Research Report

INF20010 - Database Systems

Student Name: Nguyen Nam Tung

Student ID: 103181157

Word Count: 2010

Table of Content

Abstract

Data is a crucial asset for both businesses and daily life, with unstructured data now making up over 80% of enterprise data. Unlike structured data, unstructured data lacks a set format, making it difficult to process and analyse using conventional methods. However, technologies such as blockchain, artificial intelligence, and chatbots are transforming how organisations use unstructured data, which enhance decision-making and improve operational efficiency. This literature review explores the role of these emerging technologies in managing and utilising unstructured data. Blockchain ensures data integrity and security, AI enhances complex data analysis through machine learning (ML), deep learning (DL), natural language processing (NLP), and Chatbots automate user interactions using natural language processing (NLP). Despite these ample benefits, challenges such as data quality, scalability, and regulatory hurdles remain.

Key Word: Unstructured data, Blockchain, Artificial Intelligence, Chatbots, Natural Language Processing (NLP), Machine learning (ML), Deep Learning (DL)

Introduction

In these modern days and ages, data has been becoming the essential part of not only companies and enterprises but also for daily life usage. The rapid growth of digital information has resulted in unstructured data comprising over 80% of all enterprise data generated today (Kumar et al., 2023). Unlike structured data, which refers to data that is organised in a predefined organisation, often in tabular formats like relational databases or spreadsheets according to Janos Demetrovics, Son and Akos Guban (2016), unstructured data can not be easily organised into traditional row-column databases and it can be in various formats, such as text documents, images, videos, social media posts, and audio files (Wang et al., 2016). Due to the fact that unstructured data lacks a predefined structure, it's much more complex to process and analyse it with traditional data management methods. However, it holds significant value, often containing valuable insights that can enhance decision-making, improve customer experiences, and guide strategic planning. (Kumar et al., 2023).

In these recent years, technologies like blockchain, AI, and chatbots have also become increasingly popular in many industries. When combined with unstructured data, they have made a noticeable impact in several areas.

This literature review looks at how blockchain, AI, and chatbots are being used to manage unstructured data, focusing on their practical uses and challenges. By reviewing current research, this paper gives a clear overview of the latest trends and shows how these technologies can help unlock the full potential of unstructured data.

Literature Review: Utilisation of Unstructured Data Across Blockchain, AI, and Chatbots

Unstructured data shows a lot of potential when combined with technologies such as blockchain, artificial intelligence, and chatbots. The advantages of these combinations will be demonstrated in the following sections

Blockchain and Unstructured Data

Blockchain is a decentralised transaction and data management technology developed first for the Bitcoin cryptocurrency, providing security and data integrity without third party control (Yli-Huumo et al., 2016). It is used to manage unstructured data like videos and images, with Ethereum and the internet of vehicles being the largest underlying blockchains and fields of application, respectively (Ceron, Cristian Tinipuclla and Shiguihara, 2023). Blockchain's decentralised architecture allows data to be distributed across multiple nodes, creating a secure record of all the data entries. This feature is especially valuable in scenarios where data reliability is critical, such as in healthcare or in supply chain management.

In healthcare, blockchain is used to securely manage unstructured medical records, such as clinical notes and diagnostic images, ensuring sensitive patient data stays protected while allowing authorised healthcare providers to access it when needed (Dagher et al., 2018). This technology greatly improves interoperability by enabling providers to securely share and access patient data across different platforms.

In the supply chain industry, blockchain is employed to track unstructured data related to product histories, such as shipping documents, inspection reports, and product images (Liu, Ran and Liu, 2023). Through this integration, a highly reliable and secure system can be established, effectively addressing numerous challenges related to visibility and transparency (Azzi, Chamoun and Sokhn, 2019).

However, on the other hand, despite its potential, using blockchain for collecting and storing unstructured data presents several challenges. In the energy sector, for instance, blockchain adoption faces obstacles like technical limitations, market barriers, and regulatory issues. Andoni et al. (2019) point out that these challenges must be addressed for the technology to be widely adopted.

AI and Unstructured Data

Before exploring this topic further, it's important to clarify a key concept - the definition of AI, which is frequently debated and understood in different ways. According to Kaplan and Haenlein (2019), AI is a system's ability to interpret external data, learn from it, and achieve specific goals and tasks through flexible adaptation. With its powerful capabilities in machine learning (ML), deep learning (DL), natural language processing (NLP), and computer vision, AI has become one of the most effective ways to transform unstructured data into valuable knowledge that drives decision-making and innovative solutions. It offers a variety of techniques that can process and analyse those data in ways that traditional methods cannot, which comprises the use of ML and DL models to recognize patterns, perform object detection and classification and even create prediction.

For instance, in healthcare and biomedical application, Convolutional Neural Networks (CNNs) - which are a class of deep learning algorithms specifically designed for processing images and videos (Miao, Yao and Zhao, 2021), have achieved exceptional performance levels in many applications such as personalised biomedical data classification, early diagnosis, structural health monitoring, anomaly detection, utilising medical images from the patients (Kiranyaz, 2021), which significantly enhances the efficiency of treatments and related medical procedures.
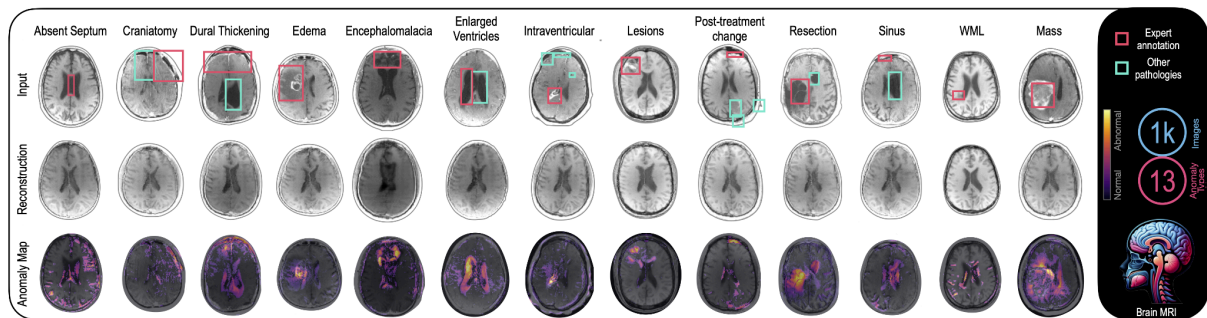

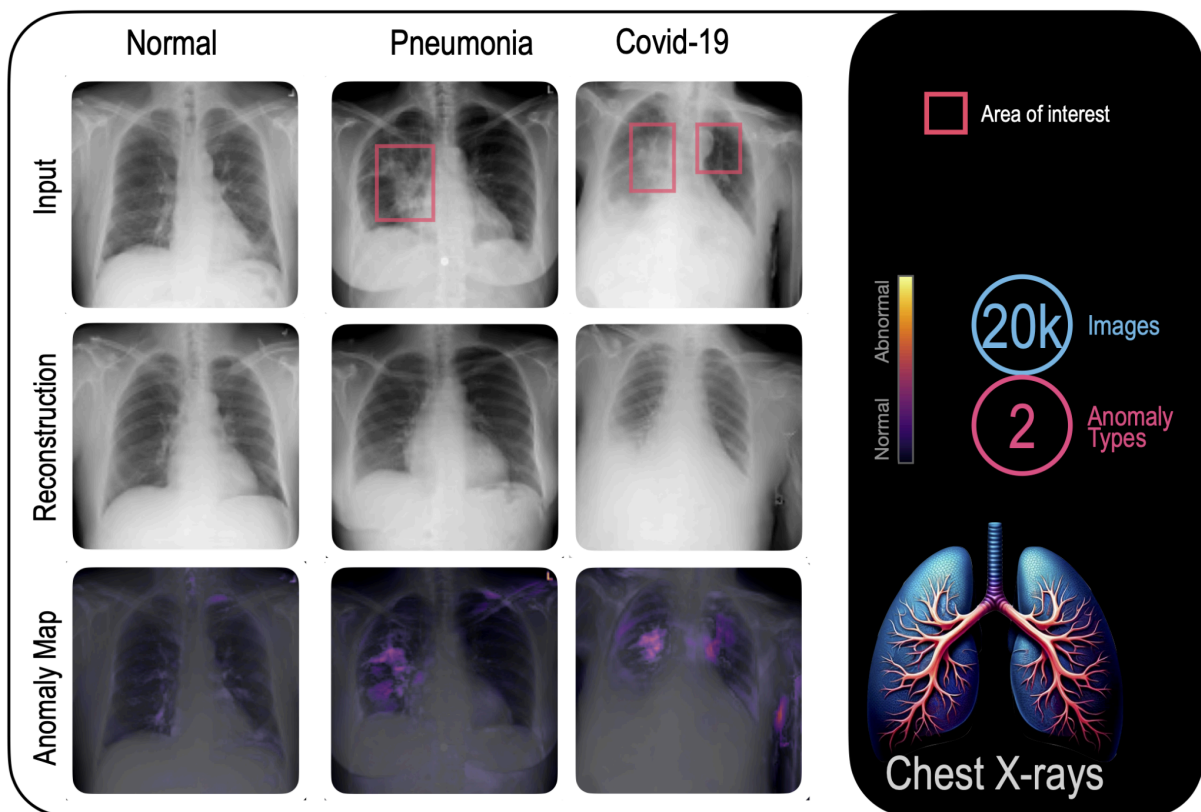Figure 1: Anomaly detection in brain MRI using DL (Arxiv.org, 2024)

Figure 2: Anomaly detection on chest X-rays using DL (Arxiv.org, 2024)

Additionally, AI-powered licence plate recognition systems use advanced computer vision, especially CNNs, to detect and read vehicle licence plates from live traffic camera footage. These systems can accurately identify plate numbers in real time and extract important

details such as the owner's information, registration status, outstanding fines, and any related criminal records by cross-referencing databases. This technology is crucial for law enforcement, enabling police to quickly track vehicles involved in crimes and improve response times, regardless of weather or lighting conditions. (Shashirangana et al., 2021)
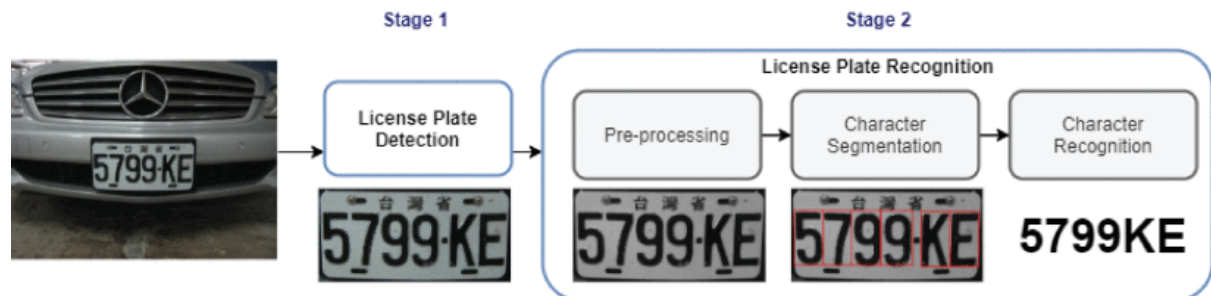

Figure 3 : Key phases in a licence plate recognition system (Shashirangana et al., 2021)

On the other hand, although AI has great potential, it still faces several challenges when dealing with unstructured data, with data quality being one of the main issues. This can be explained by the fact that AI systems often have difficulty managing the different formats and layouts that unstructured documents come in. To overcome this, a solid framework is needed to process and improve the quality of the data, but creating such a system can be costly and time-consuming (Baviskar et al., 2021).

Chatbots and Unstructured Data

Chatbots are devices that use artificial intelligence and machine learning to communicate with users and provide automated responses to their requests in natural language (Suhaili, Salim and Jambli, 2021). They have become highly effective in managing unstructured data, especially text and speech. Using NLP and neural networks, chatbots analyse human language, extract key information from unstructured data, and generate relevant and appropriate responses, as noted by Han (2023).

For example, many large corporations, such as Amazon and Google, utilise chatbots to handle customer service and product questions. When a customer submits a text inquiry, the chatbot analyses the unstructured data, stores it in a database, identifies key phrases, and retrieves the relevant answers from its knowledge base. (Kavitha & Murthy, 2019).

On the other hand, chatbots also face several challenges when processing unstructured data. Human language is complex and often includes slangs or cultural references that can be difficult for chatbots to understand, which results in irrelevant and inappropriate responses generated (Cai et al., 2023).

Conclusion

In conclusion, technologies such as blockchain, AI, and chatbots are transforming how unstructured data is handled across businesses. Blockchain ensures secure and transparent data administration, AI enables enhanced data analysis, and chatbots enhance real-time

interactions with unstructured data. However, there are still some difficulties such as data quality and scalability, and technical limitations. As a result, continuous innovation is necessary for fully exploiting the potential of these technologies in dealing with the difficulties of unstructured data.

Reference

Andoni, M., Robu, V., Flynn, D., Abram, S., Geach, D., Jenkins, D., McCallum, P., & Peacock, A. (2019). Blockchain technology in the energy sector: A systematic review of challenges and opportunities. *Renewable and Sustainable Energy Reviews.* https://doi.org/10.1016/J.RSER.2018.10.014.

Arxiv.org. (2024). Towards Universal Unsupervised Anomaly Detection in Medical Imaging. https://arxiv.org/html/2401.10637v1/#S5.

Azzi, R., Kilany, R., & Sokhn, M. (2019). The power of a blockchain-based supply chain. *Comput. Ind. Eng.*, 135, 582-592. https://doi.org/10.1016/J.CIE.2019.06.042.

Baviskar, D., Ahirrao, S., Potdar, V., & Kotecha, K. (2021). Efficient Automated Processing of the Unstructured Documents Using Artificial Intelligence: A Systematic Literature Review and Future Directions. *IEEE Access*, 9, 72894-72936. https://doi.org/10.1109/ACCESS.2021.3072900.

Cai, Z., Haslett, D., Duan, X., Wang, S., & Pickering, M. (2023). Does ChatGPT resemble humans in language use? *ArXiv*, abs/2303.08014. https://doi.org/10.48550/arXiv.2303.08014.

Ceron, J., Cristian Tinipuclla, & Shiguihara, P. (2023). A Survey of Blockchain for Video Integrity. https://doi.org/10.3390/engproc2023042004.

Dagher, G.G., Mohler, J., Milojkovic, M., & Marella, P.B. (2018). Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. *Sustainable Cities and Society*, 39, 283–297. https://doi.org/10.1016/j.scs.2018.02.014.

Demetrovics, J., Hua, N., & Gubán, Á. (2016). A Formal Representation for Structured Data. https://doi.org/10.12700/APH.13.2.2016.2.4.

Han, Z. (2023). The applications of chatbot. *Highlights in Science, Engineering and Technology.* https://doi.org/10.54097/hset.v57i.10011.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons.* https://doi.org/10.1016/J.BUSHOR.2018.08.004.

Kavitha, B., & Murthy, C. (2019). Chatbot for Healthcare System Using Artificial Intelligence. *International Research Journal of Modernization in Engineering Technology and Science.* https://doi.org/10.56726/irjmets34164.

Kiranyaz, S., Avcı, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D. (2019). 1D Convolutional Neural Networks and Applications: A Survey. *ArXiv*, abs/1905.03554. https://doi.org/10.1016/j.ymssp.2020.107398.

Kumar, P., Tveritnev, A., Jan, S.A., & Iqbal, R. (2023). Challenges to Opportunity: Getting Value Out of Unstructured Data Management. [online] *OnePetro.* https://doi.org/10.2118/214251-MS.

Liu, R., Ran, W., & Liu, S. (2023). Blockchain Technology Applied to Supply Chain Management: A Systems' Analysis. *Mobile Information Systems*, 2023, 1–23. https://doi.org/10.1155/2023/6046503.

Miao, F., Yao, L., & Zhao, X. (2021). Evolving convolutional neural networks by symbiotic organisms search algorithm for image classification. *Appl. Soft Comput.*, 109, 107537. https://doi.org/10.1016/J.ASOC.2021.107537.

Shashirangana, J., Padmasiri, H., Meedeniya, D., & Perera, C. (2021). Automated License Plate Recognition: A Survey on Methods and Techniques. *IEEE Access*, 9, 11203-11225. https://doi.org/10.1109/ACCESS.2020.3047929.

Suhaili, S., Salim, N., & Jambli, M. (2021). Service chatbots: A systematic review. *Expert Syst. Appl.*, 184, 115461. https://doi.org/10.1016/J.ESWA.2021.115461.

Wang, J., Shen, X., Sun, Y., & Qu, A. (2016). Classification With Unstructured Predictors and an Application to Sentiment Analysis. *Journal of the American Statistical Association*, 111(515), 1242–1253. https://doi.org/10.1080/01621459.2015.1089771.

Yli-Huumo, J., Ko, D., Choi, S., Park, S., & Smolander, K. (2016). Where Is Current Research on Blockchain Technology?—A Systematic Review. *PLOS ONE*, 11(10), e0163477. https://doi.org/10.1371/journal.pone.0163477.