BIG DATA ANALYSIS FOR DISEASE PREDICTION AND PREVENTION

Loso Judijanto *[1]
IPOSS Jakarta, Indonesia
losojudijantobumn@gmail.com

Lola Yustrisia
Fakultas Hukum Universitas Muhammadiyah Sumatera Barat
yustrisialola@gmail.com

Entin Solihah
Persatuan Terapis Gigi dan Mulut Indonesia (PTGMI) Kabupaten Sambas

Abstract

Big Data analytics in the context of disease prediction and prevention stands out as a critical issue in today's digital age, given its unique capacity to process and analyze massive volumes of health data with unprecedented speed and precision. Through the collection of extensive data from various sources, including electronic medical records, wearable devices, and other digital inputs, Big Data analysis enables researchers and healthcare practitioners to identify patterns and trends before diseases develop, forecast outbreaks, and respond proactively to potential health crises. Its ability to integrate and map health data at scale opens up opportunities for smarter prevention and personalized approaches to disease management, significantly shifting the landscape of disease prevention from reactive to proactive, which in turn could save millions of lives and reduce the economic burden on the global health system. The study in this research uses the literature research method. The results show that the use of Big Data and machine learning has great potential in strengthening health systems through disease prediction and prevention. Key findings show that the integration of extensive health data enables more effective identification of disease patterns and trends. With these technologies in place, the ability to diagnose and forecast diseases becomes faster and more accurate, which in turn, can help in designing appropriate and evidence-based interventions. In addition, improved machine learning methods continue to push the boundaries of predictiveness, providing new insights into proactive disease management and prevention.

Keywords: Big Data, Prediction, Disease Prevention.

[1] Correspondence author

Introduction

In recent decades, the world has witnessed an unprecedented explosion of data, which is mainly driven by advances in information and communication technologies. The healthcare sector, as one of the most significant areas for the quality of human life, is no exception to this phenomenon (AL-Rummana et al., 2021). "Big data" in the context of healthcare encompasses huge volumes of information generated from electronic health records, laboratory results, imaging data, genetic data, and various other sources. The potential to utilize this data for disease prediction and prevention is a promising but complex area, which requires detailed analysis (Aljanobi & Lee, 2021).

The potential for utilizing big data in disease prevention lies in its ability to detect health patterns and trends in large populations quickly and accurately. Through big data analysis, health systems can identify disease risk factors early, enabling targeted preventive interventions and personalization of care based on individual risk profiles (Ardakani & Cheshmehzangi, 2023). This capability not only improves the overall effectiveness of disease prevention programs but also contributes to reduced healthcare costs by reducing the need for more expensive and complex medical interventions later on (Ardakani & Cheshmehzangi, 2023). As such, the integration of big data into public health strategies offers a path towards a more proactive approach to healthcare, which can ultimately improve quality of life and patient safety while curbing the growth of healthcare costs globally.

While big data analytics offer significant opportunities for health sector transformation through more accurate disease prediction and more effective preventive health interventions, there are still various challenges to overcome (Arjaria et al., 2021). Concerns about privacy, data security, and ethics, along with technical barriers such as the need for sophisticated data analysis technologies and expertise to interpret results, are often barriers to the full utilization of Health data (Asri & Jarir, 2022).

Additionally, disease prediction and prevention through big data analysis relies on the ability to not only handle the large volume, velocity and diversity of data but also to identify patterns and relationships in the data that may not be immediately apparent. This marks the need for innovative and multidisciplinary approaches in big data analysis (Bonomo et al., 2021).

The importance of disease prediction and prevention is becoming increasingly evident amidst the current global pandemic, where the ability to anticipate and manage the spread of disease is critical to the well-being of Society

(Cahyadi & Forshaw, 2021). Therefore, this review aims to explore how big data analytics can be applied in a healthcare context to improve disease prediction and prevention efforts, discussing the potential, challenges, and solutions that exist in the current literature (Capobianco & Deng, 2022).

Thus, this study is expected to not only provide theoretical and applicative insights for academics and healthcare practitioners but also contribute effective strategies to address public health challenges through technological innovation.

Research Methods

The study conducted in this research uses the literature research method. The literature research method is a systematic approach to collecting, reviewing, and analyzing scholarly publications-such as journals, books, and conference articles-to gain an in-depth understanding of a topic. (Abdussamad, 2022); (Adlini et al., 2022).

Results and Discussion

Definition of Big Data

Big data refers to extremely large and complex data sets, which exceed the processing capabilities of traditional software applications to handle, store, and analyze quickly (Capobianco & Deng, 2022). Its characteristics are often summarized into three 'V's: Volume, which indicates the large size of the data; Variety, which refers to the diverse types and sources of data, including text, images, video, and sensor data; and Velocity, which signifies the speed at which data comes in and must be processed. In its development, the concept has also expanded to include two additional 'V's: Veracity, which relates to the accuracy and trustworthiness of the data, and Value, which emphasizes the importance of extracting business value and useful insights from the data (Dai et al., 2022).

Big data emerged as a result of the digital information explosion triggered by technological advancements and the adoption of mobile devices, social media, the internet of things (IoT), and cloud computing services. Sectors such as business, healthcare, science, and government generate ever-increasing volumes of data every day, offering unprecedented potential to analyze patterns, trends, and relationships in that data (Debal & Sitote, 2022). For example, in healthcare, big data analysis can optimize patient care outcomes by predicting diseases and tailoring treatments based on analysis from large clinical datasets (Deshpande & Pham, 2022).

The challenge in big data lies not only in its collection and storage, but especially in data analysis and interpretation. This requires advanced technologies and data analysis methods, including machine learning and predictive analytics, to uncover actionable insights and provide a competitive advantage (Dritsas & Trigka, 2022). In addition, concerns regarding data privacy and security demand the development of strong legal and ethical frameworks. With the immense potential it offers, big data continues to redefine how organizations, in both the public and private sectors, operate and make strategic decisions, making it an essential pillar in the modern information economy (Duan, 2022).

The utilization of big data excels in identifying correlations, uncovering hidden trends, and providing detailed analytical insights, which used to be impossible or very difficult to achieve with smaller, heterogeneous data volumes. For example, in retail, big data analytics can use customer data traces to personalize the shopping experience and improve customer satisfaction, while in finance, it can be used to detect fraud in real-time by analyzing transaction patterns (ED-DAOUDY et al., 2022). In smart cities, data from sensors and IoT are used to optimize traffic, waste management, and energy consumption. Each of these examples demonstrates the power of big data in transforming data into evidence-based policies or strategic decisions (Farashah et al., 2021).

However, utilizing big data also comes with significant challenges, ranging from technical issues such as data storage, processing, and security, to ethical concerns regarding privacy and bias in data and algorithms. Privacy considerations become especially critical when personal data is involved, triggering the need for regulations such as the General Data Protection Regulation (GDPR) in the European Union (Golande & Pavankumar, 2023). On the other hand, machine learning algorithms and models that process big data can reproduce or even reinforce bias and discrimination if not properly designed or trained, which demands a more responsible and transparent approach to their development (Hansun et al., 2022).

In conclusion, big data represents an information revolution that has the potential to improve economic activity, innovation and quality of life. However, achieving that potential requires more than just advanced technology; it also requires a holistic approach involving policy, ethics and education to navigate the complexities inherent to handling big data. Successful implementation of big data solutions requires a balance between the technical ability to manage and analyze data and consideration of the social and ethical implications of using that data. Therefore, to maximize the benefits and minimize the accompanying risks, collaboration between disciplines, industries, and regulators will certainly be key in the continued evolution of the big data landscape.

Characteristics and Types of Big Data

Big data is recognized through several key characteristics that distinguish it from traditional data sets, known as the "3 Vs":

First, Volume: The quantity of data generated today is so large that it is one of the hallmarks of big data. Volume refers to the size of the data that must be processed. This can range from terabytes to exabytes of data and beyond. These large volumes arise from various sources such as social media, sensors, satellites, videos, transaction logs, and more. Large volumes of data present challenges in terms of storage, analysis, and management. Second, Variety: Big data includes different types of data that can come in a variety of formats (Hu et al., 2021). These varieties include structured data (such as those found in relational databases), semi-structured data (such as XML, JSON), and unstructured data (such as text, images, audio, and video). In the context of big data, this component emphasizes the complexity of standardizing, cleaning, and consolidating different types of data for effective analysis. Third, Velocity: Velocity refers to the speed at which new data is generated and the rate at which data moves. This can include real-time data that must be processed quickly, such as information coming from IoT sensors or the flow of clicks on a website. Speed demands systems that are able to collect, process, and make decisions based on data in near real-time (Jordan et al., 2021).

In addition to the basic "3 Vs", there are also additional characteristics that are sometimes mentioned, such as Veracity which expresses uncertainty in the data (such as accuracy, trustworthiness, and credibility), and Value which relates to the ability to extract useful business intelligence value from big data (K.MANOHARI, 2023).

Types of big data can be classified depending on the source: 1) Social Data: This data comes from social media platforms and between humans, including textual found in tweets, blogs, status updates, and comments. 2) Transaction Data: This is transaction history data typically found in financial records, purchase logs, and sales. 3) Sensor or IoT Data: Generated by sensors or nodes in an IoT network, this includes weather data, traffic, telemetrics, health data, and others. 4) Open Data: Big data sets that are freely available to the public by governments, research organizations, and businesses fall under this category (Kojima et al., 2022).

Each type of data contributes to the complexity of big data, requiring specialized tools to extract useful information and provide insights that can drive more efficient business and operational decisions.

Technology and Tools in Big Data Analytics

To meet the challenges of big data analysis, various technologies and tools have been developed to manage, process, and distill insights from these large and complex volumes of data. Among the most well-known is Apache Hadoop, an open-source framework that enables distributed data processing through computer clusters using a simple programming model (Kumar et al., 2023). Hadoop is designed to handle large volumes of data in a highly scalable manner. The Hadoop ecosystem includes various components such as Hadoop Distributed File System (HDFS) for storage, YARN for resource management and job scheduling, and MapReduce for distributed data processing. While Hadoop has been a pillar in the big data ecosystem, technological evolution has introduced other frameworks such as Apache Spark, which offers faster data processing thanks to its in-memory processing capabilities and support for real-time analytics (Kumar et al., 2023).

Meanwhile, for large-scale data storage and management, solutions such as NoSQL databases (e.g., MongoDB, Cassandra, and Couchbase) are often adopted due to their horizontal scaling capabilities, schema flexibility, and high performance in the face of intensive read/write loads. NoSQL databases provide an elastic alternative to more rigid relational databases, making them more suitable for the different types of big data that must be managed and analyzed (Lbrini et al., 2021).

On the analytics side, tools like Apache Kafka are used for efficient management of streaming data, while advanced analytics tools like Apache Flink focus on real-time data processing for applications that require quick responses, such as recommendation systems or fraud detection (Lei, 2024). For data exploration and visualization, tools such as Tableau, QlikView, or Power BI enable the creation of interactive dashboards and visualizations that aid in understanding data and sharing insights. At the predictive analytics and machine learning stage, platforms such as TensorFlow, PyTorch, and scikit-learn are used to build and train predictive models that can leverage large volumes of data to forecast future trends, optimize operations, or identify previously unseen patterns (Leinonen, 2021).

Overall, the ecosystem of technologies and tools in big data analytics is vast and growing, with each tool having specific strengths and advantages depending on the data analysis needs. Choosing the right technology is crucial and often depends on the type of data at hand, the scale of operations, and the type of insights to be extracted.

Definition and Importance of Disease Prediction and Prevention

Disease prediction and prevention are key elements in efforts to strengthen health systems and improve individuals' quality of life. These two aspects synergize to not only identify health risks before they occur but also implement strategies to avoid the development of these diseases (Yang et al., 2021). Disease prediction involves using a patient's clinical, historical, and genetic data to estimate the risk of a particular health condition occurring in the future. This is done through analyzing data trends, identifying patterns, and using statistical or machine learning models that can associate certain genetic markers or health indicators with the potential for disease occurrence. The goal is to provide accurate predictive information to healthcare professionals so that interventions can be made earlier (Manjula et al., 2021).

The importance of this approach is particularly evident in the management of chronic diseases such as diabetes, hypertension, and heart disease, where early detection and preventive intervention can significantly reduce the risk of severe complications or mortality. By identifying individuals at high risk of a particular disease, preventive measures such as lifestyle changes, nutritional management, and drug therapy can be taken to delay or even completely avoid disease onset (Mishra et al., 2022). This predictive and preventive approach also emphasizes the use of vaccinations for the prevention of infectious diseases, demonstrating effectiveness in reducing the prevalence and impact of diseases such as influenza, measles, and HPV on the population (Miyoshi, 2021).

Furthermore, in the context of disease prevention, health information technology, including electronic medical record systems and big data analytics tools, has transformed the way we understand and manage population health (Montesi et al., 2020). With extensive data analysis, researchers and health professionals can now identify health trends, risk factors, and social determinants of health that influence disease at a larger scale. This enables the development of targeted public health programs, preventive interventions, and more effective educational campaigns. Through disease prediction and prevention, we are able to transition from a reactive model of care to a proactive model of health, where prevention is key to reducing the burden of disease, improving quality of life, and optimizing the use of health resources (Muthulakshmi & Parveen, 2024).

In conclusion, disease prediction and prevention are fundamental components in improving health system efficiency and individual quality of life. Through predictive methods, we can identify disease risks early and implement preventive measures to reduce the chances of developing the disease. This approach not only helps in chronic disease management but also in the prevention of infectious diseases through vaccination and public health interventions. With the support of health information technology and big data analytics, disease predictions are becoming more accurate and preventive interventions more targeted, leading us to a more proactive era of healthcare. Implementing these strategies widely can significantly reduce the burden of disease in society, maximize the allocation of health resources, and promote healthier and longer lives for the population.

Predictive Modeling in Healthcare Based on Big Data

Predictive models in healthcare based on big data are a significant breakthrough that allows healthcare professionals to make decisions based on extensive and comprehensive information. By collecting and analyzing large volumes of data from various sources-from electronic medical records, laboratory results, genetic data, to information about patients' lifestyles-these models are able to identify patterns, trends, and relationships that have never been seen before (Xu et al., 2023). The use of machine learning algorithms and artificial intelligence in predictive models further improves the accuracy of predictions regarding the likelihood of disease occurrence, the progression of a patient's health condition, and the patient's response to certain treatments. This enables a more personalized medical approach, where treatment and health recommendations can be tailored specifically to individual health needs and risks (Nibareke & Laassiri, 2020).

Furthermore, the application of big data-based predictive models goes beyond individual disease prediction, making a substantial impact in epidemiological research and population health management. With the ability to process and analyze data from multiple populations on a large scale, predictive models aid in the identification of lifestyle-related, environmental, and genetic risk factors that contribute to the emergence of chronic diseases in the population (Nikiforakis, 2021). This facilitates the development of more effective and targeted public health prevention strategies, allowing health authorities to more efficiently allocate resources and interventions. Thus, big data-based predictive models not only promise advances in the care and management of individual health but also in the improvement of public health at large (Ning, 2021).

In conclusion, the use of big data-driven predictive models in the healthcare sector provides the ability to make more accurate predictions and more personalized interventions for individual risks and diseases. This enables the delivery of care that is tailored to a patient's specific genetic profile and risk factors. More broadly, these models support public health research by identifying risk factors for chronic diseases and guiding the development of population-focused prevention strategies. As a result, big data-based predictive models have great potential to improve the effectiveness of personal healthcare and overall public health, while optimizing the allocation of healthcare resources.

Technology Innovation and Latest Analytics Tools

Recent technological innovations have changed the landscape of data analysis, with the development of new tools and platforms that are more sophisticated and accessible. One prime example is artificial intelligence (AI) and machine learning, which are now an integral part of current analytics tools (Ordonez et al., 2022). AI allows computers to 'learn' from the data fed into them, automatically identifying patterns and making predictions without explicit programming. This enables companies and organizations to process and analyze huge volumes of data with unprecedented speed and accuracy. AI-based analytics tools such as tensor processing units (TPUs) and deep learning frameworks have accelerated computing and analytics capabilities, and these continue to grow (Pejic-Bach et al., 2022).

In addition, cloud-based data analytics platforms have made big data analytics more affordable and scalable for businesses of all sizes. Services such as Amazon Web Services, Google Cloud Platform, and Microsoft Azure offer widely customizable infrastructure as a service (IaaS), as well as platform as a service (PaaS) equipped with various tools for data analysis and processing. Thanks to secure cloud storage and real-time collaboration, geographically separated teams can now work together on the same dataset efficiently, allowing organizations to draw insights from their data and apply them to make more informed and rapid decisions (Pontes & Benjannet, 2021).

In the field of data visualization, innovations such as Tableau, Qlik, and Power BI enrich the way we understand and communicate with data. These tools bring numbers to life through interactive dashboards, engaging graphs, and informative heat maps (Prasad, 2021). With intuitive user interfaces and drag-and-drop capabilities, data visualization tools allow users from any discipline to explore the complexity of data and discover important insights. These tools are also often integrated with AI and machine learning to provide smarter visualization recommendations, making it easier for users to interpret data and make evidence-based decisions (Prehofer & Mehmood, 2020).

Augmented intelligence is another recent innovation that combines human intelligence with artificial intelligence to produce better results in data analysis. These tools are designed to enhance human capabilities rather than replace them, providing enhanced insights and analysis that help users make more informed and strategic decisions (R. et al., 2022). Examples of implementations of this include systems that detect anomalies in financial transactions or industrial processes, and predictive analytics platforms that aid in resource planning. With the ability to learn continuously and adapt to new inputs and changing situations, Augmented Intelligence reduces errors, increases efficiency, and often simplifies complex processes for business practitioners (Rawat et al., 2023).

Bockchain also plays an important role in analytics innovation, by facilitating high data security and integrity during the analysis process. As a distributed ledger technology, blockchain allows data to be stored in many decentralized nodes, making it almost impossible to manipulate (Roberts & Segev, 2020). This is particularly useful in the context of data transparency and auditability in highly regulated industries such as banking, healthcare, and supply chain. The implementation of blockchain in data analysis tools not only increases trust in the data being analyzed, but also strengthens security in sharing data between entities (Ruan et al., 2020)..

Challenges and Barriers in Implementing Big Data for Disease Prevention

The application of big data in disease prevention offers great potential, but also faces various challenges and obstacles. One of the biggest challenges is the issue of data privacy and security. Health data is highly sensitive and personal information (Sasikala & Sheela, 2020). When health data is collected, stored, and analyzed for big data purposes, the risk of data breaches increases. Patients may hesitate to share their information if they are not confident that the data will be kept safe. Compliance with standards and regulations such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General

Data Protection Regulation (GDPR) in the European Union is an essential need to maintain patient privacy (Sewell, 2022).

Interoperability and data standardization are also barriers to the application of big data in healthcare. Health data is often fragmented and stored in various incompatible formats between different health systems and facilities. This hinders the ability of analytics systems to integrate and interpret data from different sources. Deficiencies in data standardization can lead to inaccurate or incomplete conclusions, which in turn can affect disease prevention efforts (Shahrivari et al., 2022).

Furthermore, quantity of data does not necessarily mean quality. Another major challenge is ensuring the quality, accuracy and relevance of data used for big data analysis in healthcare. Incomplete, inaccurate, or biased data can lead to misleading insights and may adversely affect preventive interventions. Therefore, there must be an intensive effort to validate and cleanse data before it is used in big data models, which of course requires significant resources, time, and expertise (Shankhdhar, 2022).

Given the complexity of big data, implementing it on a large scale requires adequate resources and trained manpower. Human resources competent in big data are often limited, especially in the healthcare domain in many countries. Ensuring that healthcare teams are equipped with the skills to manage and analyze big data is an important issue to address (Shinde et al., 2022). Additionally, implementing a technology infrastructure capable of managing and analyzing large and growing volumes of data is another challenge that requires significant investment from healthcare institutions (Shivani & Rao, 2021).

Addressing these challenges requires close collaboration between stakeholders in various sectors, including the government, technology industry, and medical community. Governments can play an important role by creating regulations that strike a balance between privacy protection and the use of data for healthcare research (Simon, 2022). These regulations should be flexible enough to accommodate technological advancements and innovations, while still maintaining high standards of data security and privacy. For interoperability issues, standardization of health data formats and data exchange protocols is needed so that different systems can communicate efficiently (Singh et al., 2023).

In the area of human resource development, investment in the education and training of healthcare professionals and data analysts is needed. Educational programs should be designed to equip students with expertise in big data analytics as well as a deep understanding of ethics and data privacy. These efforts can be strengthened by providing ongoing training and certification for working professionals, so that they remain relevant to the latest developments in healthcare technology and practice (Skretting & Gronli, 2020).

Also, to address technological infrastructure challenges, the adoption of cloud computing models and AI-based solutions can provide a more efficient and scalable way to manage big data. This model offers advantages in terms of data storage, processing capacity, and flexibility, allowing healthcare institutions to access and analyze data from anywhere, anytime. It also facilitates easier collaboration between institutions, allowing for large-scale research and integrated data analysis from multiple sources (Sorell, 2020).

Therefore, while the challenges are significant, through collaboration, innovation and investment in the right resources, the potential of big data in driving disease prevention efforts can be maximized. This includes not only applying new technologies and methods but also ensuring that the humanitarian and ethical aspects of big data utilization are not overlooked. In this way, big data can be a transformational force in disease prevention, leading to a more responsive, effective and patient-centered health system.

Conclusion

Big Data analytics has become a vital instrument in the healthcare field, especially for the purpose of disease prediction and prevention. Research activities aimed at collecting, storing, and analyzing health data by utilizing Big Data technology provide great opportunities in understanding disease trends or patterns, which in turn helps in devising more effective disease prevention strategies. With the ability to process huge amounts of data from various sources, this technology enables information mining and visualization of trends or discovery of certain phenomena at high processing speeds, whether it is related to past, current, or predicted data for the future.

Research in this field not only focuses on analyzing disease trends, but also includes the development of machine learning methods to improve the accuracy of disease prediction and prevention. Web-based applications of these methods offer widespread access and democratize disease prevention, providing tools for communities and governments to reduce the social and economic burden caused by disease. As such, Big Data analysis and machine learning are becoming an important part of identifying health risks and enabling timely interventions to prevent disease progression.

The integration of Big Data in health research bridges the gap between big data and informative health policies. Applying the insights gained from analyzing this data enables the development of more proactive public health strategies, targeting health problems before they develop into larger issues. This phase is a joint effort between researchers, health practitioners, and policymakers to optimize the use of data to create a healthier society and be able to tackle future health challenges more effectively.

References

Abdussamad, Z. (2022). *Buku Metode Penelitian Kualitatif*. *Query date: 2024-05-25 20:59:55*. https://doi.org/10.31219/osf.io/juwxn

Adlini, M. N., Dinda, A. H., Yulinda, S., Chotimah, O., & Merliyana, S. J. (2022). Metode Penelitian Kualitatif Studi Pustaka. *Edumaspul: Jurnal Pendidikan*, *6*(1), 974–980. https://doi.org/10.33487/edumaspul.v6i1.3394

Aljanobi, F. A., & Lee, J. (2021). Topological Data Analysis for Classification of Heart Disease Data. *2021 IEEE International Conference on Big Data and Smart Computing (BigComp)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/bigcomp51126.2021.00047

AL-Rummana, G. A., Al-Ahdal, A. H. A., & Shinde, G. N. (2021). The Role of Big Data Analysis in Increasing the Crime Prediction and Prevention Rates. *Intelligent Data Analytics for Terror Threat Prediction*, *Query date: 2024-06-09 05:26:26*, 209–220. https://doi.org/10.1002/9781119711629.ch10

Ardakani, S. P., & Cheshmehzangi, A. (2023). On-Board Unit Freight Transport Data Analysis and Prediction: Big Data Analysis for Data Pre-processing and Result Accuracy. *Big Data Analytics for Smart Transport and Healthcare Systems*, *Query date: 2024-06-09 05:26:26*, 45–61. https://doi.org/10.1007/978-981-99-6620-2_3

Arjaria, S. K., Rathore, A. S., & Cherian, J. S. (2021). Kidney disease prediction using a machine learning approach: A comparative and comprehensive analysis. *Demystifying Big Data, Machine Learning, and Deep Learning for Healthcare Analytics*, *Query date: 2024-06-09 05:26:26*, 307–333. https://doi.org/10.1016/b978-0-12-821633-0.00006-4

Asri, H., & Jarir, Z. (2022). *Toward a Smart Health: Reality Mining and Big Data Analytics for Real-Time Disease Prediction*. *Query date: 2024-06-09 05:26:26*. https://doi.org/10.21203/rs.3.rs-1621912/v1

Bonomo, M., Placa, A. L., & Rombo, S. E. (2021). Prediction of Disease–lncRNA Associations via Machine Learning and Big Data Approaches. *Knowledge*

Modelling and Big Data Analytics in Healthcare, *Query date: 2024-06-09 05:26:26*, 203–226. https://doi.org/10.1201/9781003142751-14

Cahyadi, & Forshaw, M. (2021). Hard Disk Failure Prediction on Highly Imbalanced Data using LSTM Network. *2021 IEEE International Conference on Big Data (Big Data)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/bigdata52589.2021.9671555

Capobianco, E., & Deng, J. (2022). Editorial: Big Data Analytics for Precision Health and Prevention. *Frontiers in Big Data*, *4*(Query date: 2024-06-09 05:26:26). https://doi.org/10.3389/fdata.2021.835353

Dai, P., Chen, Y., & Feng, Y. (2022). Big Data Analysis of Applying Artificial Intelligence to Criminal Justice and Their Prevention. *2022 International Conference on Computation, Big-Data and Engineering (ICCBE)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/iccbe56101.2022.9888156

Debal, D. A., & Sitote, T. M. (2022). Chronic kidney disease prediction using machine learning techniques. *Journal of Big Data*, *9*(1). https://doi.org/10.1186/s40537-022-00657-5

Deshpande, U., & Pham, T. (2022). Storage Capacity Prediction using Population Analytics. *2022 IEEE International Conference on Big Data (Big Data)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/bigdata55660.2022.10020706

Dritsas, E., & Trigka, M. (2022). Machine Learning Techniques for Chronic Kidney Disease Risk Prediction. *Big Data and Cognitive Computing*, *6*(3), 98–98. https://doi.org/10.3390/bdcc6030098

Duan, Y. (2022). Statistical Analysis and Prediction of Employee Turnover Propensity Based on Data Mining. *2022 International Conference on Big Data, Information and Computer Network (BDICN)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/bdicn55575.2022.00052

ED-DAOUDY, A., Maalmi, K., & ouaazizi, A. E. (2022). *A scalable and real-time system for disease prediction using big data processing*. *Query date: 2024-06-09 05:26:26*. https://doi.org/10.21203/rs.3.rs-1567163/v1

Farashah, M. V., Etebarian, A., Azmi, R., & Dastjerdi, R. E. (2021). A hybrid recommender system based-on link prediction for movie baskets analysis. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00422-0

Golande, A. L., & Pavankumar, T. (2023). Optical electrocardiogram based heart disease prediction using hybrid deep learning. *Journal of Big Data*, *10*(1). https://doi.org/10.1186/s40537-023-00820-6

Hansun, S., Wicaksana, A., & Khaliq, A. Q. M. (2022). Multivariate cryptocurrency prediction: Comparative analysis of three recurrent neural networks approaches. *Journal of Big Data*, *9*(1). https://doi.org/10.1186/s40537-022-00601-7

Hu, H., Xin, Y., Wu, X., Bao, C., Yang, L., & Bai, L. (2021). Research on Analysis and Prediction of Big Data of Chinese Medicinal Materials in R+Hadoop. *2021*

International Conference on Artificial Intelligence, Big Data and Algorithms (CAIBDA), *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/caibda53561.2021.00044
Jordan, G., Brimicombe, A., & Li, Y. (2021). *Big Data Quantitative Risk Analysis Method for Machine Health Indicator Prediction*. *Query date: 2024-06-09 05:26:26*. https://doi.org/10.14293/s2199-1006.1.sor-.pp55dcw.v1
K.MANOHARI. (2023). An Efficient Exploration on Big Data Analysis in Adolescent Diabetic Prediction with Deep Learning Techniques. *International Journal of Information Technology, Research and Applications*, *2*(2), 33–40. https://doi.org/10.59461/ijitra.v2i2.51
Kojima, R., Legaspi, R., & Wada, S. (2022). Trip Destination Prediction by Cross-City Exploratory Data Analysis Approach in People Flow Data. *2022 IEEE International Conference on Big Data (Big Data)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/bigdata55660.2022.10020611
Kumar, A., Singh, K. U., & Kumar, M. (2023). A Clinical Data Analysis Based Diagnostic Systems for Heart Disease Prediction Using Ensemble Method. *Big Data Mining and Analytics*, *6*(4), 513–525. https://doi.org/10.26599/bdma.2022.9020052
Lbrini, S., Fadil, A., Aamir, Z., Khomali, M., Oulidi, H. J., & Rhinane, H. (2021). Big Health Data: Cardiovascular Disease Prevention Using Big Data and Machine Learning. *Studies in Computational Intelligence*, *Query date: 2024-06-09 05:26:26*, 311–327. https://doi.org/10.1007/978-3-030-72065-0_17
Lei, L. (2024). Research on Disease Prediction Based on Daily Data Analysis of the Elderly Based on Neural Network Algorithms. *2024 IEEE 3rd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/eebda60612.2024.10485910
Leinonen, J. (2021). Improvements to short-term weather prediction with recurrent-convolutional networks. *2021 IEEE International Conference on Big Data (Big Data)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/bigdata52589.2021.9671869
Manjula, G., Gopi, R., Rani, S. S., Reddy, S. S., & Chelvi, E. D. (2021). Firefly—Binary Cuckoo Search Technique based heart disease prediction in Big Data Analytics. *Applications of Big Data in Healthcare*, *Query date: 2024-06-09 05:26:26*, 241–260. https://doi.org/10.1016/b978-0-12-820203-6.00007-2
Mishra, S., Pandey, M., Rautaray, S. S., & Chakraborty, S. (2022). Multiclass Prediction of Heart Disease Patients Using Big Data Analytics. *Studies in Big Data*, *Query date: 2024-06-09 05:26:26*, 177–193. https://doi.org/10.1007/978-981-19-5154-1_11
Miyoshi, T. (2021). *Fusing Big Data and Big Computation in Numerical Weather Prediction*. *Query date: 2024-06-09 05:26:26*. https://doi.org/10.52843/cassyni.n6qqk0

Montesi, D., Girdzijauskas, S., & Vlassov, V. (2020). Repeating Link Prediction over Dynamic Graphs. *2020 IEEE International Conference on Big Data (Big Data)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/bigdata50022.2020.9378360

Muthulakshmi, P., & Parveen, M. (2024). Big Data Analytics for Heart Disease Prediction using Regularized Principal and Quadratic Entropy Boosting. *Indian Journal Of Science And Technology*, *17*(6), 533–547. https://doi.org/10.17485/ijst/v17i6.2928

Nibareke, T., & Laassiri, J. (2020). Using Big Data-machine learning models for diabetes prediction and flight delays analytics. *Journal of Big Data*, *7*(1). https://doi.org/10.1186/s40537-020-00355-0

Nikiforakis, G. (2021). The Use of Data Collection and Big Data Analysis in Neurodegenerative Disease Prevention. *GeNeDis 2020*, *Query date: 2024-06-09 05:26:26*, 181–181. https://doi.org/10.1007/978-3-030-78775-2_21

Ning, F. (2021). Prediction and Detection of Urban Trajectory Using Data Mining and Deep Neural Network. *2021 International Conference on Big Data Analysis and Computer Science (BDACS)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/bdacs53596.2021.00017

Ordonez, C., Fund, I., & Bellatreche, L. (2022). Comparing Association Rules and Deep Neural Networks for Heart Disease Prediction. *2022 IEEE International Conference on Big Data (Big Data)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/bigdata55660.2022.10020522

Pejic-Bach, M., Pivar, J., & Krstić, Ž. (2022). Big Data for Prediction. *Research Anthology on Big Data Analytics, Architectures, and Applications*, *Query date: 2024-06-09 05:26:26*, 1192–1215. https://doi.org/10.4018/978-1-6684-3662-2.ch057

Pontes, E. L., & Benjannet, M. (2021). Contextual Sentence Analysis for the Sentiment Prediction on Financial Data. *2021 IEEE International Conference on Big Data (Big Data)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/bigdata52589.2021.9672027

Prasad, Dr. S. (2021). Machine Learning Era In Heart Disease Prediction- An Intense Learning Analysis With Big Data. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(5), 203–208. https://doi.org/10.17762/turcomat.v12i5.874

Prehofer, C., & Mehmood, S. (2020). Big Data Architectures for Vehicle Data Analysis. *2020 IEEE International Conference on Big Data (Big Data)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/bigdata50022.2020.9378397

R., R., P., K., E., M., V., K., & Pon, H. (2022). Disease Analysis and Prediction Using Digital Twins and Big Data Analytics. *New Approaches to Data Analytics and Internet of Things Through Digital Twin*, *Query date: 2024-06-09 05:26:26*, 98–114. https://doi.org/10.4018/978-1-6684-5722-1.ch005

Rawat, V., Singh, D. P., Singh, N., & Negi, S. (2023). Heart Disease Prediction Using Machine Learning and Big Data. *Big Data, Cloud Computing and IoT*, *Query date: 2024-06-09 05:26:26*, 93–102. https://doi.org/10.1201/9781003298335-7

Roberts, H., & Segev, A. (2020). Animal Behavior Prediction with Long Short-Term Memory. *2020 IEEE International Conference on Big Data (Big Data)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/bigdata50022.2020.9378184

Ruan, J., Wu, W., & Luo, J. (2020). Stock Price Prediction Under Anomalous Circumstances. *2020 IEEE International Conference on Big Data (Big Data)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/bigdata50022.2020.9378030

Sasikala, P., & Sheela, L. M. I. (2020). Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS. *Journal of Big Data*, *7*(1). https://doi.org/10.1186/s40537-020-00308-7

Sewell, D. K. (2022). Network-Informed Constrained Divisive Pooled Testing Assignments. *Frontiers in Big Data*, *5*(Query date: 2024-06-09 05:26:26). https://doi.org/10.3389/fdata.2022.893760

Shahrivari, H., Papapetrou, O., & Fletcher, G. (2022). Workload Prediction for Adaptive Approximate Query Processing. *2022 IEEE International Conference on Big Data (Big Data)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/bigdata55660.2022.10020614

Shankhdhar, A. (2022). Visualization and Prediction of Heart Disease using Big Data Analytics. *2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/smart55829.2022.10046877

Shinde, S., Satav, S., Shirole, U., & Oak, S. (2022). Comprehensive Analysis of Parkinson Disease Prediction using Vocal Parameters. *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/com-it-con54601.2022.9850857

Shivani, B., & Rao, S. P. G. (2021). Stock Market Analysis & Prediction. *2021 International Conference on Forensics, Analytics, Big Data, Security (FABS)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/fabs52071.2021.9702549

Simon, S. D. (2022). Centers for Disease Control and Prevention (CDC). *Encyclopedia of Big Data*, *Query date: 2024-06-09 05:26:26*, 158–161. https://doi.org/10.1007/978-3-319-32010-6_258

Singh, A., Vij, D., Jijja, A., & Verma, S. (2023). Prediction of Heart Disease Using Various Data Analysis and Machine Learning Techniques. *Springer Proceedings in Mathematics & Statistics*, *Query date: 2024-06-09 05:26:26*, 23–35. https://doi.org/10.1007/978-3-031-15175-0_3

Skretting, A., & Gronli, T.-M. (2020). Baseline for Performance Prediction of Android Applications. *2020 IEEE International Conference on Big Data (Big Data)*, *Query date: 2024-06-09 05:26:26*. https://doi.org/10.1109/bigdata50022.2020.9377882

Sorell, T. (2020). Policing with Big Data: DNA Matching vs Crime Prediction. *Big Data and Democracy*, *Query date: 2024-06-09 05:26:26*, 57–70. https://doi.org/10.3366/edinburgh/9781474463522.003.0005

Xu, J., Liu, J., Yao, T., & Li, Y. (2023). Prediction and Big Data Impact Analysis of Telecom Churn by Backpropagation Neural Network Algorithm from the Perspective of Business Model. *Big Data*, *11*(5), 355–368. https://doi.org/10.1089/big.2021.0365

Yang, Y., Li, Y., Chen, R., Zheng, J., Cai, Y., & Fortino, G. (2021). Risk Prediction of Renal Failure for Chronic Disease Population Based on Electronic Health Record Big Data. *Big Data Research*, *25*(Query date: 2024-06-09 05:26:26), 100234–100234. https://doi.org/10.1016/j.bdr.2021.100234