Data Analytics in the Internet of Things Era: Tools, Approaches, Challenges, and Solutions

Yun Liu

School of Mathematics and Computer Science, Quanzhou Normal University, Quanzhou, 362000, China
Fujian Provincial Key Laboratory of Data-Intensive Computing, Quanzhou Normal University, Quanzhou, 362000, China
E-mail: bbd_yy@163.com

Received 25 March 2025; Accepted 03 December 2025

Abstract

Rapid growth in the number of devices connected to the Internet of Things (IoT) and the exponential surge in data usage clearly suggest that the development of big data is inextricably linked with the IoT. In an ever-expanding network, big data raises concerns regarding data access efficiency. This study critically reviews IoT data analytics, tools, techniques, and challenges in extracting meaningful information from IoT device-generated massive data sets. IoT data analysis approaches, including real-time analysis, predictive analysis, and anomalous behavior analysis, are discussed in detail. How big data platforms and cloud computing can tackle IoT data and why IoT data preprocessing, integration, and storage matter are explored in this paper. Additionally, it covers issues and future research directions in IoT data analytics, including data security, scalability, and privacy.

Keywords: Internet of Things, big data, data analytics, review.

Journal of ICT Standardization, Vol. 13_4, 427–448.
doi: 10.13052/jicts2245-800X.1344

1 Introduction

The Internet of Things (IoT) has revolutionized device-to-device communications and interactions, as well as led to a colossal explosion in data creation [1]. With trillions of smart devices and sensors creating real-time information, IoT velocity, volume, and variety have increased significantly [2]. With such a boom in data, processing, storing, and analysis have become a challenge in its aftermath [3]. Traditional processing approaches cannot tackle IoT data produced in a voluminous and changing manner; hence, complex analysis of data is necessary to discover meaningful insights [4].

IoT data analysis enables companies and researchers to make informed decisions through real-time information streaming, predictive modeling, and anomalous behavior analysis [5]. Analytic approaches in IoT can be utilized in many industries, including smart cities, medicine, industrial automation, environment, and smart transportation [6]. For example, in intelligent health care, IoT analysis of patient health data identifies anomalies or abnormalities. In the industry, predictive maintenance models prevent failures via sensors [7].

Despite its potential, IoT big data analysis is challenged by various impediments, including diversity in data, scalability, security, and privacy [8]. High-velocity streaming data requires efficient real-time processing architectures, and IoT device distribution requires efficient storage and integration techniques [9]. In addition, the security and maintenance of data integrity is critical, with IoT networks being at a high risk of cybersecurity and unauthorized access [10].

This paper presents a thorough analysis of data analytics in IoT, covering the tools and challenges involved in the study of large data sets generated. The paper further highlights different platforms, machine learning algorithms, and real-time analysis platforms that use data analytics in IoT effectively. Moreover, the paper summarizes the primary challenges encountered in data analytics in IoT.

The main body of the research is organized in the following way: Section 2 presents insights into data analytics in IoT, covering various forms of data analytics. Section 3 describes the different tools and platforms used in IoT data processing. Section 4 gives an overview of the most critical data analysis methodologies in IoT. Section 5 discusses key challenges in IoT data analysis. Section 6 proposes possible research areas for the future. Section 7 concludes the research paper.

2 IoT Data Analytics: An Overview

The IoT has experienced exponential development in terms of data generated by sensors, intelligent devices, and IoT-related systems [11]. The data analytics process in the IoT domain has emerged as the most relevant process in converting data into meaningful information and achieving intelligent decision-making in different domains like healthcare [12], smart cities [13], industrial automation [14], and environmental monitoring [15]. The classification of IoT data analytics is depicted in Figure 1.

IoT data analysis comprises data gathering, processing, storage, and analysis aimed at gaining valuable information from the data [16]. Figure 1 depicts the major types of IoT analysis: descriptive analysis, real-time analysis, predictive analysis, and prescriptive analysis. Various analysis processing

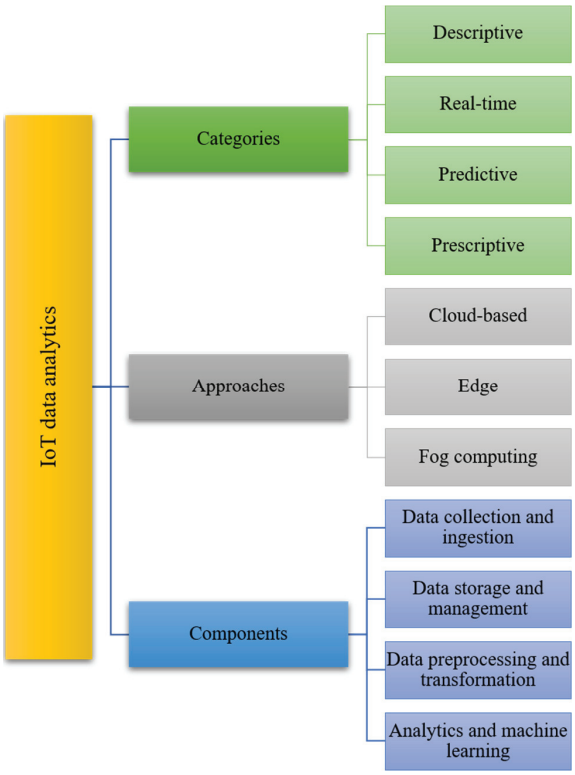

Figure 1 Overview of IoT data analytics categories, approaches, and components.

techniques like cloud computing analysis, edge analysis, and fog computing analysis are effective at analyzing IoT data based on latency requirements and analysis complexity [17]. Moreover, IoT data analysis involves crucial factors such as data gathering/ingestion functionality, data storage/management functionality, data preprocessing/transformation functionality, and machine learning functionality in the analysis. These factors are necessary in IoT data analysis to ensure effective analysis and utilization of IoT data.

2.1 Categories of IoT Data Analytics

IoT data analytics can be classified into four broad categories based on the type of analysis or insight obtained from data analysis. There are four types of analysis: descriptive, predictive, real-time, and prescriptive [18]. Descriptive analytics involves the analysis of data from the past, addressing questions like "What happened?" through statistical analysis and simple aggregation [19]. Descriptive analytics in IoT entails analyzing data from sensors in many IoT applications.

Real-time analytics facilitates actions and insights while the data is still being generated [20]. This kind of analytics is most critical in time-critical use cases like self-driving cars, automated industries, and health tracking. Real-time analytics uses stream processing engines like Apache Kafka, Apache Flink, and Apache Storm for processing high-speed data streams [21].

Predictive analytics relies on machine learning and modeling methods for forecasting future events. The question being answered here is "What's likely to happen?" through the identification of correlations found within IoT data [22]. Examples of predictive analytics applications include predictive maintenance in Industrial IoT or forecasting based on environmental IoT sensors. Prescriptive analytics not only makes predictions but also provides optimized recommendations.

Prescriptive analytics responds to the query: "What should be done?". Prescriptive analytics techniques include optimization methods, reinforcement learning, and decision support systems. Prescriptive analytics applications in IoT include automating traffic in smart cities, resource allocation in manufacturing processes, and consumption in smart grids.

2.2 IoT Data Processing Approaches

Efficient processing of IoT data is imperative considering the velocity, variety, and volume of data streamed [23]. IoT data is processed using cloud analytics, edge computing, and fog computing.

Cloud computing provides scalability and flexibility in resource utilization for processing and analysis of IoT data. Cloud platforms like Amazon Web Services (AWS), Microsoft Azure, and Google Cloud deliver powerful data analysis solutions for companies. This permits companies to process vast volumes of data effectively [24]. Cloud analysis offers effective solutions to applications requiring intensive computation, like deep analysis in IoT and predictive modeling. However, there are constraints to cloud analysis, like latency and bandwidth, in applications requiring real-time decision support.

Edge analytics refers to the process of performing computational analysis on data close to the source or on IoT devices and gateways, hence requiring less data transfer back to distant servers in the cloud [25]. This process reduces latency levels and boosts real-time decision-making effectiveness. Edge computing applications include industry automation, self-driving cars, and smart surveillance systems. Edge analytics provides lower bandwidth requirements and enhanced privacy levels. However, there are constraints in terms of processing power and memory for this type of analysis.

A fog computing environment performs calculations and data storage closer to the IoT devices, but maintains connectivity with the central cloud systems [26]. This process further boosts the level of scalability and efficiency in terms of data processing performed across various hierarchical levels. Fog computing can be best utilized in smart cities and healthcare IoT applications.

2.3 Fundamentals of IoT Data Analytics

IoT data analysis involves various stages from data gathering to analysis of the needed information [27]. IoT data analysis modules are data gathering or acquisition, data storage and management, data preprocessing or transformation, analysis and machine learning, and decision interpretation.

IoT devices generate continuous data through sensors, actuators, or other connected systems. The process of data gathering through various frameworks like MQTT (Message Queuing Telemetry Transport), CoAP (Constrained Application Protocol), and RESTful APIs in HTTP proves effective in transferring data [28]. Tools like Apache NiFi, Cloud Pub/Sub from Google Cloud Platform, and Kinesis from Amazon Web Services are utilized for data gathering from IoT devices.

IoT data needs scalable and high-speed data storage solutions. Based on the type of data processed in IoT applications, various data storage models are deployed. These include time series databases (InfluxDB and TimescaleDB) for storing data from sensors with timestamps; NoSQL databases (MongoDB

and Cassandra) if the IoT data is unstructured or semi-structured; and Cloud Storage solutions (Amazon S3 or Google Cloud Storage) in cases of data archival [29]. IoT data collected initially contained noisy data, missing values, and many inaccuracies. Data preprocessing involves:

- Cleaning: Removing duplicate, corrupted, or irrelevant data points.
- Normalization: Standardizing data formats for consistency.
- Feature engineering: Extracting relevant features for machine learning models.

IoT data is analyzed using various machine learning and statistical techniques:

- Supervised learning (e.g., regression, decision trees) for predictive analytics.
- Unsupervised learning (e.g., clustering, anomaly detection) for pattern recognition.
- Deep learning (e.g., CNNs, LSTMs) for complex data processing tasks.

More advanced AI methods, such as federated and reinforcement learning, are being incorporated into IoT analytics for smarter decision-making. The final step of IoT data analysis involves providing meaningful results for decision makers. Result visualization tools such as Power BI, Tableau, or Grafana make real-time IoT system monitoring possible. A significant step in IoT data analysis involves data visualization with various tools and methods.

3 Tools for IoT Data Analytics

The level of complexity and scale of data generated by IoT devices make it necessary to have specialized tools or platforms in order to handle or process data effectively [30]. This data or information is necessary in decision-making or prediction models in IoT applications.

Big data platforms are necessary in the processing of voluminous data from IoT sources, in addition to providing distributed computing infrastructure and high-throughput data ingestion functionality [31]. From Table 1, it can be seen that Apache Hadoop and Apache Spark are the big data platforms that provide scalable solutions in the processing of voluminous data. Big data platforms have malleable data analysis and storage functions necessary in the analysis of voluminous data from IoT sources. IoT applications involve the processing of data streams in real time.

The application of stream processing frameworks helps in the analysis of events with low latency. The different stream processing frameworks

Table 1 Big data platforms for IoT analytics

Platform	Description	Key features
Apache Hadoop	Open-source framework for distributed storage and batch processing of large datasets.	HDFS storage, MapReduce processing, and scalability.
Apache Spark	In-memory data processing engine for on-demand and batch analytics.	High-speed processing, machine learning libraries (MLlib), and support for multiple languages (Scala, Python, Java).
Google BigQuery	Cloud-based data repository optimized for large-scale analytics.	Serverless architecture, SQL-based querying, and automatic scaling.
Amazon EMR	Managed Hadoop framework on AWS for processing big data workloads.	Elastic scaling, integration with S3, Spark, and Presto.
Microsoft Azure Synapse Analytics	Cloud-based analytics service that integrates big data and data warehousing.	Advanced security, SQL-based analysis, and big data integration.

Table 2 Stream processing frameworks for IoT analytics

Framework	Description	Key features
Apache Kafka	Distributed event-streaming platform for handling real-time data feeds.	High throughput, fault tolerance, and real-time event processing.
Apache Flink	Stream-processing framework for real-time and batch analytics.	Stateful stream processing, low latency, and machine learning integration.
Apache Storm	A real-time computation system is designed for distributed stream processing.	Low latency, scalability, and fault tolerance.
Google Cloud Dataflow	Managed stream and batch processing service based on Apache Beam.	Fully managed, auto-scaling, and real-time analytics.
AWS Kinesis	Scalable real-time data streaming service for IoT analytics.	High availability, real-time insights, and serverless integration.

that support the analysis of data streams in real-time in IoT are shown in Table 2, including Apache Kafka and Amazon Kinesis. Stream processing frameworks support real-time analysis of events in IoT applications in terms of event detection.

Table 3 IoT-specific analytics platforms

Platform	Description	Key features
AWS IoT Analytics	Fully managed service for processing and analyzing IoT data at scale.	Built-in ML support, automated data processing, and integration with AWS cloud.
Google Cloud IoT Core	Cloud-based solution for connecting, monitoring, and analyzing IoT data.	Secure communication, real-time analytics, and AI-powered insights.
Azure IoT Hub	Microsoft's IoT platform for device management and data analytics.	Secure cloud integration, edge analytics, and AI-driven insights.
IBM Watson IoT Platform	AI-powered IoT analytics and automation platform.	Cognitive computing, ML integration, and real-time monitoring.
ThingSpeak	MATLAB-based IoT analytics platform for time-series data.	Data visualization, MATLAB analytics, and IoT device management.

Several cloud-based and enterprise-level platforms are specifically designed to manage IoT data. These platforms offer data collection, device management, and analytics capabilities. Table 3 describes the platforms that are especially IoT-oriented in terms of analytics solutions like AWS IoT Analytics and Azure IoT Hub. IoT-oriented platforms are complete solutions in terms of IoT device management, data intake, or AI-powered analytics. IoT-oriented platforms are complete solutions in terms of IoT devices' management, data intake, or AI-powered analytics.

Machine learning (ML) and artificial intelligence (AI) are being used extensively in predictive modeling and decision-making in the IoT. These platforms are highly effective at unearthing key information from IoT data. The utilization of ML and AI platforms like TensorFlow, PyTorch, and Scikit-learn in IoT analytics is depicted in Table 4. ML and AI platforms make IoT analytics more effective in automation, anomaly detection, and intelligent decision-making.

Data visualization is crucial to IoT analytics, allowing stakeholders to interpret and monitor IoT data in real-time. BI tools provide interactive dashboards, reporting, and trend analysis. As presented in Table 5, Tableau, Power BI, and Grafana enable interactive monitoring and reporting of IoT data trends. These tools help transform raw IoT data into actionable insights, supporting industry decision-making.

Table 4 Machine learning and frameworks for IoT analytics

Framework	Description	Key features
TensorFlow	An open-source ML framework widely used for deep learning applications.	Neural networks, scalability, and GPU acceleration.
PyTorch	Deep learning framework optimized for AI-driven IoT applications.	Dynamic computation graphs, cloud integration, and AI model training.
Scikit-learn	Lightweight ML library for clustering, regression, and classification.	Wide range of ML algorithms, user-friendly API, and data preprocessing tools.
Microsoft ML.NET	ML framework for .NET developers integrating AI into IoT applications.	Scalable ML models, deep learning support, and IoT automation.
Keras	High-level API to build deep learning solutions with TensorFlow backend.	Simple API supports multiple ML backends and rapid prototyping.

Table 5 Data visualization and tools for IoT analytics

Tool	Description	Key features
Tableau	Powerful visualization tool for big data and IoT analytics.	Interactive dashboards, continuous IoT monitoring, and drag-and-drop analytics.
Power BI	Microsoft's BI tool for data visualization and reporting.	AI-driven insights, integration with Azure IoT, and real-time analytics.
Grafana	Open-source visualization tool for monitoring IoT sensor data.	Real-time graphs, an alerting system, and support for multiple data sources.
Google Data Studio	Free cloud-based visualization tool for IoT analytics.	Customizable dashboards, Google Cloud integration, and real-time reporting.
Kibana	A visualization tool for Elasticsearch IoT data analytics.	Time-series analysis, security monitoring, and anomaly detection.

4 Techniques for IoT Data Analytics

The scale and complexity of IoT data require advanced analytics methods that enable effective and timely discovery of valuable information. The process of IoT data analysis involves various phases, starting with data preprocessing and integration, as explained in Table 6, including preprocessing methods like data cleaning, transformation, fusion, and compression. Raw data from the

Table 6 Techniques for IoT data analytics

Category	Technique	Description	Applications
Data preprocessing and integration	Data cleaning	Removes noise, missing values, and duplicate records to improve data quality.	Smart healthcare, industrial IoT, and anomaly detection.
	Data transformation	Converts raw data into structured formats (normalization, encoding).	Sensor fusion and machine learning preprocessing.
	Data fusion	Combines multiple data sources to enhance accuracy and reliability.	Smart cities, autonomous systems, and environmental monitoring.
	Data compression	Reduces data size while maintaining information integrity.	IoT cloud storage and energy-efficient edge processing.
	Data synchronization	Ensures real-time consistency between distributed IoT data sources.	Industrial automation, logistics, and supply chain management.
Machine learning	Supervised learning	Labeled datasets train predictive models (e.g., decision trees, SVM, neural networks).	Predictive maintenance, healthcare monitoring, and traffic forecasting.
	Unsupervised learning	Identifies patterns in unlabeled data using clustering and anomaly detection (e.g., K-means, DBSCAN).	Network intrusion detection, fraud detection, and sensor clustering.
	Deep learning	Uses multi-layered neural networks for feature extraction and classification (e.g., CNN, RNN, LSTM).	Image and video analytics, speech recognition, and autonomous vehicles.
	Reinforcement learning	Optimizes decision-making through trial and error in dynamic environments.	Robotics, smart grid optimization, and adaptive traffic control.
	Federated learning	Decentralized ML where IoT devices train models without sharing raw data, enhancing privacy.	Healthcare IoT, edge AI, and distributed computing.
Real-time and edge analytics	Complex event processing	Detects patterns in streaming data and triggers real-time actions.	Smart surveillance, fraud detection, and industrial automation.
	Edge AI	Deploys AI models directly on IoT edge devices for decentralized analytics.	Smart homes, autonomous vehicles, and wearable health monitoring.

(*Continued*)

Table 6 Continued

Category	Technique	Description	Applications
	Fog computing analytics	Distributes computation between IoT edge and cloud for reduced latency.	Smart manufacturing and intelligent traffic management.
	Data stream processing	Continuously processes IoT data streams for instant insights.	Financial transactions and real-time anomaly detection.
	Distributed computing	Uses multiple nodes to process IoT data in parallel.	Smart grid analytics and environmental monitoring.
Anomaly detection	Statistical methods	Use thresholds, Z-score, and standard deviation to detect outliers.	Network traffic monitoring and industrial equipment failures.
	Machine learning-based anomaly detection	Uses classification and clustering models to detect abnormal patterns.	Cybersecurity, fraud detection, and predictive maintenance.
	Autoencoder	Learns compressed representations of normal behavior and flags deviations.	Healthcare diagnostics and smart city monitoring.
	Isolation forest	Detects anomalies by isolating outliers using decision trees.	Energy consumption analysis and fault detection.
	Time-series analysis	Identifies unexpected trends in IoT sensor data using ARIMA and LSTM.	Weather forecasting and industrial process monitoring.
Predictive maintenance	Time-series forecasting	Uses historical sensor data to predict future failures.	Smart manufacturing and industrial IoT.
	Condition monitoring	Continuously tracks equipment performance metrics.	Power grids and wind turbines.
	Vibration analysis	Detects faults using vibration sensor data.	Mechanical systems and transportation infrastructure.
	AI-based fault prediction	Uses ML models to assess equipment health and predict failures.	Automotive industry and aerospace.
	Sensor fusion for predictive maintenance	Combines data from multiple sensors to improve prediction accuracy.	Smart factories, the oil and gas industry.

IoT resembles undifferentiated data with attributes that include missing data, noise data, redundancy in data, and gaps in data repositories; therefore, data preprocessing must be carried out. After data preprocessing or processing, data from the IoT undergoes analysis in ML through various approaches, such as unsupervised learning and supervised learning.

Since IoT devices continuously generate high-velocity data streams, real-time and edge analytics techniques are crucial for immediate decision-making and reduced latency. For real-time applications, IoT analytics relies on streaming and edge-based processing, which includes approaches like complex event processing (CEP), edge AI, and fog computing to facilitate low-latency decision-making. Additionally, IoT systems require robust anomaly detection mechanisms to employ statistical methods, ML models, autoencoders, and isolation forests to identify irregular patterns and security threats in IoT data streams.

Finally, predictive maintenance solutions apply time series forecasting analysis, vibration analysis, and AI-enabled predictive failure analysis and predictive maintenance techniques in predictive maintenance solutions in order to anticipate failures and provide optimized functional use for Industry IoT. The analytics techniques mentioned above can increase IoT efficiency, security, and decision-making in smart cities, healthcare, manufacturing, and the environment.

5 Findings

Despite the fast evolution of IoT data analytics technologies and solutions, there are several challenges that affect the efficiency, scalability, and security of the process. This can be attributed to the vast volumes and differing velocities of data generated from IoT technologies. However, the power required to process and analyze such data streams remains a challenge in IoT data analytics. This section highlights the challenges faced in IoT data analytics.

The scalability issue arises from the sheer amount of data IoT produces. The interconnection of billions of devices generates constant data streams at high frequencies. The data cannot be processed by traditional data storage solutions or infrastructure due to the uncontrolled influx of unstructured and structured data. The issue gets more complicated if data analysis involves real-time processing. The latency involved in cloud computing infrastructures can be problematic for real-time applications like robotics or new technologies in the medical sector.

Moreover, resource-limited IoT devices, especially those in the edge domain, may not possess sufficient energy, memory capacity, or computing power either. The problem may be mitigated through edge computing in IoT devices. However, the requirement for powerful but more resource-friendly

analytics frameworks emerges as another challenge in this sector. The process of data filtration or compression would be necessary in this context.

Security remains a pressing concern in data analysis in IoT networks owing to the distributed nature of IoT networks and the sensitive data involved. This is especially true considering the fact that many IoT devices do not have inherent security measures in place; hence, they are prone to various cyber threats like data breaches, denial-of-service (DoS) attacks, and malware. The absence of standardized security protocols from various IoT providers only worsens the situation.

Privacy-related issues are more relevant in domains like healthcare, smart homes, and industry use cases of IoT. This is because personal data or business-critical data is processed constantly in such domains. This may cause significant losses if unauthorized access occurs. Further, the increased threat level with both edge computing and fog computing involves alternative threats like man-in-the-middle attacks due to multiple computing nodes in fog computing. To overcome the privacy-related challenges in IoT analysis in domains like healthcare and industry use cases, methods like multi-party encryption and blockchain-related integrity verification of data can be adopted.

IoT data is heterogeneous, inconsistent, or incomplete in nature. This causes data quality-related challenges. Data is collected from various sources like sensors, smart devices, or manufacturing machines, each with a different resolution or sampling rate. Noisy data may affect analytics or decision-making accuracy.

Moreover, IoT applications are often challenged by the issue of integrating data from legacy systems. The absence of common data protocols and standards makes data aggregation and analysis from different IoT sources difficult. For example, applications in smart cities are faced with data from different municipal authorities that use different protocols and platforms, such as transport, energy, and surveillance systems. To overcome such challenges, preprocessing techniques like cleaning, normalizing, fusing, and transforming are required. This would help ensure effective data integration by implementing standard IoT frameworks or middleware solutions.

The power constraints of IoT devices, especially those installed in distant or battery-powered setups, create many challenges in data analysis. The continuous process of data transmission, storage, and analysis consumes lots of power in IoT devices. This issue is more pronounced in various applications like animal tracking systems, industry sensing applications, and

disaster response solutions. Recharging or replacing batteries in such devices is not feasible.

Cloud computing and analysis methods involve continuous data streaming to servers. This consumes considerable bandwidth and power. Edge computing may address this particular problem by allowing IoT devices to cope with data analysis through computation done inside IoT devices. However, there would be compromises in computational power. The designer's objective would be to increase the life span of IoT devices while ensuring accuracy in analysis.

The absence of universal IoT standards causes challenges in the analysis of IoT data. This is due to the use of customized communication standards, data formats, and cybersecurity protocols by IoT device manufacturers. This caused a limitation in the creation of scalable IoT data analysis solutions across different industries. For instance, smart grids, healthcare information systems, or industry IoT communication networks may employ various communication protocols like MQTT, CoAP, OPC-UA, or HTTP. This further causes difficulties with smooth data transfer. Additional complications may be due to variations in the data structure and/or format.

For improving the level of interoperability, there are current attempts to develop common standards for IoT applications (namely IEEE P2413, oneM2M, and OpenIoT). The use of semantic Web technologies or open-source IoT platforms may also contribute to seamless data integration and analysis. Many IoT use cases require real-time decision-making, like self-driving cars, manufacturing automation, or intelligent healthcare networks. However, analysis in the cloud involves latency in transferring data over the network. The challenge in real-time IoT analysis involves performing calculations with high data volumes and reducing latency.

Real-time processing of IoT data needs optimized stream processing platforms, event-oriented architectures, and smart caching systems. By utilizing edge AI, distributed processing, and 5G networks, latency challenges can be mitigated, in addition to improving the functionality of real-time decision-making.

6 Discussion

The constantly evolving nature of IoT ecosystems requires future research in areas such as scalability, security, real-time processing, data quality, and energy efficiency. The integration of AI, edge computing, blockchain technology, federated learning, and quantum computing provides new avenues

that can be exploited in future research to further enhance data analytics in IoT ecosystems. This section presents information regarding the key research areas responsible for IoT analytics.

AI and ML technologies are transforming IoT data analysis in terms of immediate decision-making, predictive analysis-enabled maintenance, and intelligent automation. On the other hand, as shown in Table 7, there are many research gaps in current ML models in terms of improving efficiency, interpretability, and adaptability in resource-constrained IoT networks. Currently, XAI research focuses on improving the interpretability and trustworthiness of AI decision-making solutions in applications like healthcare and industry automation. Moreover, research in federated learning may contribute towards distributed ML model training and privacy preservation in IoT networks. Further research in deep learning approaches may support real-time IoT data analysis in terms of immediate decision-making with minimal computational complexity.

The use of blockchain technology brings together data management that is both decentralized and tamper-proof, and is thus ideal for securing IoT ecosystems. Currently, research into blockchain technology involves the use of simplified models of blockchain for IoT devices. Other areas include data provenance in secure tracking and AI usage in improving fraudulent and anomaly detection. Trustless automation is another benefit that results from the use of smart contracts on blockchain. This reduces dependence on centralized authorities in the IoT ecosystem.

With the ever-increasing needs of IoT applications in terms of low latency and real-time processing, fog computing has recently been projected as an alternative to traditional cloud computing analytics. Despite the growing interest in both edge and fog computing paradigms due to their feasibility in IoT applications requiring low latency in real-time processing and analysis, various challenges persist in the current state of both computing platforms. These include insecurity in fog computing and ensuring effective workload distribution in both computing paradigms while overcoming the restrictions imposed by power consumption.

Quantum computing could be revolutionary in the analysis of IoT data, with the power to substantially increase computing efficiency and support sophisticated encryption methods. Future studies should target quantum-aided ML to boost pattern detection and prediction analysis in voluminous IoT data. Moreover, quantum cryptography provides unbreakable encryption methods, making IoT even more secure from cyberattacks. Quantum-aided optimization methods may also increase resource allocation and network

Table 7 Future research directions in IoT analytics

Aspect	Research Area	Key Focus	Potential Impact
AI driven	Explainable AI for IoT	Enhancing transparency and interpretability of AI models.	Increases trust in AI-driven decisions for healthcare and industrial automation.
	Federated learning for IoT	Enabling decentralized ML without sharing raw data.	Enhances privacy and reduces bandwidth usage in smart cities and healthcare.
	Lightweight deep learning models	Optimizing ML models for edge and IoT devices.	Improves real-time analytics while conserving computational resources.
	Self-learning AI models	Adaptive AI systems that continuously learn from new data.	Enhances predictive accuracy in dynamic IoT environments.
Blockchain based	Lightweight blockchain for IoT	Reducing computational overhead for IoT devices.	Enhances security without overloading resource-constrained sensors.
	Blockchain-enabled data provenance	Tracking and verifying IoT data sources.	Ensures data authenticity in critical sectors (e.g., supply chain, healthcare).
	Integration of AI and blockchain	Using AI for blockchain-based anomaly detection.	Improves fraud detection and cybersecurity in IoT networks.
	Smart contracts for IoT automation	Enabling automated decision-making via blockchain-based contracts.	Reduces reliance on centralized authorities, ensuring trustless operations.
Edge and fog computing	Energy-efficient edge AI	Optimizing ML models for edge devices.	Reduces power consumption while maintaining analytical accuracy.
	Edge-based federated learning	Privacy-preserving distributed ML at the edge.	Enables AI-driven analytics while minimizing data transmission.
	Fog-enabled collaborative analytics	Distributing analytics across fog nodes.	Improves real-time processing for industrial and healthcare IoT.
	Edge security mechanisms	Enhancing security at edge computing nodes.	Prevents cyber threats and data breaches in decentralized architectures.

(*Continued*)

Table 7 Continued

Aspect	Research Area	Key Focus	Potential Impact
Quantum computing	Quantum-enhanced ML for IoT	Using quantum algorithms to accelerate ML tasks.	Improves pattern recognition and predictive modeling for large-scale IoT data.
	Quantum cryptography for IoT security	Leveraging quantum encryption techniques.	Enhances security against quantum-resistant cyber threats.
	Quantum-assisted optimization for IoT	Solving complex optimization problems in IoT networks.	Improves routing, load balancing, and resource allocation.
Standardization	Unified IoT data standards	Establishing common data structures and protocols.	Enhances interoperability across heterogeneous IoT devices.
	Semantic Web for IoT integration	Using ontologies to enable machine-readable IoT data.	Facilitates automated and intelligent IoT analytics.
	Cross-industry IoT frameworks	Developing unified architectures for different sectors.	Enables interoperability across healthcare, smart cities, and industrial IoT.
Energy efficiency	Green AI for IoT analytics	Developing AI models with reduced energy consumption.	Enables sustainable AI-driven analytics for IoT.
	Self-powered IoT sensors	Exploring energy harvesting technologies.	Enhances the longevity of IoT deployments in remote areas.
	Adaptive data sampling and compression	Reducing redundant data transmission.	Lowers bandwidth usage and extends battery life.

routing in IoT. At this stage of development, quantum computing holds vast potential in dealing with IoT challenges with enhanced speed and security.

The standardization of data types, communication mechanisms, and security systems in IoT applications has emerged as one of the key challenges toward the successful implementation of IoT solutions. Future research and development endeavors should be centered on ensuring data standardization in IoT applications. This would provide support for the seamless transfer of data among various IoT ecosystems. The use of Semantic Web technologies

would also help support machine data transfer in IoT applications. The standardization process would be essential in improving the scalability and reliability of IoT applications in the future.

There is a rising need for more sustainable and energy-efficient analytics in the IoT network. Future research should be based on the idea of implementing more green AI models that use less computational power and are more effective in analysis. Furthermore, IoT sensors that are self-powered using different forms of harvesting methods may increase the lifetime of the devices and eliminate the need for periodic replacements of batteries. Techniques that are adaptive in data sampling and compression may further enhance data usage in IoT networks and hence minimize each IoT analysis system's carbon footprint.

7 Conclusion

The IoT has experienced rapid growth in recent years, allowing data generation in volumes hitherto unexplored. This necessitates the use of effective data analytics strategies for data analysis. IoT data analysis tools have significant applications for real-time data analysis in IoT scenarios such as smart cities, medical, industry, and environmental monitoring. The voluminous nature of IoT data, however, makes it challenging to analyze in real-time. This paper provides an in-depth review of IoT data analytics, concentrating on the tools, techniques, and challenges of processing and analyzing large-scale IoT-generated data. We explored various analytics approaches, including descriptive, real-time, predictive, and prescriptive analytics. We discussed the role of big data platforms, cloud computing, and ML frameworks in IoT data processing. Furthermore, we examined key challenges such as data security, privacy risks, computational constraints, and standardization issues, which must be addressed to ensure the efficient deployment of IoT analytics solutions. To overcome these challenges, future research directions should focus on:

- AI-driven IoT analytics emphasizing explainable AI, federated learning, and lightweight ML models to enhance predictive capabilities while preserving privacy.
- Blockchain-based IoT security that ensures data integrity, transparency, and decentralized authentication to mitigate cybersecurity risks.
- Edge and fog computing architectures improving real-time data processing and reduce network congestion by distributing computational workloads closer to IoT devices.

- Quantum computing applications leveraging quantum algorithms for faster data processing, optimization, and quantum-safe encryption.
- Standardization efforts, establishing common protocols and frameworks to enable seamless IoT data integration across industries.
- Sustainable IoT analytics, developing energy-efficient AI models, and self-powered IoT sensors to support eco-friendly and long-term IoT deployments.

With the afore-mentioned key areas in mind, IoT data analytics solutions may gain more optimized efficiency solutions in terms of more secure systems and more scalable solutions for bringing about the next generation of intelligent data ecosystems through IoT. Breakthroughs in areas like AI, edge computing, blockchain, and quantum computing will define the future of IoT data analytics.

References

[1] J. S. Yalli, M. H. Hasan, and A. Badawi, "Internet of things (iot): Origin, embedded technologies, smart applications and its growth in the last decade," *IEEE access*, 2024.

[2] T. Ahmad and D. Zhang, "Using the internet of things in smart energy systems and networks," *Sustainable Cities and Society*, vol. 68, p. 102783, 2021.

[3] J. Zandi, A. N. Afooshteh, and M. Ghassemian, "Implementation and analysis of a novel low power and portable energy measurement tool for wireless sensor nodes," in *Electrical Engineering (ICEE), Iranian Conference on*, 2018: IEEE, pp. 1517–1522.

[4] R. Chataut, A. Phoummalayvane, and R. Akl, "Unleashing the power of IoT: A comprehensive review of IoT applications and future prospects in healthcare, agriculture, smart homes, smart cities, and industry 4.0," *Sensors*, vol. 23, no. 16, p. 7194, 2023.

[5] D. Adhikari, W. Jiang, J. Zhan, D. B. Rawat, and A. Bhattarai, "Recent advances in anomaly detection in Internet of Things: Status, challenges, and perspectives," *Computer Science Review*, vol. 54, p. 100665, 2024.

[6] A. O. Al-Mashhadani and M. Al-Khafajiy, "Big data analytics for IoT: Technologies, importance, and algorithms," in *Empowering IoT with Big Data Analytics*: Elsevier, 2025, pp. 15–44.

[7] B. Pourghebleh, V. Hayyolalam, and A. Aghaei Anvigh, "Service discovery in the Internet of Things: review of current trends and research challenges," *Wireless Networks*, vol. 26, no. 7, pp. 5371–5391, 2020.

[8] T. M. Mengistu, T. Kim, and J.-W. Lin, "A Survey on Heterogeneity Taxonomy, Security and Privacy Preservation in the Integration of IoT, Wireless Sensor Networks and Federated Learning," *Sensors*, vol. 24, no. 3, p. 968, 2024.

[9] S. S. Mahadik, P. M. Pawar, and R. Muthalagu, "Heterogeneous IoT (HetIoT) security: techniques, challenges and open issues," *Multimedia Tools and Applications*, vol. 83, no. 12, pp. 35371–35412, 2024.

[10] D. Fawzy, S. M. Moussa, and N. L. Badr, "The internet of things and architectures of big data analytics: Challenges of intersection at different domains," *IEEE Access*, vol. 10, pp. 4969–4992, 2022.

[11] B. Pourghebleh, N. Hekmati, Z. Davoudnia, and M. Sadeghi, "A roadmap towards energy-efficient data fusion methods in the Internet of Things," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 15, p. e6959, 2022, doi: https://doi.org/10.1002/cpe.6959.

[12] W. Li et al., "A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system," *Mobile networks and applications*, vol. 26, pp. 234–252, 2021.

[13] M. Talebkhah, A. Sali, M. Marjani, M. Gordan, S. J. Hashim, and F. Z. Rokhani, "IoT and big data applications in smart cities: recent advances, challenges, and critical issues," *IEEE Access*, vol. 9, pp. 55465–55484, 2021.

[14] Y. Himeur et al., "AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives," *Artificial Intelligence Review*, vol. 56, no. 6, pp. 4929–5021, 2023.

[15] B. Pourghebleh and V. Hayyolalam, "A comprehensive and systematic review of the load balancing mechanisms in the Internet of Things," *Cluster Computing*, vol. 23, no. 2, pp. 641–661, 2020.

[16] M. Paramesha, N. L. Rane, and J. Rane, "Big data analytics, artificial intelligence, machine learning, internet of things, and blockchain for enhanced business intelligence," *Partners Universal Multidisciplinary Research Journal*, vol. 1, no. 2, pp. 110–133, 2024.

[17] C. S. Babu, G. M. AV, S. Lokesh, A. Niranjan, and Y. Manivannan, "Unleashing IoT data insights: Data mining and machine learning techniques for scalable modeling and efficient management of IoT," in *Scalable Modeling and Efficient Management of IoT Applications*: IGI Global, 2025, pp. 153–188.

[18] I. H. Sarker, "Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective," *SN Computer Science*, vol. 2, no. 5, p. 377, 2021.

[19] B. Pourghebleh and N. J. Navimipour, "Data aggregation mechanisms in the Internet of things: A systematic review of the literature and recommendations for future research," *Journal of Network and Computer Applications*, vol. 97, pp. 23–34, 2017, doi: https://doi.org/10.1016/j.jnca.2017.08.006.

[20] S. Patil, S. Chintamani, B. H. Dennis, and R. Kumar, "Real time prediction of internal temperature of heat generating bodies using neural network," *Thermal Science and Engineering Progress*, vol. 23, p. 100910, 2021.

[21] B. G. Deepthi, K. S. Rani, P. V. Krishna, and V. Saritha, "An efficient architecture for processing real-time traffic data streams using apache flink," *Multimedia Tools and Applications*, vol. 83, no. 13, pp. 37369–37385, 2024.

[22] D. Roy, R. Srivastava, M. Jat, and M. S. Karaca, "A complete overview of analytics techniques: descriptive, predictive, and prescriptive," *Decision intelligence analytics and the implementation of strategic business management*, pp. 15–30, 2022.

[23] R. Al-amri, R. K. Murugesan, M. Man, A. F. Abdulateef, M. A. Al-Sharafi, and A. A. Alkahtani, "A review of machine learning and deep learning techniques for anomaly detection in IoT data," *Applied Sciences*, vol. 11, no. 12, p. 5320, 2021.

[24] V. Hayyolalam, B. Pourghebleh, M. R. Chehrehzad, and A. A. Pourhaji Kazem, "Single-objective service composition methods in cloud manufacturing systems: Recent techniques, classification, and future trends," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 5, p. e6698, 2022.

[25] V. Hayyolalam, B. Pourghebleh, A. A. Pourhaji Kazem, and A. Ghaffari, "Exploring the state-of-the-art service composition approaches in cloud manufacturing systems to enhance upcoming techniques," *The International Journal of Advanced Manufacturing Technology*, vol. 105, pp. 471–498, 2019.

[26] M. Laroui, B. Nour, H. Moungla, M. A. Cherif, H. Afifi, and M. Guizani, "Edge and fog computing for IoT: A survey on current research activities & future directions," *Computer Communications*, vol. 180, pp. 210–231, 2021.

[27] S. Mukherjee, S. Gupta, O. Rawlley, and S. Jain, "Leveraging big data analytics in 5G-enabled IoT and industrial IoT for the development of sustainable smart cities," *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 12, p. e4618, 2022.

[28] A. Medina-Pérez, D. Sánchez-Rodríguez, and I. Alonso-González, "An internet of thing architecture based on message queuing telemetry transport protocol and node-red: A case study for monitoring radon gas," *Smart Cities*, vol. 4, no. 2, pp. 803–818, 2021.

[29] V. Hayyolalam, B. Pourghebleh, and A. A. Pourhaji Kazem, "Trust management of services (TMoS): investigating the current mechanisms," *Transactions on Emerging Telecommunications Technologies*, vol. 31, no. 10, p. e4063, 2020, doi: https://doi.org/10.1002/ett.4063.

[30] K. Ngcobo, S. Bhengu, A. Mudau, B. Thango, and M. Lerato, "Enterprise data management: Types, sources, and real-time applications to enhance business performance-a systematic review," *Systematic Review|September*, 2024.

[31] Y. Liu, W. Yu, W. Rahayu, and T. Dillon, "An evaluative study on IoT ecosystem for smart predictive maintenance (IoT-SPM) in manufacturing: Multiview requirements and data quality," *IEEE Internet of Things Journal*, vol. 10, no. 13, pp. 11160–11184, 2023.

Biography

Yun Liu graduated from University of South Florida with a master's degree in computer science. She currently works in School of Mathematics and Computer Science, Quanzhou Normal University as a teaching assistant. She has published three papers. Her research interest includes IoT, deep learning, image recognition and intelligent computing.