International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Artificial Intelligence and Internet of Things Integration Taxonomy

Poorana Senthilkumar S [a*], Anand T R [b], Mohana Priya T [c] and Karthik S [d]

[a*] *Assistant Professor, Department of Computer Applications, Dr. N.G.P. Arts and Science College, Coimbatore, India*
[b] *Assistant Professor, Department of Computer Technology, Dr. N.G.P. Arts and Science College, Coimbatore, India*
[c] *Assistant Professor, Department of Computer Science, Christ (Deemed to be University), Bengaluru, India*
[d] *Assistant Professor, Department of Computer Science with Data Analytics, PSG College of Arts and Sceince, Coimbatore, India*
DOI: https://doi.org/10.55248/gengpi.5.0424.1082

A B S T R A C T

The Internet's transformation into the Internet of Things (IoT) is becoming the core of real-time and advanced technology appliances nowadays in all the fields. These real-time connected devices generate a large quantity of data. Artificial Intelligence (AI) represents the new direction of computer science, encompassing many branches including machine learning (ML), deep learning (DL), natural language processing (NLP), and many more. It can be defined by the term 'automated intelligence' as it is a program that handles prediction, forecasting, and decision-making without human interference. The automated intelligent processing and analysis of this large data (Big Data) is the key to developing smart IoT applications. ML may be used in cases where the desired effect is defined (supervised learning), where data itself is not defined beforehand (unsupervised learning), or where learning is the outcome of the interaction between the learning model and the environment (reinforcement learning). In this article, we present and discuss an important taxonomy of ML algorithms that can be used with IoT applications. Furthermore, we illustrate how various ML techniques are used to derive higher-level information from the data. Finally, it provides a basic line of investigation into real-world IoT data characteristics that involve an integration of AI with IoT.

Keywords: Artificial Intelligence, Data Source, Deep learning, Internet of Things, Machine Learning,

1. Introduction

The automated smart applications domains are always looking for ways to improve efficiency, fault tolerance, minimize downtime, costs. Today's digital conveniences attained by end ¬users are due to the endless efforts of intelligent researchers across the world. These scientists have made humans life smart and comfortable. For example, let's start with daily life that deals with different types of routine tasks and activities. The time of wake up and till we get ready for our work, we use various smart gadgets like mobile chargers, microwaves, hairdryers, and espresso makers, to name a few. All these appliances were developed to reduce human efforts and to save time. Similarly, various Internet -enabled smart devices such as computers, smartphones, and printers are connected together in order to improve workplace operations as the Internet enables intelligent communication and control by end ¬users anywhere, anytime through wire or wireless network connection. [1] Hence this example conclude that the Internet plays an important role in humans' daily life. This interconnection of smart devices with Internet connection is termed as the IoT as summarized in Figure 1.

These Internet connected objects (Things) are popularly known as IoT Devices. These inter connected devices were sharing endless information which can be embedding the intelligence into the devices is the task of artificial intelligence (AI). In other aspects, AI cutting edge approaches such as federated learning is used to aid in privacy preservation of IoT systems data. The concept AI was introduced in the 1950s, when a group of founders in the computer science research field started to challenge whether machines could be "think" - A query, the ramifications of which are still being discussed today. A succinct description of AI - the attempt to automate the intellectual tasks usually performed by the humans. AI is therefore a general field involving ML and DL, but it also comprises several other methods that do not require learning ever again. In the 1960s, the US Department of Defense became involved and started teaching computers to emulate basic human thought. Long before Siri, Alexa and Cortana had become common names, DARPA created smart personal assistants in 2003 [2].

This earlier work is recognized to automation and formal thought, involving DSS and smart search systems to add and improve human expertise that we see on today's computers. AI is now more widely popular due to higher volume of data, advanced analytics, and improved computational and storage capacity. It is a generic term for technologies which have intelligent interfaces that communicate actively. If things connect with each other and have smart interfaces, they can have new functionality outside their own current characteristics. With the growth of IoT, applications have grown smarter and interconnected devices are being used in all areas of various applications. Because the amount of data generated grows, ML approaches can be used to further improve the intelligence and effectiveness of applications [3].

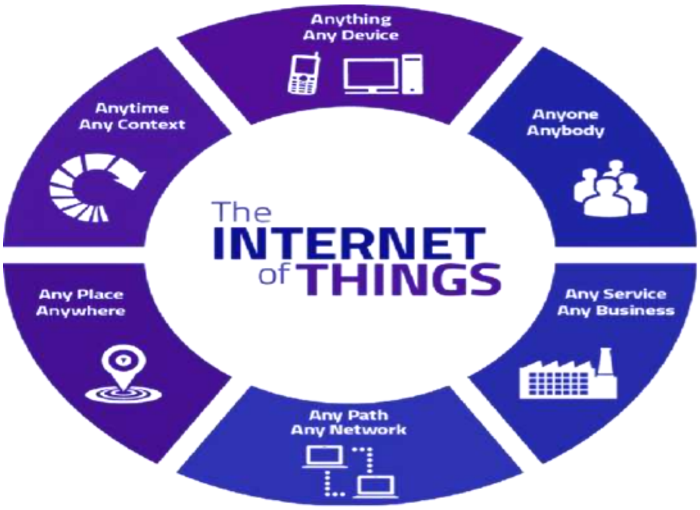

Figure 1. Internet of Things

The AI is not a single component it contains another two more interrelated techniques which are known as ML and DL. The following Figure 2 illustrate the relationship between AI, ML and DL and AI employments requires the related interest technique to accomplish the intelligence task.

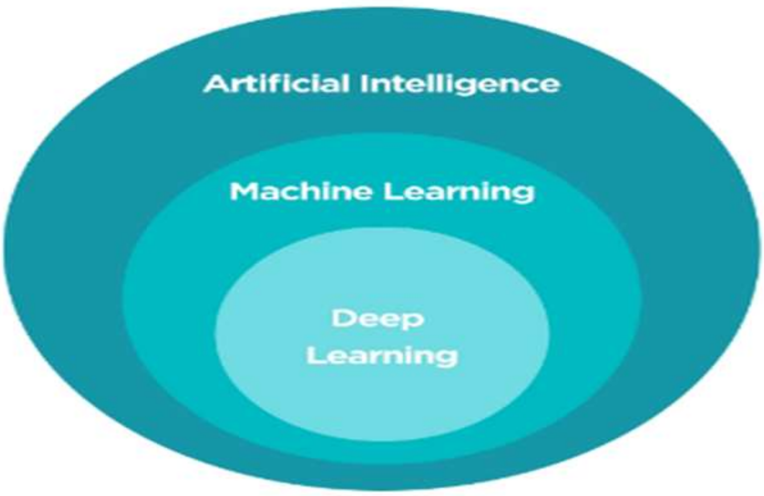

Figure 2. Relationship among AI, ML and DL

1.1. Machine Learning (ML)

ML emerges from its query: can a machine go beyond “what we know how to order it to do” And understand how to execute a specific task on its own way? Does the machine take us by surprise? Can a system automatically acquire such rules by analyzing data, rather than developers writing data processing rules by hand? That challenge unlocks the door to a modern paradigm of programming. The model of representational AI, human input rules (a program) and data to be processed according to these rules are used in traditional programming, and output responses, and the same is presented in Figure. 3. With ML, humans input data, the predicted data responses, and the rules come out. To generate original responses, these rules can then be applied to the new data. For more than a century, traditional computer programming has existed, dating back to the 1800s with the first known computer programming. Traditional programming corresponds to any programs created manually that utilizes data input and executes to produce output on a computer. But for years now a specialized form of programming has revolutionized industry, especially in the areas of intelligence and integrated analytics. Input data and results are supplied to an ML algorithm, often referred to as augmented analytics. This produces good observations which will be used to predict potential outcomes [4].

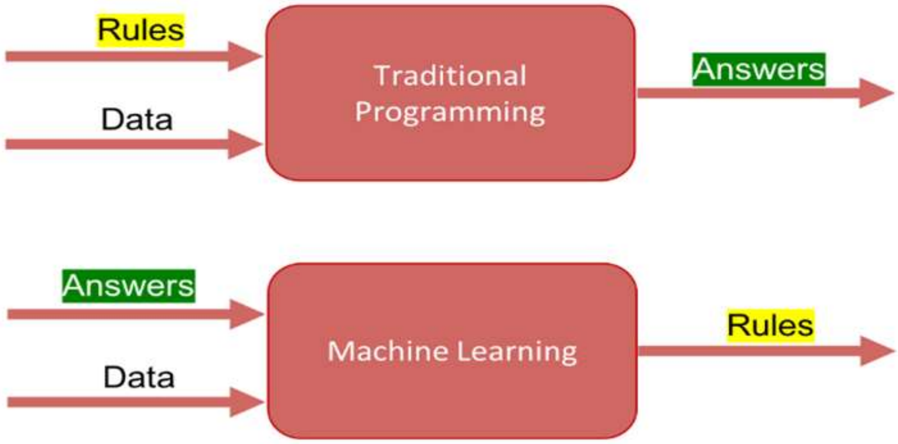

Figure 3. Traditional Programming vs Machine Learning paradigm.

1.2. Deep Learning (DL)

DL is a specific subfield of ML: a modern method to learn data representations that focuses learning continual stages of progressively expressive representations. Deep learning does not mean a deeper understanding of the methodology; relatively it refers to the concept of successive levels of representation. How many layers contribute to a model is called the model's depth. Layered representations of learning and hierarchical representations of learning may have been other suitable names for the field. In modern deep learning, multiple successive layers of the order of ten to hundred layers representations are frequently involved, and they all learn automatically from exposure to training data [5].

2. IoT Data Sources

The data generated by the IoT applications, users, devices, and sensors are fetched by the IoT infrastructure. IoT processes the data in such a manner that the meaningful information is retrieved, which in turns help initiate the smart activity. IoT data can be generated by any enterprise or organization or device by using IoT service or application. Based on the IoT application deployed, the IoT data sources are categorized as Industrial Data, Business Applications, Sensors and Devices [6]. These all the generated data should be processed for the efficient manner of utilize the IoT applications for refining the data [7]. The architecture for collecting data in IoT applications typically involves several layers and components working together to gather, process, and transmit data from various devices and sensors. Here's a breakdown of a typical IoT application data collection architecture:

Sensors and Devices: At the lowest level of the architecture are the sensors and devices embedded with various types of sensors (e.g., temperature, humidity, motion, GPS). These devices collect real-world data from their surroundings.

Edge Devices/Gateways: Edge devices or gateways act as intermediaries between the sensors/devices and the central processing system. They perform initial data filtering, preprocessing, and sometimes even basic analytics at the edge of the network. Edge computing helps reduce latency, conserve bandwidth, and improve data privacy and security.

Communication Protocols: IoT devices and edge gateways communicate with each other and the central system using various communication protocols such as MQTT, CoAP, HTTP, or proprietary protocols. These protocols ensure efficient and reliable data transmission over the network.

Connectivity: IoT devices typically connect to the internet or local network using wired (Ethernet, PLC) or wireless (Wi-Fi, Bluetooth, Zigbee, LoRaWAN, cellular) connectivity technologies, depending on the application requirements and environmental constraints.

Cloud Platform or Data Center: Data collected from IoT devices is transmitted to a centralized cloud platform or data center for storage, processing, and analysis. Cloud platforms provide scalable storage and computing resources, enabling organizations to handle large volumes of IoT data efficiently.

Data Ingestion and Storage: In the cloud platform or data center, incoming data is ingested, stored, and organized in databases or data lakes. Time-series databases are commonly used for storing IoT data due to their efficient handling of time-stamped data.

Data Processing and Analytics: Once the data is stored, it can be processed and analyzed to derive actionable insights. This may involve batch processing, real-time stream processing, or a combination of both. Machine learning algorithms may also be applied to analyze the data and extract patterns or predictive models.

Application Layer: The application layer consists of various software applications and services that consume the processed data and provide value-added services to end-users. This could include real-time monitoring and visualization dashboards, predictive maintenance systems, smart city applications, or industrial automation systems.

Security and Privacy: Security measures such as encryption, authentication, access control, and data anonymization are crucial components of IoT data collection architecture to protect sensitive data and ensure the integrity and confidentiality of communications.

Scalability and Flexibility: The architecture should be designed to scale seamlessly as the number of IoT devices and data volume grows. It should also be flexible enough to accommodate new devices, protocols, and applications as technology evolves.

An effective IoT application data collection architecture integrates hardware, software, communication, and security components to enable efficient and reliable data collection, processing, and utilization in various IoT applications. Data collections and analysis ensures that the data is ready to fit into the models of AIs ML [8]. The data at this stage is the level of computer responsiveness which flows from the processed data to make it more efficient and the IoT data sources are presented in Figure 4.

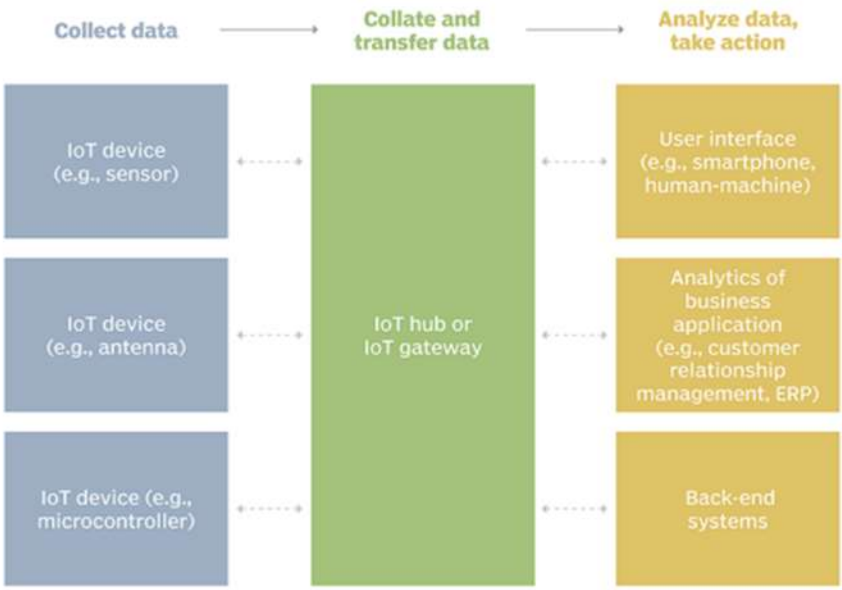

Figure 4. The IoT data Sources

There are four main categories of data analytics described as follows:

Descriptive: It defines the past by gathering raw data from various data sources to extract useful information. The result describes what has happened but does not describe why it happened. For example, a student may gather the marks of all his classmates to judge who has scored the maximum.

Diagnostic: It is the process to analyze the reason behind some event which has already occurred. It gives us the answer to why it has happened. For example, a student finds the reason why he had scored poorer marks.

Predictive: It is the technique of forecasting the future by considering the past instances. It is done to know what will happen. It considers various techniques like machine learning, which uses various models for prediction. For example, forecasting the temperature of another city after two days for a visit.

Prescriptive: It is the method to learn from the experiences to avoid the near future risk. It gives us the answer as to what should we do. It involves various techniques. For example, a student learns new techniques to answer the questions in an examination to achieve good marks.

3. ML Data Model

The data which is processed previously is collected and analyzed using the ML technique. This technique exploits the hidden insights of data and predicts the new result based on this self-learning model experience. There are various ML models like classification, regression, and clustering models. We train the processed data into one of these models, and then the model predicts the unseen data by learning from the previous data. This technique helps us to avoid failure and losses and improve the performance [9]. For example, predicting tomorrow’s whether and temperature can be done using ML for the IoT based real time applications [10]. For the prediction model the collected raw data were could not support the do the right decision. The data validation the following data-based processing are required to employee the ML model in the AI concept.

3.1. Data Preprocessing in ML

The direct collected data could not support to obtain the optimized result. The collected raw data are required some of the preprocessing to ensure the better result. The following steps are involved to perform the Data preprocessing [11].

3.1.1. Data Formatting

This step refers to the procedure required for formatting the data. As the data is collected from various sources like sensors and then converted by ADC, the format should be the same and consistent for all the data. For example, when you are practicing machine learning and download the dataset from Google, you may observe different types of files being downloaded, such as, .txt, .pl,. graph, and .csv. All the data needs to be formatted before it is fed to the machine learning models. Then using of software like R or RStudio or python requires a .csv file for the processing. Therefore, the data needs to be formatted in the .csv file.

3.1.2. Data Cleaning

The data should be clean enough for further processing. Data cleaning is the crucial step of data preprocessing that ensures the detection of duplicate and redundant values, irrelevant values, and missing values. All these issues should be resolved, by implementing of cleaning otherwise, it leads to the failure cases of machine learning models.

3.1.3. Duplicate Observations

Sometimes, the IoT sensors capture the same data again and again. For example, if you are leaving your child sleeping at home under CCTV footage and you leave for office. After returning home hours later, it won't be possible to see the complete footage of hours. The CCTV should be smart enough so that it should record only when a child moves, and not in the sleeping position is vary from time to time. The repetition of the same action should not be recorded. hence saving cost and time.

3.1.4. Irrelevant Observations

Collected data should not contain irrelevant information. For example, suppose you download the dataset of websites, and it contains information about the color of the website, which is irrelevant. It is best to detect such irrelevant observations and it is used to identify the unwanted data it will give right inputs to the ML models and avoid the irrelevant execution of ML.

3.1.5. Handling Missing Data

The collected data is in free form, and the values may consist of some 0s or NULL. It is very important to treat such values before the data is further processed. There are various ways to handle this constraint. A few of these are described as follows.

• Deleting Rows: Delete the complete column which contains approximately 70% to 90% of zero or NULL values.

• Replacing with Mean/Median/Mode: Substitute the missing value or the null value with some aggregate value, such as the mean.

3.1.6. Detecting Outliers

An Outliers are the instances which does not lie within a particular range. It is difficult to detect the outliers just by visualization. It depends upon the data and its features as to how relevant and useful the data is.

3.1.7. Feature Selection

Selecting the most relevant features can improve model performance and reduce computational complexity. Techniques such as correlation analysis, feature importance scores, and recursive feature elimination (RFE) can be employed for feature selection. It's essential to strike a balance between model complexity and predictive power.

Realtime data is always collected randomly. As in the analog form, it is difficult to say the number of classes it will form or the number of dimensions it will converge to after conversion. The resulting of digital form of data may contain an unbalanced number of cases in various dimensions. Balancing the data is done under the mentioned data preprocessing steps. These data preprocessing technique is effective and essential for preparing data for ML models, improving their performance, and ensuring their robustness. These techniques tailored to the specific characteristics of the real time dataset and the requirements of the ML task. In IoT applications, we collect datasets of different natures from various levels. All this data should be preprocessed for the integration of IoT and AI because AI always accomplishes tasks based on defined models.

4. Conclusion and Feature Direction

This article is primarily targeted for the beginners, researchers, and industrialists, those who are planning to design an IoT infrastructure with emphasis on how AI can help the system. The perception is the power to emphasize the upcoming facts of the technical world. So, intellectual computing is also an integral part of the system to make it smarter. This article provides information about potential IT strategies, IoT structure, and machine learning. It covers different types of ML dataset preprocessing models and it covers the details of machine learning terminology along with IoT applications. Data analytics play an important role as that is concerned with data processing activities such as data gathering, data preprocessing, and simulation technologies. Working on IoT devices and their applications are the reason behind the technical innovations the driving force behind technical innovations lies in exploring IoT devices and their applications. Looking ahead, the article aims to support the advancement of IoT applications by enabling real-time data integration for constructing simulation environments with preprocessed datasets.

REFERENCES

[1] D. Mocrii, Y. Chen, and P. Musilek, "IoT-based smart homes: A review of system architecture, software, communications, privacy and security," Internet of Things, vol. 1–2, pp. 81–98, 2018, doi: 10.1016/j.iot.2018.08.009.

[2] K. S. Mohamed, The Era of Internet of Things. Cham: Springer International Publishing, 2019. doi: 10.1007/978-3-030-18133-8.

[3] R. R. Kanna, "Machine Learning Algorithm to Identify the Fault Data Identification Using Multi-Class Support Vector Machine," Int. J. Futur. Revolut. Comput. Sci. Commun. Eng., vol. 12, no. 3, pp. 471–475, 2016, [Online]. Available: http://www.ijfrcsce.org/index.php/ijfrcsce/article/view/1709%0A

[4] V. Raut, A. Dhok, J. Sagwadiya, and S. More, "ADVANCEMENT IN AGRICULTURE WITH AI , ML AND IOT : AN OVERVIEW," vol. 10, no. 12, pp. 140–145, 2022.

[5] A. Ghosh, D. Chakraborty, and A. Law, "Artificial Intelligence in Internet of Things," pp. 1–11, 2018.

[6] M. Jindal, "Machine learning methods for IoT and their Future Applications," 2019.

[7] J. Azar, A. Makhoul, M. Barhamgi, and R. Couturier, "An energy efficient IoT data compression approach for edge machine learning," Futur. Gener. Comput. Syst., vol. 96, pp. 168–175, 2019, doi: 10.1016/j.future.2019.02.005.

[8] S. Ahmed, "IoT Based Smart Systems using Machine Learning (ML) and Artificial Intelligence (AI): Vulnerabilities and Intelligent Solutions," no. Icsit, pp. 56–61, 2022.

[9] A. Elhanashi, "Integration of Deep Learning into the IoT : A Survey of Techniques and Challenges for Real-World Applications," Electron., vol. 12, no. 24, pp. 3–10, 2023, doi: https://doi.org/10.3390/electronics12244925.

[10] K. Y. Lee and H. Park, "Data , Privacy , and Artificial Intelligence," SSRN Electron. J., no. July, pp. 1–43, 2023, doi: 10.2139/ssrn.4246578.

[11] N. Pandey, P. K. Patnaik, and S. Gupta, "Data Pre-Processing for Machine Learning Models using Python Libraries," Int. J. Eng. Adv. Technol., vol. 9, no. 4, pp. 1995–1999, 2020, doi: 10.35940/ijeat.D9057.049420.