working with them, existing health IT environments are heterogeneous and thus hard to fully utilize; a tight integration of them into one would certainly result in a very difficult development process. The different clinical and biomedical domains with their diverse standards for databases result in confusion on data representation, data - entry procedures, lab measurements in different units, and failed interoperability. In such a complicated world of various systems and domains, a well - designed architecture for biomedical research is extremely valuable for reaping the full potential benefit of these commercial systems. It can widen the control of most basic clinical business processes (i. e., clinical data collection and storage) from the various sources of data - mining AI engines in the future analysis stage. i2b2's open - sourced clinical data integration technology enables mining huge amounts of clinical big data and thus making them sharable. Unlike systems that automatically store clinical records in a number of disparate tables and only serve a limited number of queries, i2b2 provides an entire architecture consisting of clinical data integration and a variety of plug - and - play analytical engines, ranging from statistical and machine learning packages to deep learning environments. It therefore provides a basis for more complex biomedical data - mining challenges. It has the potential functionality to go beyond the i2b2's clinical data warehouse and be integrated with imaging, semantic/metabolic, and genetic data - mining engines.

In summary, the i2b2 architecture complements existing systems rather than competing with them. It is nevertheless clear that the intensive efforts made by other groups and a tight integration among these base - layer systems will be necessary to address the MACS 2014 challenge, which is to go beyond these base - layer 'unclean' data sources.

4.1 Integrating Genomic Data

Most of the challenges for the integration of genomic data within EHRs center on the open access to genomic data. In recent years, a large number of companies, research institutions, and hospitals have started to offer completely online services for ownership, storage and interpretation of genome variants, gene panels, and exomes. Although there are plans to allow genomic data to be shared openly and accessed by researchers and healthcare professionals worldwide, the lack of proper laws that govern the ownership of genomic data, data privacy and ethical matters are creating issues.

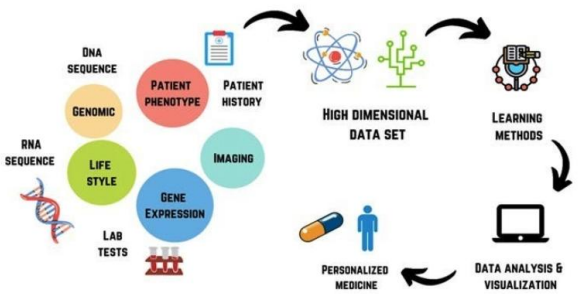

Figure 3: Integrating Genomic Data in Personalized Medicine

The debate over whether the genomic variant should belong to the person from whom it was sequenced, the clinic that performed the sequencing, the health care institution that offered the testing, the researcher that analyzed it, or the company that developed the testing, is a hot topic at meetings related to social and ethical issues in genomics, but at the moment there are no clear answers. The Clinical Laboratory Improvement Amendments (CLIA) of the United States federal law regulate laboratory testing to ensure the safety and accuracy of laboratory tests. Because of these restrictions, a few online analysis centers have started offering non - interpretative analysis, in which raw data could be uploaded for less specific analysis, but this is currently limited to only a small number of widely adopted gene panels, it creates a burden on laboratories that may offer clinical tests and need to accommodate an unusual data format.

Genetic data could be considered as sensitive data, and some countries rule their transmission strictly. Different regulations exist in different continental areas and even in a same country, depending on the state. Even in research centered countries, huge costs are incurred by taking precautionary measures. Recently, there has been several efforts to define ways to analyze genomic data that are stored in a remote cloud without transferring them, but practical solutions are still lacking. Most companies offering genomic testing do not guarantee that the raw sequencing data of tests can be accessed by external parties, and only a few allow reparsing the raw data, while the remaining databases do not guarantee that the data remain accessible in time. Some companies, while allowing raw data text files to be downloaded, only provide an interpretation of single nucleotide polymorphisms but not of more complex types of variants such as copy number variations, structural variations, or exome variants.

4.2 Integrating Clinical Records

While an extensive amount of clinically correlated genomic data are generated daily through advanced sequencing technologies, the integration of such genomic data with clinical and biomedical records still presents a major challenge for data analysts and researchers. An ideal setting to integrate, query and analyze genomic big data in a timely manner would be a big data - based platform designed with a "distributed" architecture where genomic data are stored at different nodes of a network and the applied analysis may be conducted at each of these nodes without the requirement to move the data to one central workstation. Within such settings, particularly privacy - sensitive medical information, such as data from electronic health records (EHR), may be stored locally and not disclosed to external authorities for analysis. However, it would still be inheriting all data mining tools and functional capabilities that are available within the distributed data warehouse.

In the last decade, state - of - the - art natural language processing (NLP) methods and tools were developed. The majority of the research on NLP - based analysis of textual corpora is however based on monolingual techniques even though a large share of the available textual data is described in two or more languages. The rapid development of social media in the last decade has proved extremely beneficial for the dissemination of diseases and their symptoms by the affected citizens. The clear impact of such data on people's health by the affected individuals is also supported by a rapidly growing body of literature aiming to analyze health events using social media data. Text mining NLP methods are employed to extract epidemiological information related to the outbreak of diseases.

5. Challenges in Data Integration

The rapid growth of sides in both the initiation of clinical patient records and genomic data is creating a challenge for the pairing of such data into information systems that will facilitate joint scientific queries on the transcriptome and clinical data. This paper deals with and investigates three strategies to integrate clinical data from the i2b2 and genomic data from a diverse resource of structural genomics data and functional genomics data, examining their cardinalities and means.

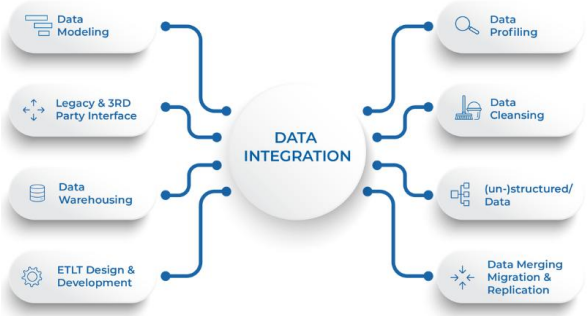

Figure 4: Data Integration Challenges.

The quality of the schemas and the integrity of the mapping rules help the data modelers provide review and feedback to the schema designers. Meaningful transformation steps, such as Normalization and Aggregation, would also have entities for in - line protocols. The combined use of querying and data integration tools can be much more productive than relying on the querying tools alone. Unlike the many tools and established techniques for schema matching and data preparation, users of querying and integration tools need to often re - learn the condition language, put - in Rule Editor criteria, and manage the execution of relevant agents.

With careful planning, particularly with Schema Assembly controls and direct agent execution of pre - planned redundancy - removing schema queries, a much increased productivity of pathological query processing can be attained. Data transformation based on a set of mapping rules, applied to correspondence relations in coupling rules, requires referring to the research of fine - grained corrections of

transformation and coupling rules. Many existing approaches transform and couple schemas at the level of database relation and schema representation. Due to schema heterogeneities that arise from both differences in semantics and ordinalities of patient history.

5.1 Data Privacy Concerns

The collection and use of genomic data is expected to be a driver of the next breakthroughs in healthcare and personalized medicine, otherwise this data will be effectively wasted. Health data is an invaluable asset that might improve population health and the quality of care while reducing overall costs. In the past decade, this realization has led to substantial investments to collect genomic data. For example, the U. K. government has announced plans to sequence the genomes of five million people. The European Union funded the H2020 project focusing on data sharing and big data analytics in biomedical research and patient care. Markets in Europe and Asia have emerged to store and share health data. Unfortunately, at the same time, questions about privacy have gained new importance.

There is ample reason for concern regarding data privacy. A study demonstrated the feasibility of de - identifying medical prescriptions with supervised machine learning. With respect to genomic data, research showed that 99.9% of all U. S. individuals can be uniquely identified from the combination of just 64 single - nucleotide polymorphisms (SNPs). Furthermore, an adversary might obtain such SNPs through free state - of - the - art sequencing services. Because genetic information is static and shared with relatives, privacy breaches will become a much more serious problem than in the case of traditional data. Thus, without a proper, rigorous privacy framework, data holders of genomic data will hesitate to share them.

5.2 Data Quality Issues

Data quality issues are among the main challenges for your research field objectives. Data quality refers to the grade of data concerning use of data quality indicators: accuracy, completeness, consistency, credibility, precision, and timeliness. In the health big data era, data is collected from various heterogeneous sources with diverse formats, semantics, and protocols. Quality evaluation of big data in the health domain is hence a challenge. Two aspects under investigation include big data quality evaluation and health informatics interventions promoting data quality improvement. The first study attempts to identify and formalize quality indicators for health big data using literature - based knowledge modeling. A framework is proposed. A conceptual framework and prototype tools called iBigCue (Intelligent Big Data Quality Evaluation) are developed for mining, evaluating, and visualizing health big data quality in accordance with the proposed framework. The framework and tools are a first step towards supporting evidence - based data quality evaluation and quality data use in health systems. With the wide adoption of electronic health records (EHRs), immense amounts of health data are systematically collected with in - depth details about patient healthcare over time. Secondary use of EHR data for research and decision support is of great value for improving data - driven healthcare services, patient outcomes, and health systems performance. Major challenges include data quality issues with respect to both obesity predicted and related predictors. This paper reports such first - hand experience with focusing on issues and methods for the secondary use of EHR data for survival analysis. Data quality issues include variable - related, entity - related, issue - related, record - related, and time - related issues. These issues lie in the collected data extraction from EHR sources, secondary use data from raw EHR data, and translating subject domain knowledge into engineering logic. Development of external informatics algorithms such as rule - based approaches, pre - trained machine learning models, and resampling techniques, is essential. The promising experimental results suggest future work to further improve work performance through systematic evaluation and integration of more informatics methods.

6. Case Studies in Personalized Medicine

Prediction of breast cancer recurrence in clinical state This study aimed to predict breast cancer recurrence with integrated genomic and clinical data. Legacy clinical data resources as well as genomic data such as uPA/PAI - 1 test, IHC test, and microarray test were collected from EHR, PACS, and previous tests. In this case, physician - derived features would play vital roles in prediction accuracy. Hormone receptor status, tumor stage, and lymph node involvement were selected as features detected from physicians' comments. Additional features like LP had to be computed with prognosis interval and tolerance. Baselines were built using six classic ML algorithms. Random forest, support vector machine, and decision tree data mining algorithms outperformed other schemes. Tree - based model was used as the meta - ML model here since it allows understanding of how the input variables affect predictions. In this case, a personalized medicine question was presented where the goal was to decide treatment recommendations for chronic myeloid leukaemia (CML) patients based on genomic sequences. Each CML patient sits in a unique position in the variable space, and treatment regimens vary according to the spatial placements. The approach focused on the treatment recommendations based on inflammatory markers. The first task was to extract the features from blood test records, classify the new records into one of the 4 classes, and reclassify the potential false positive records from the first stage. Using a class imbalance correction method, multi - class imbalanced classification was studied, and a few techniques for specific clinical domains were proposed. The network relies on multi - class classification mechanisms; nevertheless, in real - world scenarios, a patient may have multiple possible treatment regimens offered by multiple different doctors, which may contradict each other. Both the number of candidate treatment regimens and the time taken for recommending a decision would significantly increase with a large number of sampling requirements. Moreover, conflicting recommendations would confuse the patients.

6.1 Successful Implementations

Next - generation sequence and clinical data integration has been signaled out as the next step for the genomic data hosting and processing ecosystem. To support this move, a

proteomics data management platform designed for compatibility. Using a publicly available MS data set on rhinovirus - A16 infection, different levels of complementarity between software for data import, sample quality control, data processing, and batch correction, tested. In terms of technical implementation of MS workflows, one - third of evaluated algorithms permitted binning of complementary tools in Texture, PeptideQuant, and SumOf in Quasar, respectively. However, the integration of tools from different software bundles and merging of analysis results across samples is usually not possible due to a lack of common file formats and well - defined minimum requirements of input data. To ensure seamless compatibility of batch analysis workflows, the output of open - source software is. In addition to capability testing of biophysical binding assays, modalities of use to adapt available models/additional data types to experimental design and consequently increase the size of applicability domain are illustrated. These approaches are compatible with available open - source implementations and commercial tools in this field and compared with respect to their training load, available options for probing agent design and the type of data output. We have shown the potential of mass - spectrometry - based proteomics as a tool in personalized medicine, utilizing an analytical platform with a high - throughput protein quality control assay, a multiplexed biointeraction assay for simultaneous screening and target nomination, and a classifying pKD framework complemented with additional machine - learning algorithms. The steps necessary for the application of this platform to drug discovery and development dosing regimens tailored to patients are considered.

6.2 Lessons Learned

Big data is an emerging field that encompasses enormous volumes and varieties of structured and unstructured data across systems and organizations that cannot be captured, stored, managed, or analyzed without using new innovative technologies. In health care, big data has been harnessed to assist decision - making, manage knowledge, identify risk factors, determine treatment options, generate health and care strategies tailored to individuals' needs, and predict health status trajectories. Precision medicine, or personalized medicine, refers to tailoring a specific therapy to an individual based on their genomic and clinical records. Precise and individualized therapies are acknowledged as the next frontiers in therapeutics. The cost of high - throughput sequencing technologies and the lift in data volumes collected has stimulated the growth of bioinformatics, which is assessed as a prime application of big data in biomedicine. Simulation criteria for temporal multiplex networks may assist with big - data - driven health management and treatment monitoring. Recent tethered smart devices generating substantially comprehensive bio and clinical health data present new opportunities and challenges for PM solutions.

Omics technologies span multiple domains, including genomics, epigenetics, transcriptomics, proteomics, metabolomics, microbiomics, etc. Omics are relevant to multiscale PM since the data measures on different biological systems and processes, on different time scales, and on information, knowledge, and evidence contents. However, such collections of heterogeneous big data pose major challenges to how best integrate and analyze the data. Though multiscale PM requires ad hoc and integrated analytics, there is a lack of methods and systems that allow a multi - dimensional and multi - scale analysis at the same time. Few studies appraise multi - level integration of omics, clinical procedures, and other information sources on health development, disease progression, and treatment reactions. An effort was also made to introduce network frameworks into this multidimensional data integration problem, nevertheless focusing mostly on the methodological aspect rather than a concrete biomedical application.

7. Future Directions in Personalized Medicine

The potential of big data in biomedical research of the future will depend on its bulk, accuracy, and richness. Genomic, ontological, phenotypic, and therapeutic big data should be individualized and collectively stored and accessible in secure and interoperable formats. New here ontology and format - conversion efforts will be needed to enable this compilation. Furthermore, data - mining technologies and bespoke AI algorithms should be deployed to mine the high - dimensional heterogeneous big data correctly. This will alter the course of personalized practice (s) in medicine, leading to smarter diagnosis and precision drug design to bring miracles at an affordable cost worldwide. Computational biomedicine is at the interface of biology and computer science, including areas such as bioinformatics, biostatistics, biomedical data mining, computational medicine, network biology, omic - data science, and systems biology, wherein mathematical, statistical, and computational modeling and simulation are utilized to solve problems in biomedicine. Multi - omics will be capitalized to understand diseases and drug processes' etiology and mechanism at multiple molecular levels. Further, studying tumor evolution and tumor - immune interactions will enhance the effectiveness of cancer immunotherapy.

Figure 5: Future Directions of ML in Personalized Medicine.

Modernization will entail smart, facilitated, and citizen - consented delivery of biomedical data to patients, clinicians, researchers, and pharmaceutical companies for collaborative and pooled opportunities, curtailing time, resources, and costs. Feasible codecs and/or converters should be devised for omics big data in binary forms that fit on affordable consumer electronics globally. There should be guaranteed free online downstream services worldwide that automatically convert and provide processed genomic data in disease - ontology, drug - reaction - ontology - based formats depending on the patient's consent. Enhanced free follow - up services should fulfill precision medical needs before deciding treatment. Should design - a - drug - need arise, automatic matching of

receptors' pharmacogenetic information with existing drug structures will be provided following the algorithmic opening of associated molecular medicinal chemistry bio specifications. These capabilities will upgrade childhood to clinical genome - to - drug - to - home big data analytics and preclusion opportunities for any disease domain. Similar algorithms will catch up and speed with other compensating biomedicines curious about unexplored diseases. Personalizable and affordable ubiquitous initiatives to enhance detection and treatment prospects could unify the world.

7.1 Emerging Technologies

Healthcare is entering a new era of data - driven discoveries and personalized applications. New data acquisition devices such as digital pathology, wearable biosensors, and molecular imaging are generating huge amounts of high - dimensional data like genomic sequences, molecular structures, and hybrid records of clinical notes, imaging, and pathology. The wealth of data containing information related to drug mechanisms and patient response to specific treatments is motivating big data - driven precision medicine applications. Data analytics techniques such as computer vision, natural language processing, deep learning, and reinforcement learning have been successfully applied and widely adopted in other industries. However, the long tail of 90/90 rule applies that it is typically a 90 - to - 10 rule in terms of cost and effort with 90% funding requirement directed toward 90% of the effort in data acquisition, preprocessing, cleaning, integration, and normalization. The supply problem remains a critical challenge in healthcare AI.

Personalized medicine is a new healthcare paradigm that integrates the Natural history of disease (NHD) model and its data structure in developing precise data - driven predictive analysis. By tailoring treatment to individual patients based on integrated genomic/clinical/microbiome exposure data, it aims to optimize treatment efficacy and patient health outcome. A NHD model that integrates the knowledge of genomic and clinical data with ML techniques is studied to personalize therapy for bladder cancer patients based on their hlA genotype. Pre - processing and integration approaches that transform genomic and clinical records into structured representations are highlighted to enable answers to sophisticated queries that present a unified view of comprehensive information.

7.2 Potential Impact on Patient Care

Big Data is presently on the tip of everyone's tongue. This term encapsulates an era where information has become ubiquitous, generated by various devices, in growing volumes, with increasing velocity and variety. Suddenly, databases, hard drives, and even the cloud feel too small, and new options such as Blockchain are now in vogue. Big Data is currently driving scientific advancement in many fields, with scientists looking for new, efficient, and actionable ways to understand large amounts of unstructured data from genetic sequences to Twitter. New technologies such as Next Generation Sequencing (NGS) generate vast amounts of information to better understand biology and pathophysiology, and at the same time, treatments are becoming increasingly individualized.

Some large - scale data - driven projects to leverage Big Data are already in place in the fields of health and healthcare research. The concept of "Population Health" and commenting on "Big Data in Digital Health - Care" focuses on NHS Digital's new Data Services Platform, which is expected to support research on and improve patients' health outcomes. However, Big Data still has immense potential advantages to unlock in the field of healthcare and clinical practice in order to create more personalized and effective treatments. Some large - scale datasets even derive from public or partially open sources; including, but not limited to, patient records from EMRs, wearables, genetic assays, and clinical reports. The global healthcare stack also faces a unique infirmity; the digitization processes and the transition from paper - based records to exclusive EHRs still represent a major task. Compliance and standardization to abide by CURES and preventive health regulations in order to be able to share records among national endpoints is time - consuming and resource - wasting. Moreover, Big Data means putting together "2D" or "X - Y - Z" datasets, rather than solely storing text files or images into data lakes. The amounts of Big Data are too - often unmanageable, as follows. In the present era, Machine Learning (ML) enables the development and deployment of new approaches to leverage Big Data. ML, a discipline of AI, provides the means to understand, solve, and improve processes using data rather than following strict instructions. Actionable machines based on data and algorithms already impact individuals' every - day lives, producing better or targeted recommendations on the amount of time spent on a specific website, or in decoding the propensity of a contact network node to develop a disease. The Healthcare sector has not been alien to this wave, with Modelling Learning Easy (MLE) and much experimental literature currently affecting patients, physicians, payers, researchers, and companies. All sorts of datasets have been studied and leveraged, from structured patient records to text - based clinical notes.

EHRs are databases where patients' records are stored. Following standards, they store heterogeneous information such as structured records, clinical notes, laboratory tests, and images. EHRs have become more accessible through remote electronic access and easy data manipulation, driving the discovery of biases in demographic data, confirming the "social determinants of health" (SDS - H) and illustrating the possibility of actionable analysis on datasets. These new resources have also made an impact on the collection and now predicted cost - effectiveness of drugs on heterogeneous data. Drug models integrated with NGS, routinely collected data such as EHRs, or datasets are filtering patterns in the treatment effects of NGS based on EMERGE OncoCube, as well as evaluating/repurposing treatment regimens based on EHRs and CDRs.

8. Ethical Considerations

Federated data technology enables virtual unification of data from different sources under a uniform data model, while the underlying data stores operate autonomously. This would allow currently isolated genomics training datasets to become accessible to machine learning models. Privacy - preserving technologies enabling safe and ethical data access should be pursued. Machine learning models have several applications in genomic medicine such as to recommend diagnostic tools and pharmacogenomic therapies based on the patient's genetic makeup. It is critical to address the under representation of many ethnic groups and the social, environmental, and health disparities prevalent in clinical research and healthcare datasets. Machine learning algorithms may exacerbate inherent biases. Clinically underserved communities are unlikely to develop trust in machine - learning - guided genomic - based treatment plans unless health disparities research is incorporated from the start of the model - building process. To engender trust and build a culture of ethical and transparent machine - learning applications, partnerships among stakeholders should be promoted. Machine - learning model developers in clinical settings should understand health disparities research as a prerequisite for applying their models to patient data. Other ELSI considerations include establishing standards for explainability, transparency, reproducibility, trustworthiness, and accountability regarding machine learning applications in genomic medicine. ELSI research at the interface of machine learning and genomic medicine reveals a multitude of scenarios in need of further research support. The risk of "black box" algorithms, which cannot be interpreted, is a nationwide concern that is amplified in clinical practice with severe implications for screening, diagnosis, and treatment. The academic community must respond to the demand for advanced models while ensuring that patients' rights or social goals are not compromised and that trusted access, use, and sharing of health, genetic, and other sensitive data is enabled. Explanatory algorithms must complement predictions through uncertainty quantification, sensitivity analysis, probing, intuition building, imitation, and visualization.

8.1 Informed Consent

Many genomic research initiatives that triage and store anonymized and broad consent evidence in biobanks require re - invitation for new studies. Furthermore, many patients wish to notify secondary use once healthy again or change treatment. Using a 3D scoring system to visualize types of consent categories in easy - to - understand form and with flexibility in a dashboard enables medical doctors and biobank staff to target and select categories in a direct user - friendly way. Over time, genes defining the heritable risk predisposition for common diseases in modern populations have been uncovered using various methods.

These genetic factors influenced the development of individualized risk scores for clinical usage. Molecular medicine translates genetic information into individual structure–function derangements of genetic landscape. Both approaches are hampered by phenotypic measures which are imprecise and non - determined biological and environmental exposure factors apart from sex and age and ethnic ancestry. Using microarrays to gather deep - level histological and pathological, gene and microRNA transcriptomic, and integrated epigenomics data sets from supervised medical health records analysis, several proofs of the prospects and challenges of polygenic risk score and molecular medicine usage for health have been published. Up to 35% of initial phonemes of acute coronary syndrome, breast, prostate, and colorectal cancer were found sharable with research biobanks. Large - size biobanks of diverse ethnic ancestry and diagnoses exist and are expected to get annually feasible biobanks of diverse ethnic ancestry and diagnoses by creating individual genomic information - based rapid health improvement.

8.2. Equity in Healthcare Access

Access to healthcare is a universal human right essential for achieving ever - growing life expectancy and quality of life across the globe. Significant disparities across the world exist in access to protective health services. Such disparities matter immensely when translating into avoidable deaths or associated morbidity. Equal access to healthcare means that individuals should not be privileged for inherent characteristics such as biological makeup, birthplace, or socioeconomic conditions. The philosophy of universal equity calls for health systems to be designed in a way that minimises inequalities of access, even if this goal cannot be achieved in its entirety. The World Health Organization describes health equity as ensuring that "everyone has a fair and equal opportunity to be as healthy as possible."

As an equity analysis of Big Data - Driven Precision Medicine, it must be highlighted how AI tools can also worsen equity issues. The symposium reports a lot regarding population diversity. But who has access to the data and the associated tools? For example, the European general public often expresses discontent with cancelled or delayed health care services. It is often presented in the political domain and in the media that there are not enough doctors in Europe, let alone burdens such as migratory crises. In the meantime, it appears that billions of patient records are captured. In many hospitals, for example, patient records are digitised, but restrictive policies on general access apply. In the USA, facing far broader burdens, patients can switch clinicians, but this is almost impossible in Europe. In short, ZIN and the symposium present great opportunities, but it should also be stated who owns the data and resources and, more importantly, why they should benefit from using it.

This equity analysis of Big Data - Driven Precision Medicine will focus on relying on deep networks, as these require significantly less casewise knowledge than other AI algorithms. Still, state - of - the - art deep learning systems remain either poorly interpretable or high - complexity. Therefore, possibilities should be sought to alleviate these issues. For example, there appears to be no intrinsic limitation in algorithm data compressed in digital signal - info frames. Perhaps large networks could be dissected into manageable chunks. Practical computational complexity, interpretability, robustness, and risk attenuation or behaviours should be considered. Nevertheless, even if interpretability was assured, the dose–response curve might still be showing outcomes that correspond to the field statistics scale of application, and here lies a profound ethical concern. In summary, the equity analysis by which Big Data Precision Medicine could yield

more equitable healthcare services is participation in data/resources by patient rights.

9. Regulatory Framework

As Big Data and artificial intelligence are increasingly harnessed to improve health outcomes, regulators will be challenged to keep pace with innovation. Present efforts are focused on analyzing data for designing algorithms that accelerate healthcare. Emergent algorithms must be subjected to scrutiny akin to drug models to ensure safety and efficacy prior to implementation. For models already adopted, mechanisms must be instituted to validate efficacy against benchmarks once in use. For regulators, cause for optimism lies in the fact that the strategies needed are well known since the fields of computational biology and bioinformatics have been advancing for decades. Solutions to issues brought forth by treating health data as a commodity will also be found by evaluating ethical and philosophical frameworks that are emerging.

For example, much in the way antibodies are evaluated for safety, the FDA already has a framework for how Big Data models will be evaluated and regulated. This framework must be updated to allow for the unique challenges and emergent concepts that will arise. Nevertheless, it is surely preferable to writing regulations from scratch at the outset.

As regulators devise new frameworks however, care must be taken to ensure the resultant strategy is not worse than the existing approach of treating health data and derived adaptations as immutable secrets. This defective stance has not proven to be sufficient in preventing bias from being incorporated into models, nor has it been successful in preventing discriminatory use of the biometric data itself. At present, the patchwork of policies around the world is inadequate to address issues ranging from competition to algorithmic transparency, and biomedicine will be left at the mercy of tech monopolies. Meanwhile, the information asymmetry created by allowing companies proprietary access to health data is antithetical to the idea of a health - focused digital society.

9.1 Current Regulations

Developments in human genome sequencing and the collection of large - volume clinical records that follow patients' longitudinal outcomes act in a synergistic manner to revolutionize the practice of medicine, allowing for the development of precision health systems that provide upgraded transparency in care. The rapid development of genomic technologies is matched by the advent of smart sensors and synchronous miniaturization of non - invasive medical devices. Data - driven machine learning algorithms are deemed a necessity to sculpt the colossal data stored in cloud computing approaches into useful health information. New dimensions of infectious and social disease predictions, interactable holographic visualizations, and autonomous robotic platforms are expected to emerge with the technological advancements in medicine. These developments call for a radical conceptual redesign of health systems, with the potential of changing the current design shipping permanently. National sequencing projects and a global interest in health data sharing have the potential of opening the Pandora box of unethical automated algorithms that uncloak and broaden the extent of human biometrics upon invisible internet - operated machines. Artificial intelligence (AI) is transforming health systems into information monoliths that are increasingly controlled by potent black - box closed - source algorithms, deployed by monopolizing benched corporations in the business of mass information collection and extraction. The widespread use of wearable health sensors in health care deepens the existing dilemma, posing significant challenges for ethics, regulations, and culture. Genome - data - powered biobanks are set to obstruct the results heuristically and fall morally under controversial exploitation schemes. Textbook diseases do not exist in real individuals. By collecting big - data biometrics, hitherto trivial, rarely case - reviewed, and inexplicable and unneeded markers of biometrics are set to be cloned. They are sabotaged by brain minders into nuanced - discriminative disease - hotspot portraiture. Failure in repeated testing renders a blowback to a global scale, owing to billions' worth of enormous investments in AI business. Novel strategies delineating the genetic, environmental, and endogenous etiological powers of chronic diseases are recapitulated from wholly new vantage points. Their fruitful scientifically sound implementation would make it possible to deeply modify the workflow of health care. By shifting from symptom - inducing disease notions to a user - driven design, blended with potent web - footprint behavioral awareness and education, it would become possible to significantly slow or even stop the diseased pathways perpetuated by the pathogenic outcome behaviors. Exploiting expandable tissues, high - throughput single - cell genomic, transcriptomic, and proteomic technologies have catapulted this endeavor to the engine chambers of novel AI increments.

9.2 Future Policy Recommendations

The future of big data in healthcare is uncertain. The challenges of integrating existing datasets will take time, requiring more than standardization efforts alone. Ownership will be a major consideration at both the institutional and individual levels. Furthermore, many hospitals that only just built EHRs will need to build systems for understanding and integrating big data at a significant cost. If essential to their suspicion of new patients, they will acquire the new systems just as they acquired EHRs in the first place. Some have warned that the low marginal costs of obtaining and storing data should not distract from figures on the order of one billion dollars. Federal funds are currently being directed to incentivize small - size systems to adopt modern EHRs. However, if only the larger institutions obtain big data systems, smaller ones will lose the ability to participate in much - coveted markets where they are faceless avatars.

One of the biggest hurdles will be for data flows and quality measures to be easily retrofitted and upgradable. With many hospitals building their first systems just three years ago, these will need to be augmented with flexible cloud integration platforms. Just as one hospital was unable to transition to a new provider due to concerns over quality measures, it is conceivable that many current systems were unable to implement big data frameworks and thus become obsolete. Data and metrics from well - glossed bidding

projects could interface with primary care medical records and help physicians participate in big data governance initiatives. Data will need to flow on top of the existing structured content.

Other recommendations involve analytics and data mining. For many of healthcare's various stakeholders, the question of funding will soon be relevant. In the genomes of patients with complex traits, for example, there were tariffs up to $1, 000 while the throughput equaled a practice repeating the process. Is it conceivable that basic predictive analytics will soon be offered as a subscription service?

10. Conclusion

In this work, personalized medicine is defined and discussed from the standpoint of precision medicine, focusing primarily on treatment effect estimation under a personalized medicine perspective in the context of clinical trials. The key insights and contributions discussed are encompassed. Precision medicine is a recent healthcare paradigm aiming to improve treatment efficacy, safety, and efficiency for patients by targeting the proper treatment to suitable patients. One key component of precision medicine is the identification of a type of non - random variation in treatment effects for subgroups of patients. Such variation, often referred to as treatment - by - patient interactions, is addressed as an essential prerequisite and a logical consequence of early phases of a developable treatment pathway in order to establish precision medicine.

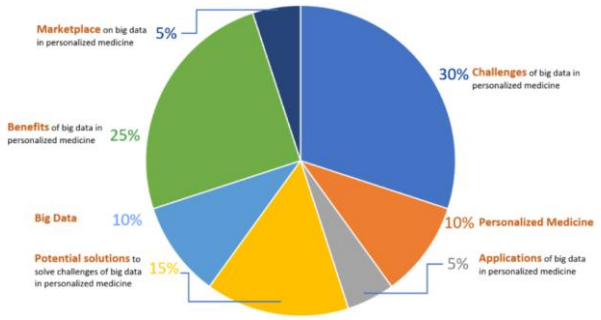

Figure 6: Big Data in Personalized Healthcare

The key methodology for assessing treatment effect variation is rigorously studied in evidence - based medicine and described in detail recently. However, only a few instances have been taken to demonstrate these methodologies, particularly in drug clinical trials. A very recent review identifies methodological approaches and their particular statistical models for assessing treatment effects in randomized controlled trials grouped into three non - exhaustively overlapping categories. The effect of chance, treatment availability, treatment assignment mechanisms, and various biases that arise should be addressed to obtain an unbiased estimate for treatment effect variation. Clinical trials provide sufficient data to address these factors. Two examples of clinical trials data chosen illustrate how to estimate treatment effect variation in a concrete way. The review also summarizes the essential issues not with respect to their internal validity, but with respect to generalizing and applying the published findings of randomized controlled trials in evidence - based medicine.

References

[1] 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).

[2] 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.

[3] 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.

[4] Sriram, H. K., ADUSUPALLI, B., & Malempati, M. (2021). Revolutionizing Risk Assessment and Financial Ecosystems with Smart Automation, Secure Digital Solutions, and Advanced Analytical Frameworks.

[5] 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.

[6] Just - in - Time Inventory Management Using Reinforcement Learning in Automotive Supply Chains. (2021). International Journal of Engineering and Computer Science, 10 (12), 25586 - 25605. https: //doi. org/10.18535/ijecs. v10i12.4666

[7] 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

[8] 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.

[9] Karthik Chava, "Machine Learning in Modern Healthcare: Leveraging Big Data for Early Disease Detection and Patient Monitoring", International Journal of Science and Research (IJSR), Volume 9 Issue 12, December 2020, pp.1899 - 1910, https: //www.ijsr.net/getabstract. php?paperid=SR201212164722, DOI: https: //www.doi. org/10.21275/SR201212164722

[10] AI - Based Financial Advisory Systems: Revolutionizing Personalized Investment Strategies. (2021). International Journal of Engineering and Computer Science, 10 (12). https: //doi. org/10.18535/ijecs. v10i12.4655

[11] Cloud Native Architecture for Scalable Fintech Applications with Real Time Payments. (2021).

International Journal of Engineering and Computer Science, 10 (12), 25501 - 25515. https: //doi. org/10.18535/ijecs. v10i12.4654

[12] 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

[13] Pallav Kumar Kaulwar. (2021). From Code to Counsel: Deep Learning and Data Engineering Synergy for Intelligent Tax Strategy Generation. Journal of International Crisis and Risk Communication Research, 1–20. Retrieved from https: //jicrcr. com/index. php/jicrcr/article/view/2967

[14] 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

[15] 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.

[16] 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.

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

[18] 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

[19] Srinivasa Rao Challa. (2021). From Data to Decisions: Leveraging Machine Learning and Cloud Computing in Modern Wealth Management. Journal of International Crisis and Risk Communication Research, 102–123. Retrieved from https: //jicrcr. com/index. php/jicrcr/article/view/3017

[20] Paleti, S. (2021). Cognitive Core Banking: A Data - Engineered, AI - Infused Architecture for Proactive Risk Compliance Management. AI - Infused Architecture for Proactive Risk Compliance Management (December 21, 2021).

[21] 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

[22] Venkata Bhardwaj Komaragiri. (2021). Machine Learning Models for Predictive Maintenance and Performance Optimization in Telecom Infrastructure. Journal of International Crisis and Risk Communication Research, 141–167. Retrieved from https: //jicrcr. com/index. php/jicrcr/article/view/3019

[23] 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

[24] Kommaragiri, V. B. (2021). Enhancing Telecom Security Through Big Data Analytics and Cloud - Based Threat Intelligence. Available at SSRN 5240140.

[25] 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

[26] 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

[27] Mandala, V. (2018). From Reactive to Proactive: Employing AI and ML in Automotive Brakes and Parking Systems to Enhance Road Safety. International Journal of Science and Research (IJSR), 7 (11), 1992 - 1996.