International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-14 Issue-9, August 2025

Integration of Big Data and IoT: A Review

Amit Kumar, Kanchan Singh, Aman Jyoti

Abstract: ***In today's world, data is everywhere — whether it's coming from devices, sensors, or digital systems used across different industries. As this data continues to grow, it has become increasingly important not only to collect it but also to make sense of it in real-time. That's where the combination of Big Data and the Internet of Things (IoT) comes in. While IoT helps collect information from the physical world using connected devices, Big Data tools help process that information to identify functional patterns, trends, or warnings. This paper examines how these two technologies work together and how they're being applied in real-world situations, such as hospitals, smart cities, farms, and factories. It also discusses the benefits, such as faster decision-making and more efficient systems. Still, it doesn't overlook the challenges, including data security, handling large volumes, and ensuring that different systems work smoothly together. In short, this review examines how Big Data and IoT support each other, where they're already making a difference, and what needs to be improved going forward. The goal isn't just to explain the technology, but to give a real picture of how it's shaping the way we work and live today — and what that could mean for the future.***

Keywords: Big Data, IoT, Cloud Computing

Abbreviations:
AI: Artificial Intelligence
ML: Machine Learning
IoT: Internet of Things

I. INTRODUCTION

These days, digital technology is growing rapidly, and with it, the amount of data being generated has increased exponentially. Every time we use an app, browse social media, or work with machines in industries, data is being created in the background. Even small, everyday devices — such as thermostats or fitness bands — are constantly collecting some kind of data. And when you think about how many places this data is coming from, it's pretty evident that we need more intelligent systems to manage and make sense of it.

That is where Big Data and IoT come into the picture. Big Data primarily deals with handling large-scale information that exceeds the capabilities of regular tools [1]. On the other hand, IoT involves devices generating real-time data continuously [4].

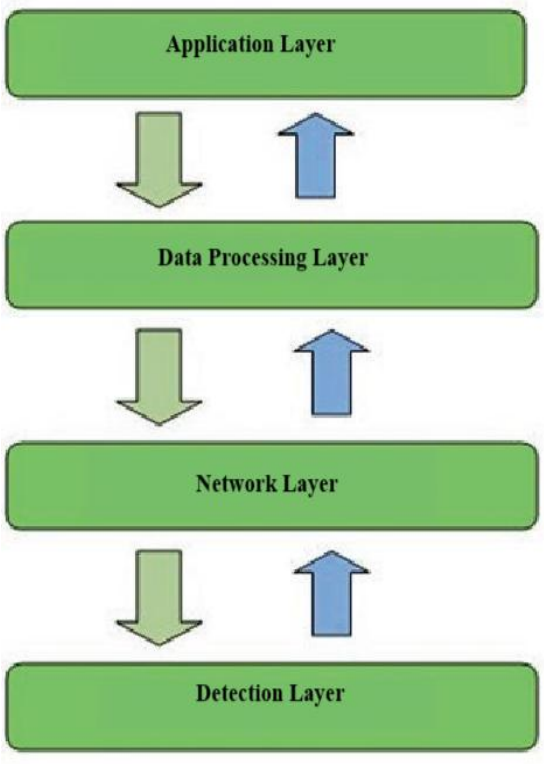
Manuscript received on 02 June 2025 | First Revised Manuscript received on 17 June 2025 | Second Revised Manuscript received on 18 July 2025 | Manuscript Accepted on 15 August 2025 | Manuscript published on 30 August 2025.
*Correspondence Author(s)
Amit Kumar*, Department of Electronics Communication Engineering, Lingaya's Vidyapeeth, Faridabad (Haryana), India. Email ID: ak6144000@gmail.com, ORCID ID: 0009-0005-2844-9090
Kanchan Singh, Department of Electronics Communication Engineering, Lingaya's Vidyapeeth, Faridabad (Haryana), India. Email ID: kanchansingh1120@gmail.com
Aman Jyoti, Assistant Professor, Department of Electronics Communication Engineering, Lingaya's Vidyapeeth, Faridabad (Haryana), India. Email ID: amanjyoti@lingayasvidyapeeth.edu.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

When used together, they complement each other by enabling real-time data collection and analysis. For example, in smart cities, this integration can help optimize traffic and utilities [3]. When both are used together, they complement each other. IoT focuses on collecting real-time data, while Big Data takes over to analyse it and help determine what's happening. For example, if a city installs IoT sensors to monitor traffic, Big Data can step in to examine the patterns and help determine how to reduce congestion or improve routes. This combination is already being utilised in various areas, including hospitals, transportation systems, and agriculture, and it's making everyday operations faster, more accurate, and more efficient [2].

This paper examines how these two technologies work in tandem, the benefits of this combination, and where it's being applied in real-life scenarios.

It also examines some of the roadblocks that still exist and explores what future developments might help alleviate them. The goal is to give a clearer picture of how this integration can lead to more efficient, automated, and responsive systems [3].



[Fig.1: IoT Architecture] [3]

II. UNDERSTANDING BIG DATA AND IOT

Big Data is characterized by the 5Vs: Volume, Velocity, Variety, Veracity, and Value [3]. IoT refers to a network of connected devices—smart appliances, sensors, and wearables—that transmit data continuously [4]. Integrating

© Copyright: All rights reserved.

big Data and IoT enables real-time analytics and automation, like predictive maintenance in industrial systems [6].

To understand how Big Data and IoT work together, it is helpful first to examine the meanings of each term. Big Data is typically described by five main features — often referred to as the 5Vs. These are: **Volume**, which is about the massive size of data being created; **Velocity**, which refers to how fast that data is coming in and needs to be processed; **Variety**, which covers the different forms it takes (like text, images, videos, etc.); **Veracity**, which means how reliable or clean that data is; and **Value**, which focuses on whether the data is valid once it's processed.

On the other hand, IoT — or the Internet of Things — is more about the devices that constantly collect this data. These could be anything from smartwatches and home appliances to factory machines and traffic sensors. What makes IoT so powerful is that these devices are all connected and can send data without any manual input. And they are doing this 24/7, creating a live stream of information.

Now, when you put Big Data and IoT together, things get even more interesting. The devices collect the data, and Big Data tools store and analyse it. This is not just about keeping records — it is about turning that information into something meaningful. For example, if a sensor in a factory machine detects an anomaly, the system can predict a potential problem before it occurs and even trigger necessary maintenance. This back-and-forth loop makes the system smarter over time, helping it learns, improve, and adjust as needed [7].

III. INTEGRATION OF BIG DATA AND IOT

Edge computing processes data close to where it is generated, thereby reducing latency, which is particularly beneficial in time-critical scenarios [8]. Cloud platforms, such as AWS and Microsoft Azure, enable scaling of storage and processing capacity [9]. Tools like Hadoop and Spark manage large, complex datasets from distributed IoT devices.

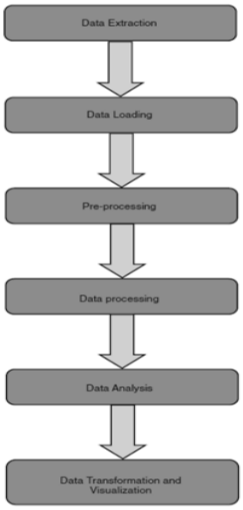
When Big Data and IoT come together, it is basically about using the best parts of both — IoT devices keep sending real-time data non-stop, and Big Data tools help make sense of it. This combo enables businesses and organisations not only to collect data but also to interpret it and act quickly.

This connection typically involves a step-by-step setup. First, you have IoT devices that are deployed in the field, such as sensors in machines, smart meters, or wearables. These devices gather data on the spot. Then, before sending everything to a central system, some basic processing occurs closer to the devices themselves — this is where edge computing or fog nodes come into play. Following this, a more detailed analysis is conducted on cloud platforms, where Big Data tools can delve deeply into trends, patterns, or alerts.

Let us say a factory is using this setup. Sensors keep checking the health of machines. If a device detects something unusual — such as a machine not functioning correctly — that information does not always have to be sent to the cloud immediately. Instead, it is often checked right where it's happening, using something called edge computing. Instead of sending every bit of data to the cloud, only the critical information — such as warnings or summaries — is sent ahead. This way, the system does not get overloaded, and everything runs quicker. It comes down to being able to respond fast, especially when something goes wrong [9].

Behind the scenes, companies use tools like **Hadoop** or **Spark** to handle and process the vast amount of data. Since much of the information is not neatly organised, databases like NoSQL are used to store it effectively. To keep things running smoothly, cloud platforms like AWS or Microsoft Azure are often utilised — they facilitate easy scaling when the data volume increases.

All of this works together to help businesses stay on top of things as they happen. It enables them to identify issues early, resolve problems before they become significant, and adjust on the fly as needed. It is not just about collecting tons of data — it is about knowing what to do with it, and doing it at the right time.



[Fig.2: Processing of Data in an IoT Environment]

IV. APPLICATIONS ACROSS INDUSTRIES

In healthcare, wearable sensors help track vital signs and send data to cloud systems, where analytics assist in early diagnosis and the development of tailored treatments [5]. Cities utilise IoT sensors for enhanced lighting, waste management, and pollution tracking—Big Data facilitates informed decision-making. In agriculture, sensor data enables precision farming strategies, such as optimised irrigation.

When Big Data and IoT come together, they have started to change how industries work on the ground. Take healthcare, for example. You will see smartwatches and health monitors almost everywhere now. These gadgets track things like heart rate, sleep quality, or oxygen levels — and they quietly send all that info to cloud systems in the background. Once the data is in, Big Data tools analyse it and look for signs of anything unusual, such as a sudden dip in health, and can even help doctors create more personalised treatment plans.

In cities, this same tech combo is helping with daily tasks, such as managing traffic lights, buses, and even checking air quality. Sensors spread across different systems collect data, which is used instantly to improve traffic flow or conserve power where possible. Bit by bit, Cities are becoming more responsive and liveable.

© Copyright: All rights reserved.

Farming has also undergone significant changes because of this. Farmers are now using sensors in the fields to receive real-time information about soil conditions, weather updates, and the progress of their crops. That way, instead of relying on intuition or past practices, they can determine the optimal time to water, fertilise, or harvest — all based on actual data. It saves time, reduces waste, and typically yields better results.

Factories are following the same idea. Machines now have built-in sensors that track performance. If something starts going wrong, Big Data systems can detect the signs early and send alerts before any issues arise. That helps reduce breakdowns and avoid delays. Even supply chains are smarter now, since tracking data helps manage stock and deliveries more smoothly.

So, whether it's hospitals, farms, factories, or big cities, the combination of IoT and Big Data enables people to work faster, make fewer mistakes, and use their time more effectively.

V. CHALLENGES AND CONSIDERATIONS

Although combining Big Data and IoT can yield numerous benefits, it's not without its challenges. One of the first issues that arises when discussing IoT and Big Data together is the concern over **security and privacy**. Many of these connected devices — especially the less expensive or basic ones — lack robust security features. They are just made to do the job and send data, and that's it. However, that also means they can be easily hacked into if someone wants to. Without proper encryption or strong passwords, the personal or sensitive information these devices collect could easily fall into the wrong hands.

Then there's the problem of **scalability**. As more devices are added to a system, the amount of data generated grows exponentially. It's not just a slight increase — we're talking about tons of data, every single second. If the company's systems aren't built to handle that kind of traffic, things can get slow, or worse, crash altogether. That's why using cloud storage that can scale with demand and having systems that can expand as needed is becoming a necessity.

There's also a **standardization** problem. Right now, different devices speak different "languages" — they use various protocols and formats. As a result, it's often challenging to integrate all these systems seamlessly. Unless there are clear, universal standards, managing and integrating everything becomes a mess.

Lastly, there's the issue of **data quality**. Just because you have a lot of data doesn't mean it's all useful. Some devices may send incomplete information, duplicate readings, or noisy data that requires cleaning. If the data isn't reliable or well-organised, then any insights derived from it could be flawed. Before running analytics, a strong process is needed to verify and clean the data first.

VI. FUTURE DIRECTIONS

Looking ahead, the combination of Big Data and IoT is poised to become even more powerful, especially with the introduction of new technologies. One big game changer is **Artificial Intelligence (AI)** and **Machine Learning (ML)**. These tools can help make sense of complex patterns in real-time IoT data, which means systems can not only respond faster but also learn from past events. For example, AI can help identify a piece of equipment that is likely to fail by analysing subtle changes in its performance over time.

Edge computing is also expected to become more common. Instead of sending all the data to the cloud first, some processing occurs right near the source, such as on the device or local gateway. This reduces delays and alleviates pressure on networks. It is beneficial in situations where every second counts, like emergency healthcare or automated vehicles.

Another area that requires serious attention is policy and regulation. As more personal and sensitive data gets collected by IoT systems, questions about who owns that data and how it can be used are becoming more critical. Governments and tech companies will need to collaborate to establish clear guidelines that protect users without hindering innovation.

There is also a lot of room for research, especially around lightweight security methods for IoT devices, better ways to clean real-time data, and making sure all kinds of devices can "talk" to each other smoothly. If these areas are improved, the integration of Big Data and IoT will only get more innovative, safer, and more useful in the future.

VII. CONCLUSION

Bringing together Big Data and IoT has changed the way industries collect and use information. Throughout this paper, we've seen how these two technologies complement each other — IoT collects data nonstop, and Big Data helps process and analyze that data to make smarter decisions.

Whether it's in healthcare, agriculture, manufacturing, or city planning, this integration is making things faster, more efficient, and more responsive. But it's not without challenges. We discussed real concerns, including security, scalability, inconsistent data, and the lack of standardised systems. These issues require attention if organisations are to achieve long-term success.

Looking forward, the use of AI, edge computing, and better privacy rules will likely shape how this integration grows. If done correctly, it can lead to systems that are not only smart but also secure and reliable. What's clear is that the combination of Big Data and IoT isn't just a tech trend — it's becoming a critical part of how businesses and society operate today.

So, while there's still work to be done, the path ahead looks promising. With the right tools and direction, this powerful duo can continue to improve the way we live and work.

DECLARATION STATEMENT

After aggregating input from all authors, I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/Competing Interests:** Based on my understanding, this article does not have any conflicts of interest.
- **Funding Support:** This article has not been sponsored or funded by any organization or agency. The independence of this research
  It is a crucial factor in affirming its

© Copyright: All rights reserved.

impartiality, as it was conducted without any external influence.

- **Ethical Approval and Consent to Participate:** The data provided in this article is exempt from the requirement for ethical approval or participant consent.
- **Data Access Statement and Material Availability:** The data and materials supporting the findings of this article are publicly accessible.
- **Author's Contributions:** The authorship of this article is contributed equally to all participating individuals.

REFERENCES

1. Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M. (2015). Internet of Things: A survey on enabling technologies, protocols, and applications. *IEEE Communications Surveys & Tutorials, 17*(4), 2347–2376. DOI: https://doi.org/10.1109/COMST.2015.2444095
2. Yadav, D. K., & Jadhav, S. (2019). Leveraging big data analytics and IoT in smart cities: A review. *Computer Science Review, 32*, 1–20. DOI: https://doi.org/10.1016/j.cosrev.2019.100307
3. Wang, G., Gunasekaran, A., Ngai, E. W. T., & Papadopoulos, T. (2016). Big Data Analytics in Logistics and Supply Chain Management: Specific Investigations for Research and Applications. *International Journal of Production Economics, 176*, 98–110. DOI: https://doi.org/10.1016/j.ijpe.2016.03.014
4. Bibri, S. E. (2019). The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability. *Sustainable Cities and Society, 38*, 230–253. DOI: https://doi.org/10.1016/j.scs.2017.12.034
5. Zhang, T., Sun, Y., Wang, Y., & Li, W. (2021). A comprehensive survey on federated learning for the Internet of Things. *arXiv preprint arXiv:2111.07494*. DOI: https://doi.org/10.48550/arXiv.2111.07494
6. Bharati, S., & Podder, P. (2022). An efficient and lightweight secure architecture for IoT using deep learning-based federated identity. *arXiv preprint arXiv:2210.13547*. DOI: https://doi.org/10.48550/arXiv.2210.13547
7. Zhang, T., Li, Y., Zhang, H., Zhou, X., & Yu, F. R. (2024). Integration of artificial intelligence with edge computing and Internet of Things: Architecture, applications, and future directions. *Artificial Intelligence Review, 57*(1), 1345–1372. DOI: https://doi.org/10.1007/s10462-024-10938-5
8. MDPI Editors. (2024). 5G and IoT integration for intelligent environments. *Computers, 14*(7), 250. DOI: https://doi.org/10.3390/computers14070250
9. Dave, V., & Mittapally, N. (2024). Data integration and interoperability in IoT: Challenges, strategies, and future direction. *ResearchGate*. https://www.researchgate.net/publication/377805078_Data_Integration_and_Interoperability_in_IOT_Challenges_Strategies_and_Future_Direction

AUTHOR'S PROFILE



Mr. Amit Kumar, I am currently pursuing a B. Tech in Electronics and Communication Engineering from Lingaya's Vidyapeeth, Faridabad, India. I am also a working professional, serving as an Assistant Section Engineer at Delhi Metro Rail Corporation. With a keen interest in the efficient handling of large-scale data and deriving meaningful insights, my research focuses on data analytics and intelligent systems. I bring practical experience from the transportation sector into my academic exploration, aiming to bridge the gap between real-world systems and emerging technologies.



Mr. Kanchan Singh completed his Diploma in Electronics and Communication Engineering from Pusa Polytechnic, New Delhi, and his B. Tech in Electronics and Communication Engineering from Lingaya's Vidyapeeth, Faridabad. His key areas of interest include Artificial Intelligence (AI), the Internet of Things (IoT), electronics, and computer systems. He is passionate about integrating AI and IoT to develop innovative, intelligent, and automated solutions for modern-day applications. With a commitment to continuous learning and practical research, he is dedicated to advancing intelligent technologies through research, development, and collaborative projects, aiming to contribute to the field by leveraging his knowledge to solve real-world problems through innovative and efficient electronic systems.



Ms. Amanjyoti Sethi, a Ph. D. Research Scholar at Symbiosis International University, Pune, is currently an Assistant Professor in the Department of Electronics & Communication Engineering at Lingaya's Vidyapeeth. She has received her bachelor's from Kurukshetra University and her master's from Thapar University. Her research focuses on the application of Artificial Intelligence in Electric Vehicles, with a particular emphasis on innovative battery management, energy optimization, and IoT-based intelligent mobility solutions. She is working on developing AI-driven models to enhance the performance, reliability, and sustainability of electric vehicle systems. Her broader research interests include power electronics for e-mobility, predictive maintenance, connected vehicle technologies, and the integration of renewable energy into smart charging infrastructures. She has published several papers in reputed journals and actively participates in conferences related to intelligent transport systems and green energy technologies.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)/ journal and/or the editor(s). The Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP) and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

© Copyright: All rights reserved.