Deep Science Publishing
https://doi.org/10.70593/978-93-49307-41-4_4

Chapter 4: From data to diagnosis: Leveraging big data and artificial intelligence for personalized healthcare

4.1 Introduction

There has been a surge in digital health applications where contemporary information and communication technologies are used to manage illnesses, health risks and to promote wellness. This includes wearable devices, mobile health, telehealth, telemedicine, etc. This evolution has the promise to improve access to healthcare, reduce inefficiencies, and provide a more personalized (patient-centric) healthcare. Artificial intelligence-based applications with the ability to provide personalized recommendations is an emerging topic in digital healthcare, which will be referred to as digital health AI. Before AI applications can be used in healthcare, they must be "trained" using clinical data or synthetic data. There is a large variety in clinical data. From less sophisticated data such as demographics, to more sophisticated data like medical notes, physical examination and clinical laboratory results. Historically, while there exist some sophisticated types of clinical data, they predominantly have been a privilege of large urban tertiary care centers. Along with the more recent emergence of advanced analytics, machine learning, and AI techniques, there are numerous possibilities for transforming this data into meaningful and actionable results. Nowadays, healthcare stakeholders can use analytical techniques to harness the power of data for analyzing historical data, predicting future outcomes, and determining the best action for the current situation. While tools exist for clinical data exploration, there is a need for more precise and focused data, with more sophisticated or richer content. This can be partly achieved by means of generating synthetic data. Synthetic data can refer to any production data applicable to a given situation that is not obtained by direct measurement.

From a sociomaterial perspective this outlines a series of practices and interventions carried out in the context of biobanks that exemplify broader trends and issues surrounding big data in medicine. Key points span from categorization of diseases to the definition and standardization of biomarkers, and from technologies that capture and produce data to emerging forms of medical data and wearables that are increasingly being embedded into ecosystems. Relevant issues include the multitude and diversity of devices and resulting data in healthcare, enrolled citizens or patients, and research studies. Intertwined with these are data processes related to discovery, collection, analysis, and sharing, which incorporate software, modeling, and statistical infrastructures, ultimately producing outputs to steer clinical practice and policy-making. Throughout these entanglements, biosociality brings apprehension and reshapes patient care, medical science, and the emergence of novel medical-tecno-health ecosystems.

Fig 4.1: AI and Big Data in Personalized Medicine

4.1.1. Background and Significance

In the past decades there has been a substantial evolution in data management and data processing techniques. (Nampalli & Adusupalli, 2024; Nandan, 2024)

On one hand, new data architectures in the form of cloud computing, distributed file systems and distributed databases made the storage and analysis of big data feasible; on the other hand, the healthcare sector, particularly in the Western world, is orienting more and more towards personalized medicine, including the adoption of certification programs promoting awareness and capacity-building. These changes to the traditional model of healthcare along with digital health initiatives driven by both government and market are resulting in an increasing trend for patients to generate and possess their own medical data using wearables, mobile apps, and Web searches. In addition, artificial intelligence (AI) is becoming more and more important — a "broad term for technologies that are capable of performing tasks that typically require human intervention". With medicine being one of the leading sectors, the application of AI technology to health is expected to significantly impact people's lives and the healthcare industry (Kaulwar et al., n.d.; Kumar et al., 2025; Kumar et al., 2025).

Despite the field being of increasing importance, AI in healthcare can be viewed as a trendy research topic, as only very few of the proposed methodologies actually manage to go beyond the research prototype stage and reach a status where they are implemented in clinical practice, characterized by being either proprietary techniques adopted by commercial entities or standardized and open technologies approved by international associations. Thus, the primary aim of this review is to provide an overview of the current research methodologies and to present an open-minded analysis of the clinical and organizational challenges for the application of AI to healthcare.

4.2. Understanding Big Data in Healthcare

Modern healthcare technologies now allow clinicians to "see" biological tissues at an extraordinary level of detail surpassing what is imaginable. Genetic sequencing can now reveal the complete coding sequence of the genome down to the level of individual nucleotides. This level of information is now increasingly used in disease diagnosis and development of treatment plans. In recent years, imaging techniques have been revolutionized by the commercialization of Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) scanners. These scanners can produce exquisitely detailed pictures of the human anatomy, which is routinely used in disease assessment and monitoring. The majority of the world population now carries cameras with them using smartphones, enabling capturing and digitization of visual information unknown to previous generations. Convergence of these trends makes it envisageable that in near future all the interactions, movements, sounds, or conversations can be recorded, creating an unprecedented amount of data generated around every individual.

4.2.1. Definition of Big Data

Objective: In healthcare, big data can be defined as datasets with $\text{Log}(n * p) \geq 7$ which raises legitimate challenges considering current computational methods. It is primarily characterized by its great variety and high velocity. Great variety is demonstrated by numerous, heterogeneous and complex datasets, both structured and unstructured; e.g, heterogeneous kinds of imaging or large-scale medical resources. High velocity results from the rapid rate at which data is created, acquired and shared, but also from data to decisions to action cycle time. High velocity is linked to great variety in that the different data flows to account for are numerous and various. These data characteristics are at the root of further specific problems. Big data is commonly associated with veracity issues, i.e. the need to process and analyse unstructured data but also to trust in analysed results. Veracity accounts for the credibility and reliability of the data. A need to extract meaningful information from the data further challenges the healthcare sector's ability to utilise big data. The need is not simply to analyse data, but also to use it to indicate optimal actions, in order to reduce costs or develop preventative strategies. Those challenges are completed by a broad necessity to share health information, while maintaining both protection of privacy and data security. Modern big data technologies, such as distributed computing or cloud platforms, may not always be fully compliant with those imperatives. Finally, most healthcare organisations are unable to exploit data from an economic perspective. More often, data is stored in different formats and locations; big data raises a further need for new computational methods that optimize data management. There are pressing ethical and legal challenges to address. On the one hand, the concept is often associated with data reuse issues, in favour of false knowledge discovery, the subsequent presentation of misleading results that can direct health policies in the wrong direction. On the other hand, the volume and variety of data being processed can raise serious doubts about patient privacy.

4.2.2. Sources of Healthcare Data

Modern medicine aims more than ever to provide personalized, individual healthcare. Personalization in the healthcare domain has been a rapidly increasing field of interest and effort over the preceding years, as healthcare provision is increasingly oriented towards the individual patient. It implies a level of precision that seeks to treat the patient as opposed to the disease, taking into account a multitude of variables, such as comorbidities, genetic predisposition, and environmental factors.

This introduces a data-processing mechanism that aims to collect, integrate, and harmonize various, otherwise, heterogeneous data sources to realize Holistic Health Records (HHRs) that will provide broader and complete data views, as well as vertically integrated data feeds for better health monitoring, risk prediction, and precise intervention. For the development of such HHRs, an advanced Health Data platform will be developed that will be able to connect and effectively manage the personalized data networks related to the health of the individual person. This platform will develop several advanced data management techniques by integrating cutting-edge Semantic Web and Machine Learning techniques that will efficiently manage the entire data lifecycle and produce personalized health insights.

4.3. Artificial Intelligence in Healthcare

There has been a surge in digital health applications where contemporary information and communication technologies are used to manage illnesses, health risks and to promote wellness. This includes wearable devices, mobile health (mHealth), telehealth, and telemedicine. This evolution has the promise to improve access to healthcare, reduce inefficiencies and provide a more personalized healthcare. Data relevant to a person's health or wellness is automatically captured by products used by the person. This can be from sensing their body, biofeedback from their body, recording their voice, or interpreting images or words. This data is processed in such a way as to provide assistance with health or wellness management. The assistance can range from simple to complex recommendations. In some cases the processing may include the generation of alerts, notifications or alarms. Before AI applications can be used in healthcare, they must be "trained" using clinical or synthetic data. There is a large variety in clinical data, such as demographics, medical notes, physical examinations, and clinical laboratory results. At the same time, analytics are used to improve decision making of different issues. Advances in this field have generated tools that enable the processing of huge data sets to obtain structured information in the form of patterns or trends. Along with the emergence of advanced analytics, machine learning, and artificial intelligence techniques, there are numerous possibilities for transforming this data into meaningful and actionable results. In the healthcare domain, both public and private entities are data rich, and using analytical techniques is conducive to harnessing the power of this data for analyzing historical data, predicting future outcomes and determining the best action for the current situation. (Nampalli & Adusupalli, 2024)

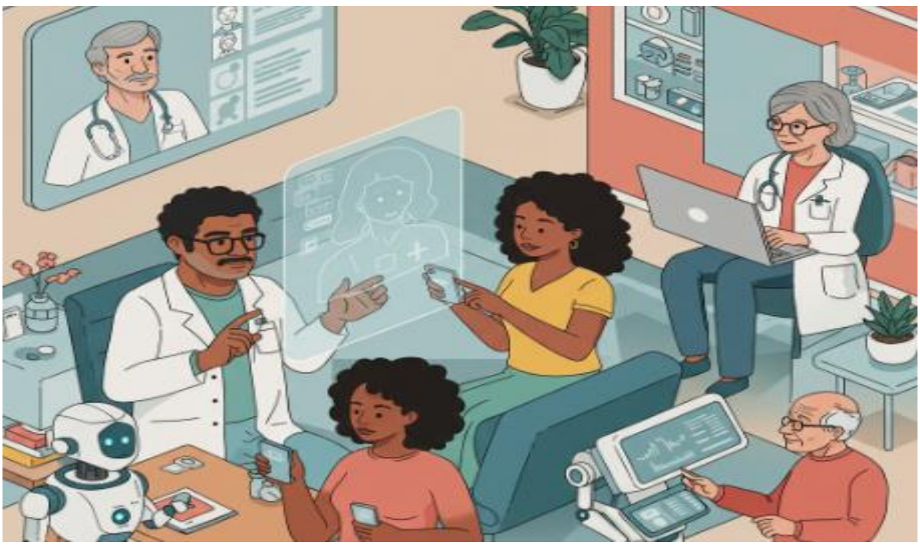

Fig 4.2: AI in Personalized Healthcare

4.3.1. Overview of AI Technologies

There has been a growing interest in the applications of various artificial intelligence (AI) technologies in digital health. This refers to a surge in digital health applications where contemporary information and communication technologies are used to manage illnesses, health risks and promote wellness. There is a growing variety of digital health including wearable devices, mobile health, telehealth, and telemedicine. This evolution has the promise to improve access to healthcare, reduce inefficiencies and provide a more personalized healthcare. However, before AI applications can be used in healthcare, it must be "trained" using clinical or synthetic data. Clinical or real-time data can be obtained using methods or devices used in the healthcare sector.

There is a large variety in clinical data, such as medical notes, demographics, physical examinations, clinical laboratory results, diagnostic results, imaging data and vital signs data. Along with the emergence of advanced analytics, machine learning, and artificial intelligence techniques, there are numerous possibilities to transform this data into meaningful and actionable results. In fact, healthcare stakeholders can leverage analytical techniques to harness the power of data not only for analyzing historical data, but also for predicting future outcomes and determining the best action given the current situation.

Despite the wide availability of clinical data, there is a need for more precise and focused data, which can be achieved by generating synthetic data. Synthetic data refers to any production data applicable to a given situation that is not obtained by direct measurement, but generated to meet specific needs or conditions. In the healthcare industry, synthetic data can be effectively used to measure the performance of a system for a particular patient or to evaluate a newly developed method for the treatment of a disease. Generation of realistic, synthetic, behaviour-based sensor data is a critical step in developing, testing, and validating machine-learning techniques for healthcare applications.

4.4. Integration of Big Data and AI

Recent years have seen a surge in digital health applications where contemporary information and communication technologies are used to manage illnesses, health risks and to promote wellness. This includes but is not limited to wearable devices, mobile health, telehealth, and telemedicine. Although it's early days for many of these applications, they have the promise to improve access to healthcare, reduce inefficiencies and provide more personalized healthcare of potential interest. Before Big Data and AI applications can be used in healthcare, they must be "trained" using clinical, healthcare information records, or synthetic data. There is a great variety in clinical data, such as demographics, medical notes, physical examinations, ECG and X-ray recordings, as well as clinical laboratory results and complication events. Along with advances in the collection of this data, machine learning and artificial intelligence techniques have sped up, and there are numerous possibilities when transforming this data into meaningful and actionable results by using analytical methods.

Nowadays, healthcare stakeholders such as hospitals, insurance vendors, and pharmaceutical firms, are increasingly using analytical techniques for analyzing historical data, predicting future outcomes and determining the best action for the current situation. Moreover, outside of the healthcare domain, academic researchers, start-ups or even governments look into these vast amounts of data, thus creating a competitive niche in the need for insights. When it is widely recognized that the key to personalized medicine, not all clinicians have the necessary tools or data to act upon this. Although there is a wide availability of clinical data, there is also a need for more precise and focused data which can be achieved by generating synthetic data. In general, synthetic data refers to any production data applicable to a given situation that is not obtained by direct measurement. For example, in communications and networking, it is used to generate realistic link metrics and interference.

4.4.1. Data Collection and Preprocessing

Recent advances in artificial intelligence (AI) hold the potential to increase patient safety, augment efficiency, and improve long-term patient outcomes. These AI technologies cover, but are not limited to, aiding physicians in differential diagnosis and selection of treatment, risk prediction and patient stratification, or matching patients with intervention. Furthermore, when fully integrated into the clinician workflow, AI systems can assist in relieving the current strain on patient-clinician-caregiver resources, thereby improving patient and clinician efficiency. It is easy to not appreciate the vast expansion of available AI technologies across almost every facet of society in the last decade and the considerable interest and exponential growth in healthcare data science. The likely strategic expansion and furtherance of these capabilities in the private and public space should continue to be monitored and supported or corrected. Nonetheless, the findings suggest there could be continued investment in AI research, software, data acquisition, civic-policy conduction, cost-benefit assessment, and ethical-public-good frameworks, in competitive alignment within both the global, and local, healthcare landscape. In the dynamic, rapidly transforming spheres of IT and healthcare, the continued adaptation through data and intelligent surveillance, health records, predictive, and network analysis will likely need to be continually adjusted, monitored, and risk-assessed.

However, most sophisticated AI models exist in high-profile, often proprietary publications viewable only to a small audience and implemented by even fewer in a clinical setting. The primary three top challenges to translational research from data science (DS) to patient care are inadequate quality of the raw data, insufficient resources, and high privacy and ethical protection concerning patients. The primary four barriers to use of the EHR data for DS are the data format heterogeneity, structured- and unstructured-data storage in different isolation, data quality, and abundant confounding technology. These and others that are the focus of a current initiative parallel ongoing efforts and will potentially heavily rely on existing and expanding interfaces. Efforts are being placed on implementing newly acquired know-how to improve distillation and care for suboptimal-situation patients. Easier immediate targets could also play an important role, such as influencing clinical practice guidelines, health policies, tender regulations, and the specifics of AI deployment. There are also collaborative efforts towards informally aligning the majority of one's institute's projects to converge on a complementary set of issues around effective countermeasures and possible upshots.

4.5. Personalized Healthcare Approaches

The next wave of healthcare innovation revolves around promoting tailored treatments to patients on an individual basis. Healthcare systems worldwide are currently burdened by unprecedented challenges. The aging global population is rapidly increasing the prevalence of chronic diseases. Inadequacies in healthcare practices perpetuate avoidable medical errors. There are estimated to be over 1.2 million of such errors related to the diagnosis annually in the United States of America. The increasing burden of chronic diseases takes more than half of the worldwide disease prevention budget. Globally, up to a fifth of people endure catastrophic or pushed into extreme poverty due to the personalized healthcare practices. Diversely directed diseases affecting one or more organ systems concurrently are greatly common amongst the community-dwelling older generation in developed countries, who tend to have 2+ disorders that require parallel tending. Centralized therapeutic methods bring forth exaggerated treatment toxicity extensively. Additionally, existing technologies hinder the exchange of biometric information outside the consulting room frequently, formalizing treatment evaluations unfeasible. The development of sophisticated healthcare engineering will help in tackling these issues. Successful business models will be oriented towards ensuring that complete treatments are engineered, ensuring the legitimacy and efficiency of pharmaceuticals and medical devices.

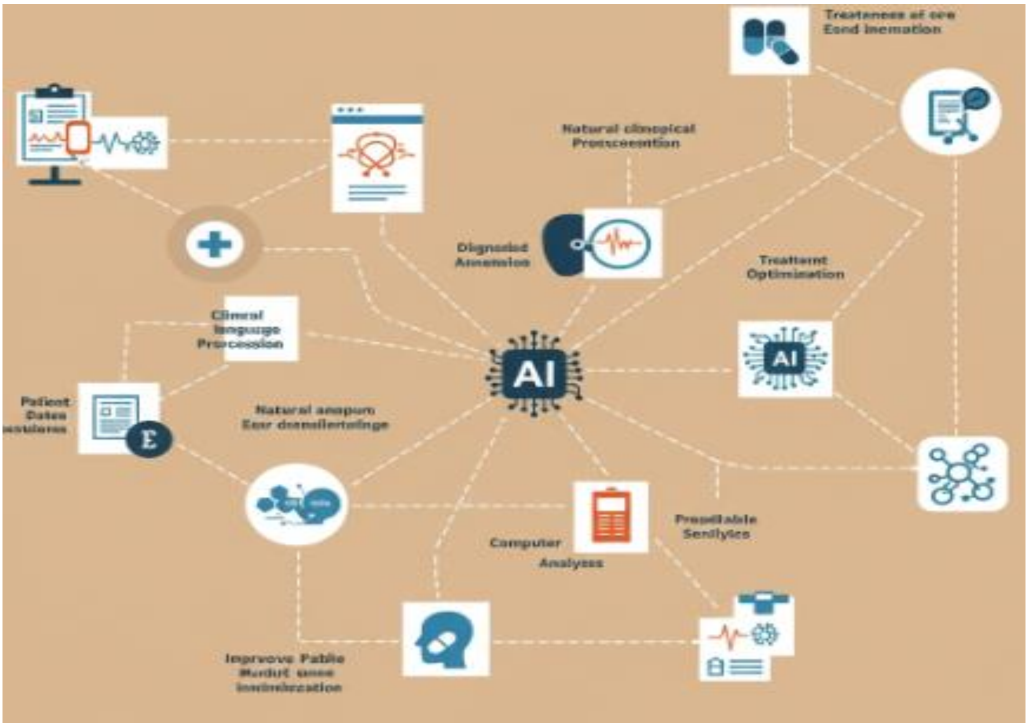

Fig : Medical Big Data and Artificial Intelligence for Healthcare

4.5.1. Definition and Importance of Personalization

Precision medicine, the fine-tuned customization of health care to the individual, is a rapidly advancing field. Although the aspirations of personalized medicine, itself a close synonym, follow and compete with various earlier visions, it is linked most closely in the public mind with the potential of new digital technologies. Often hyped as the starscape of future medicine, some envision a future in which all – or, generous funding permitting, most – aspects of health care are optimized to personal genotypic or phenotypic detail. This includes individualized prevention, diagnostics, and therapy. Moreover, the subsequent rise of great data science and the concurrent rediscovery of older techniques have seen algorithms applied to diverse aspects of health care, from the choice of therapies and dosing, to mental health assessments and intake triaging within physician practices. Though far from delivering on its more ambitious goals – however defined – precision medicine already yields benefits. It also, of course, raises many new questions .

4.6. Ethical Considerations

With the first priority of preventing life and contagion risk in a quickly developing medical crisis, it did not seem practical to ask greater ethical and legal concerns such as privacy and data security. Others counter that hacking, scam and other techno-abuses will bring tragedy and significant societal costs of their own. Covid-monitoring data was said to be like Social Security numbers of a various order at sufficiency. Experiences from a public id coding app illustrate virtual civil controversy and official uncertainty about the moral merits of the entire IT plan in some nations. Several ethics arguments have been provided, from libertarian strings in applause of persons to sacrifice personal privacy for the crowd on a voluntary basis, persons' basic rights to be concerned about the processing of personal data or the notion that privacy is a discipline premise of other arms. Person will avoid getting exposed to and infecting someone sick if they do not suppose they are. To honestly minimize exposure, they need reliable detection components, even if they can be a location of privacy and other risk. While maintaining safety from abuse, the community cannot justify technological response without revealing useful data. In this debate, tracing methods precede the use of an app empowered with spatiotemporal technology. The app uses joint architecture for contact detection and tracking, implying that it satisfies the highest technological benchmarks of confidentiality and privacy by architecture. Ethical and policy choices made hence relate to wider and more recent principles. Authorities must satisfy a sequence of criteria before a tracing tool is acquired, with agencies not possessing advanced technology, promoting the system. Ethics legislation involves using papers, as verifying compliance with the terms would now be 'against national interest.' These eight high-level principles are a manifestation of respect for democracy and human rights, the endorsement of evolution based on evidence, maximum independence, accountability, protection, honesty, implementation, and operability, and must be publicly recorded. Needed criteria may only be met by about four nations in his analysis, which offers a better view of a position attractive to a large majority of the planet.

4.6.1. Data Privacy and Security

The development of big data analytics and artificial intelligence (AI) is an umbrella for high acuity sensing devices and internet of things objects for continuous monitoring of physiology, geomedicine and other health-related issues. That breakthrough depicts extra health status statistics that enhance healthcare analysis and the possibility of being more personalized. It prevails on the introduction of an innovative paradigm of personalized healthcare delivery. Within this innovative context, the continuous data set from various sources contains confidential quantitative information. Data security and privacy are becoming the most problematic right now. An advanced approach is introduced by means of encryption, decryption, and authentication with revolutionary Data Security and Secure Data Authentication. This is a joint approach in which the encryption and decryption systems work together already, and then a secure data authentication phase within an integrated message code is integrated. The proposed enhanced design method adopted Hyper Chaos System 2.

4.7. Conclusion

On the one hand, one could see a larger chunk of structured clinical data in healthcare (e.g., as in electronic health records), and other regular or high-throughput clinical measurements aside from the targeted assays made on bio-samples. Such data is fundamental in diagnosis, follow-up, and benchmarking of disease status, staging and assessing the progress of patients, thereby necessitating well-established methods and guidelines. Health Big Data arising due to such modalities have been expected to lead to improved patient-centered personalized health management with evidence-based healthcare approaches. Here predictive analysis in health management broadly forecasts the future outcomes of patients given historical data through a variety of machine learning and statistical tools in realizing the goal of patient-centred care. Among the various application areas where Big Data is recently adopted for healthcare purposes, "screening" has received relatively less attention so far in both the medical and computer sciences literature. This is somewhat surprising given the potential of Big Data and advanced analysis in it for early detection; experience with some screening procedures has become a combination of standard, frequent, and/or popular practice. In fact, this particular phenomenon creates a bottleneck from the data collection perspective, preventing the design of studies and the analysis of "pre-biopsy" structure to assess the effectiveness and performance in the face of further advances such as AI. This is at least partly because a broad scientific consensus on standard markers and diagnostic techniques is not always immediately reached in the case of certain cancers, and patients are apparently taking an appropriate clinical intervention rather lately in comparison to the first appearance of definitive signs of the disease.

4.7.1. Future Trends

(1) The healthcare sector is facing a significant challenge as the number of tests and investigations has exploded as technology capabilities and types have wildly grown. Practices that are inclined to test heavily are struggling to keep up with the pace of new tests that are released. A financially consequential proportion of these are laborious for the doctor to understand and act upon, leading to deep conservatism and over-referral. Technologies and services that aim to help simplify the derivation of possible diagnoses from presented symptoms reduce the friction of ordering those tests and

make it simple to rapidly send on the patient with a clinical case to another medical party that can do so are fast entering the market.

Such systems reduce the need for a deep understanding of the test on the part of the ordering doctor, are much less immediately helpful to healthcare systems that are struggling to make best use of the care that patients are already given or are looking at ways to make savings via less expensive external investigations. For these aims, there will be a requirement for EMRs or EHRs to be further extended from merely acting as a record of the patient–physician interaction and results of examinations to diagnostic aids in themselves. (2) It is important to note that there are a number of pharmacological and non-pharmacological therapies for conditions that may not present as a presentation in themselves but can be treated nevertheless. Use of big data, machine learning and other AI techniques could feasibly be implemented in this direction though there has been little or no work towards this, likely due to the high legal risks involved. Companies have started to adapt, buying and pursuing the development of fully-integrated platforms encompassing diagnostic, predictive and analytical tools. These tools will likely rely heavily on AI and other machine learning techniques.

References

Kumar, B. H., Nuka, S. T., Malempati, M., Sriram, H. K., Mashetty, S., & Kannan, S. (2025). Big Data in Cybersecurity: Enhancing Threat Detection with AI and ML. Metallurgical and Materials Engineering, 31(3), 12-20. Kaulwar, P. K., Pamisetty, A., Mashetty, S., Adusupalli, B., & Pandiri, L. Harnessing Intelligent Systems and Secure Digital Infrastructure for Optimizing Housing Finance, Risk Mitigation, and Enterprise Supply Networks.

Kumar, S. S., Singireddy, S., Nanan, B. P., Recharla, M., Gadi, A. L., & Paleti, S. (2025). Optimizing Edge Computing for Big Data Processing in Smart Cities. Metallurgical and Materials Engineering, 31(3), 31-39.

Botlagunta Preethish Nandan. (2024). Revolutionizing Semiconductor Chip Design through Generative AI and Reinforcement Learning: A Novel Approach to Mask Patterning and Resolution Enhancement. International Journal of Medical Toxicology and Legal Medicine, 27(5), 759–772. https://doi.org/10.47059/ijmtlm/V27I5/096

Nampalli, R. C. R., & Adusupalli, B. (2024). Using Machine Learning for Predictive Freight Demand and Route Optimization in Road and RailLogistics. Library of Progress-Library Science, Information Technology & Computer, 44(3).