

---

## A Comprehensive Literature Review on Data Analytics in IIoT (Industrial Internet of Things)

---

B Ramasubba Reddy<sup>\*1</sup>, A V L N Sujith<sup>2</sup>

<sup>1,2</sup> Department of CSE, S V Engineering College, Tirupati

Received: 13th October 2017 Accepted: 14th November 2017, Published: 31st December 2017

### Abstract:

IoT (Internet of Things) is the paradigm in which efforts are being made to enable seamless processing of information by integrating physical and digital world devices that can be used ubiquitously. Integration of IoT and cloud computing is the major research area where the data collected from the various high data rate sensors and aggregated with data obtained by cloud-enabled services further which is delivered to the end users in an effective way. Analytical mechanisms like predictive analysis, machine learning algorithms are applied to examine and derive conclusions from the Bigdata gathered from various sources. With the advent of IIoT (Industrial Internet of Things) which is primarily concerned with IoT applications developed for the manufacturing industry, generates a huge amount of data at a greater speed where there is a need for effective data analytics. This paper provides an outline of various architectural components of IIoT and enables with a comprehensive review of various research studies addressing Big data analytics in IIoT. Furthermore, it includes contemporary technological challenges to develop IIoT systems.

**Keywords:** IoT (Internet of Things), IIoT (Industrial Internet of Things), Data analytics, Edge analytics

### 1. Introduction

Over the past two centuries, the humankind has experienced several waves of innovations of those four major innovations are remarkable for their existence. In the timeline of the eighteenth-century evolution of steam engine in the production and manufacturing enabled a path for Industrial revolution. Further, in the 20<sup>th</sup>-century advent of the electricity, it has accelerated the process of industrial evolution. Furthermore, in the decade of

considered as the appliance of standard concept of Cyber-Physical systems in which data from several industrial viewpoints are closely collected and monitored from the physical space moreover synchronized with the cyberspace. The revelation of the Industrial Internet closely depends on the procedure of adopting advanced information as well as the communication technologies in manufacturing industries.

Numerous streams of technologies are enabled in the Industrial Internet, that includes industrial sensing, networking, and control, big data, cloud computing, data analytics, and security, etc. These technologies include diverse features of the industrial manufacturing as well as the production process with sensing, connection storage, analytics, automation storage, sensing, connection, automation, machine intelligence, and manufacturing. Nowadays the sensor technology is significantly decreased in terms of size as well as the cost making the Internet which is defined as a network of networks that connects several computers. The integration of Industrial and Internet revolution provide the foundation for Industrial Internet revolution.

1950s, several modern computing technologies came into existence with the introduction of the

The IIoT gained the interest of both the industries as well as the academia with the rapid growth in the technologies like Bigdata; cloud computing, Artificial intelligence etc. IIoT is often instrumentation of machines and processes technically and financially feasible.

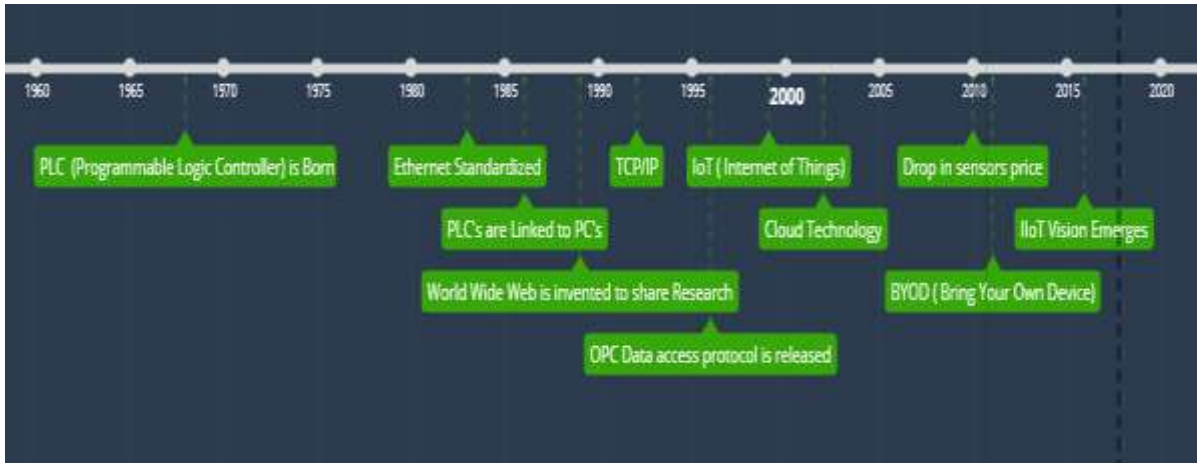


Figure 1. IIoT Vision Emergence Time line

### 1.1 IIoT Evolution:

Initially, sensors are the fundamental components of the evolution of IoT and IIoT. Sensors acquire the data in the analog format and convert them into digital data. With the advent of the ubiquitous computing as well as the data networking the usage of the sensors are dramatically increased and this transition firstly took place in the decade of 1990s. The capability of the digital sensors to transmit the data to remote servers made the storage and analysis of the data more intuitive. Figure 1 shows the timeline for the evolution of IIoT. Recent sensor technology enables the battery power as well as wireless communications using a superfluity of protocols that intelligently place the sensors in diverse locations. The drastic decrease in the cost of the IT (Information Technology) stack that includes the components to the individual system has now made the extensive use of the sensors economically feasible. The

Subsequent development is to aggregate whole sensor data to predict equipment failure or automate previously manual processes. In spite there lies the opportunity to optimize business practices based on novel insights obtained from sensor data and applied analytics. During the timeline in between 20<sup>th</sup> and 21<sup>st</sup> century the growth in ICT was very rapidly such that the era of new technologies that include sensor networks RFID, Artificial intelligence, 3D printing, cloud computing, Big data, IoT played their prominent role in enabling industrial revolution. Figure 1

shows the timeline features of the evolution of IIoT.

The significant improvement in these technologies improved the production rate in the industries by using the combination of these technologies in monitoring huge machines in the industries that may include sensing intelligence was improved by RFID and IoT, the networking levels are improved by wireless sensor network and cloud computing, multilevel decision was improved with the advent of artificial intelligence and Machine learning techniques, data monitoring and evaluation was done because of BigData analytics, controlling was done based on cyber-physical systems and finally manufacturing with 3D printing. The combination of these advanced ICT's enabled the transformation of traditional industrial monitoring to Industrial internet of things such that by enabling various challenges at both levels of academic research and industrial research communities.

The term Industry 4.0 was primarily used at the fourth industrial revolution that refers to Hanover Fair in 2011 and has gained much attention of both the industries and academia of Europe. Industry 4.0 was announced as one of the key innovation in the high technological standards by German federal government such that many international conferences, funded research projects, and technical meets are arranged to identify the significant research challenges in this area. GE (General Electric) coined Industry 4.0 as industrial internet and developed their prototype with the same vision of Industry 4.0 that made a remarkable industrial revolution [3].

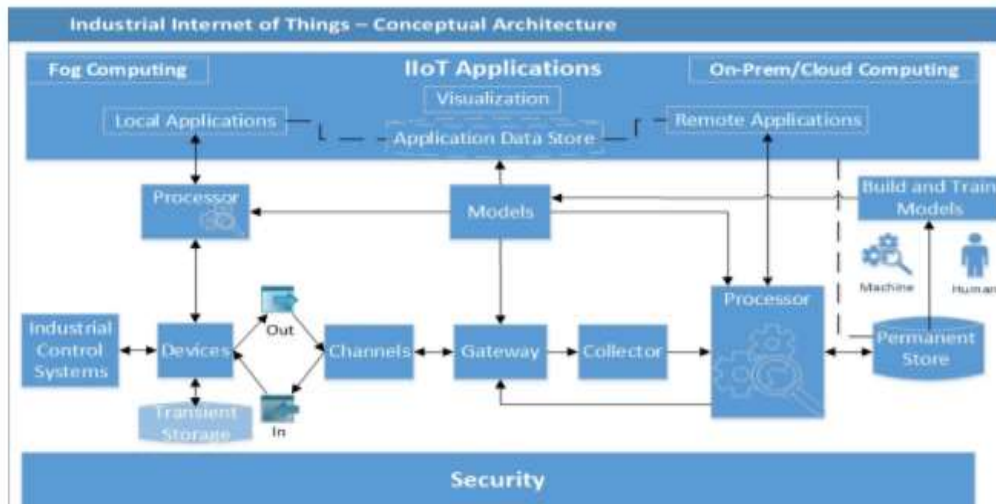


Figure 2. IIoT Conceptual architecture

### 1.2 IIoT Conceptual Architecture

IIoT (Industrial IoT) is considered as an extension of the base concept of IoT at the level of industrial scale. Multiple industries will have their own limitations in the context of devices, commercial appliances with limited interfaces in which few of them may consider the network bandwidth as a major concern. In the view of their challenges that vary from one industry to other industry, there is a lack of one single solution to these problems at the industrial scale. With their uniqueness in the level of manufacturing, every industry needs custom tailoring. Figure 2. demonstrates the conceptual architecture adopted for IIoT and its components are illustrated as follows.

#### a) Computing Environments

Computing environment at the industry level vary dynamically based on the application and industry viz. healthcare, retail, manufacturing oil and gas that may include general utilities. The most prominent and current technologies used for utility computing are enlisted as follows

- Cloud Computing – Enables global connectivity among the industries with applied analytics of data
- Fog Computing – Enable computational analytics at device level
- On-Prem Computing – Enables data crunching in HPC (high-performance computing centers)
- Hybrid Computing – It provides a hybrid technique generated with the combination of cloud, fog and, on-prem computing that optimizes the operation of industries based on their needs.

#### b) Security

Security is considered as the most central aspect of any IIoT application. It is to be enabled throughout the entire pipeline from the consumption level at

the source end as well as the destination end. Security is viewed as a critical aspect at all the levels of digital enterprises that are enabled with data processing at the small, medium and large scale industries in the world of IIoT. The IIoT enables security at the level of authentication and authorization, encryption and decryption, network level security with firewalls, masking techniques and so forth.

#### c) Models

Initially, the architectural features of IIoT enable data models and analytical models. In terms of data model, it specifically defines the structure in which the data is to be organized and in contrast the analytical models deal with custom tailored industry-specific use case designs. In general, the models provide perspective to the data at the industry level enabling the data gathered from various sources are analyzed for prediction. The analytical models are designed in such a way that they are enabled with machine learning algorithms that conducts prediction analysis with historical data. The prominent analytical models include regression and clustering techniques, statistical data analytics, Petri nets of data, and semantic analysis of the data. JSON and XML are used for providing ER (entity-relationship) mapping

#### d) Permanent Data Store and Application Data Store

A storage system in IIoT is defined as the permanent and application storage levels for long-term data storage and analysis. These data stores play a vital role in organizing the device data as well as the data from various sources to generate the data to be processed in data stores based on advanced analytical model building. These may include cloud data repositories, massively parallel processing data stores, HPC data lakes for seamless connectivity of data at the level of IT business. RDMS, Mongo DB may be considered as examples in this aspect of data storage

**e) Processors**

Processors acts as the core component at any IIoT solution. The primary function of the processor is to enable the specific business needs with automation. It includes ingestion pipelines, advanced analytical tools, machine learning algorithms, data transformers, steam processors, signal detection, analytical models with event processing of the data.

**f) Collectors**

These act as the data gatherers that generate the aggregate data from the gateways that are modelled with standard protocols and they are custom designed based on the industrial needs that vary from industry to , industry. They may include OPC data, application adapters management systems and brokers.

**g) Gateways**

Gateways enable communication over multiple networks as well as protocols that facilitate data exchange in between distributed IIoT components. Intelligent signal routers and protocol translators are the best examples of gateways

**h) Channels**

Channels act as a media for data communication between outside world and devices that includes routers network protocols, communication protocols etc.

**i) Applications**

Deployment of applications on shore and off-shore is made based on the business needs, these applications enable the insights on realtime and historical data that enable staff to compose efficient decisions, these include web based applications, alerts, notifications, dashboards, visualizations, devices etc.

**j) Local Processors**

These are considered as low latency data processor units that are geographically distributed in the industrial sites. These processors facilitate fast processing of the data that includes data filters, rule based engines, data processors, algorithms, routers and signal detectors etc.

The act of feeding data in remote applications that are distributed deployed in the industrial site is done based on the local processor and are integrated with devices for data processing

**k) Transient Store**

It is considered as a temporary and optional data store integrated to a device and its prime rationale

is to ensure system failures and data outages. These data stores include flash drives, discs etc.

**l) Devices**

In general, devices are considered as a specific components of industries that acts as an interface with analog or digital systems as well as expose data to the digital outside world. Devices enables M2M(machine to machine) and H2M(human to machine) communication with Industrial Control Systems to interchange information among the components in IIoT landscape. These include translators, interpreters, sensors, event generators, loggers etc.

**m) Industrial Control Systems (ICS)**

ICS enables the events transversely over the industrial systems to the field staff to control industrial operations and deployed in the industrial scale as programmable logical controllers, supervisory control and data acquisition systems, distributed control systems and industry specific systems.

**2. Survey Goals**

The present research aims to investigate and collect diverse credible studies addressing various mechanisms to perform data analytics on the data gathered by IIoT. In precise the purpose is to conduct systematic and comprehensive study of various techniques to enable analytics on high rated sensor data. The primary goal of the survey is

- To provide systematic study of existing techniques and mechanisms that enable analytics on IIoT data
- To analyse the anatomy of the existing studies and their experimental computations
- To assess and enumerate the data sets as well as the benchmarks identified in the studies
- To identify the scope of further research in to apply analytics in IIoT
- To present the overview of the key areas and essential insights from the studies that enables the enhancement of mechanisms to perform analytics on IIoT

**3. Research Methodology**

- Formulate questionnaire to extract the silent features of the mechanisms from the reliable research articles

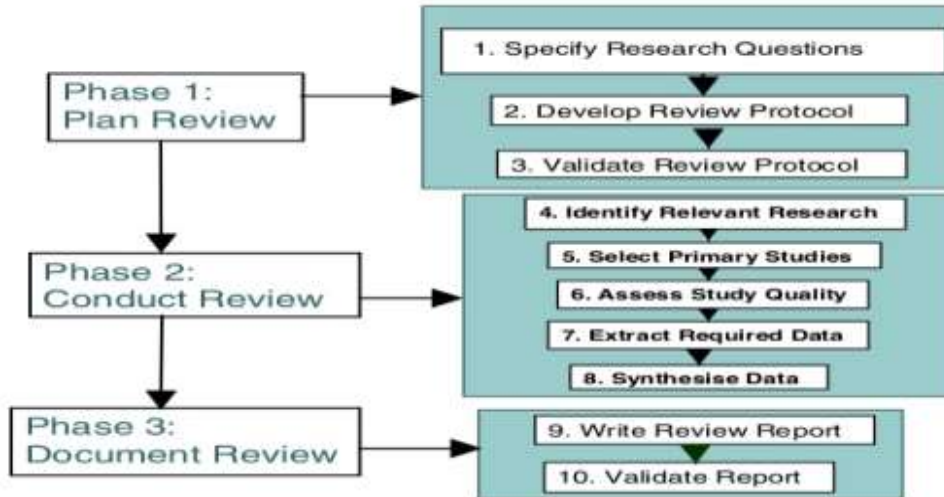


Figure 3. Research Methodology

- A review protocol is developed to enable the search strategy
- Identifying the inclusion and exclusion criteria.
- Detailed review on the selected articles
- Outline the research gaps and identify the frontier research areas.

**A. Questionnaire formulation:**

Integration of IoT with cloud computing and big data analytics enables vast range of research frontiers such that identification of the technologies within the due course of integration. Various mechanisms involved in the prediction analysis performed on the huge data. The following research questionnaire is developed to restrict the search strategy to the focused selection of the articles from reliable databases.

- RQ.1** How seamless integration of the information processing is enabled in between Industrial data and Cloud?
- RQ.2** What are the various frameworks that address the problem of Integration of IIoT and Cloud for smarter communities?
- RQ.3** Design and development of IIoT frameworks in the context of various industries
- RQ.4** Demonstrate the adoption Big data analytic tools on high rated sensor data.
- RQ.5** Discuss various Statistical techniques adopted for the improvement of the prediction

**B. Search Strategy**

The main objective of the preliminary search process is to identify reputed and reliable, journal and conference articles on IoT, IIoT and Data analytics in IIoT. The search process is conducted on scientific databases via electronic searching. Initially the following keywords and electronic

databases are identified for searching process is enumerated in table 1.

Table 1. Search Strategy

Internet of Things(Io T)	"ORed" and "ANDe d"	Google scholar ( <a href="https://scholar.google.co/">https://scholar.google.co/</a> )
Industrial Internet of Things (IIoT)		Springer ( <a href="http://link.springer.com/">http://link.springer.com/</a> )
Industry internet 4.0		IEEE explorer ( <a href="http://ieeexplore.ieee.org/">http://ieeexplore.ieee.org/</a> )
Industry Automation		Science Direct ( <a href="http://www.sciencedirect.com/">http://www.sciencedirect.com/</a> )

**4. Inclusion and Exclusion Criteria of articles**

Based on the studies retrieved by the manual and automated search strategy, each article was examined by the author and co author such that divergences were accorded among the authors as the following studies were taken into consideration if they met atleast one among the inclusion criteria enumerated below:

- Studies that explained proposed mechanism/technique/ framework clearly and obviously in an evidence based evaluation scenario
- Studies related to design and development of data analytic tools and techniques for IoT data.
- Studies that include frameworks to integrate contemporary technologies like BigData, Cloud, IoT
- Studies that provide qualitative assessment metrics such that adoption of sensors to gather diverse data, tools for evaluating data and benchmarks to perform comparative analysis.



The subsequent articles were excluded if they met atleast one of the exclusion criteria:

- Articles that are not focused on Data analytics in IoT and IIoT
- Articles that present only comprehensive analysis without any practical orientation
- Articles demonstrating frameworks without evidence based evaluation.
- Duplicated articles i.e., different sources publishing the similar studies
- Articles that are not written in the English language.

**5. Review of Research studies**

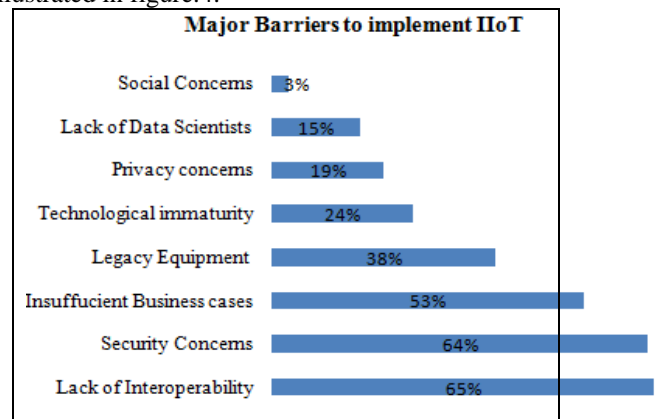
**A. Industrial Internet Review**

Shu-Ting Deng and Cong Xie [8] describes the characteristics of Industrial Internet of Things and data processing technology, which assists the better management of voluminous Industrial data. To process such huge amount of industrial data, we indeed need to understand the characteristics of industrial data. This data possesses some characteristics such as massive polymorphism, generally massive data can be in different formats which was collected by sensors node of multiple sensor networks, dynamic heterogeneity, which means data is in different quantity and formats when the data is collected by different sensors at different times and relevance, the data generated in the production and the attributes of data have definite interconnection between them etc. These features will help the data processing technology to solve the problem of data storage of industrial data. The physical objects like sensor nodes in the production phase will collect the data and the data mining and analysis will process and retrieve the valued data from raw industrial data. To store such huge amount of data, a database, data warehouse, cloud storage and network storage is needed. Wireless sensor networks have wireless database storage, but it won't be advantageous to store such massive data because of limited storage capacity, limited computing, not able to send the data back to central processing etc. so, the industrial internet of things data storage mechanism will have two forms i.e., storage and traction mechanism. Storage mechanism can store the data globally so, the data can be accessed with less query time. The traction algorithm will help to find the exact data from the data which was stored globally.

M Shamim Hossain and Ghulam Muhammad [10] proposed health based Health IIoT-enabled monitoring framework records the patients' health data and ECG through the wired and wireless networks. The recorded data such high frequency ECG signals will be sent to cloud to store through low pass filters that leads to noise so, a signal reconstruction or enhancement is to be needed. ECG watermarking has to be done on ECG to provide security and to protect from third parties

such as healthcare professionals. The cloud based storage of health care data can be extracted frequently to know the condition of patients. In spite of its usage, this method will face significant challenges on securing the data of real time patients and health professionals.

In industrial applications, the various production machines will generate enormous data, the main problem here is data is in distributed fashion. To predict the future enhancements from the available data is hard. So, Simon Duque Antón, Daniel Fraunholz, Janis Zemitis, Frederic Pohl, and Hans Dieter Schotten [9] proposes an approach to gather all the relevant data from the heterogeneous networks. This method will collect and interconnect the data from different sources which consumes more computing time. It is highly scalable i.e., it allows to extend the sensor networks whenever it needs. Due to aggregation concept, the approach can handle rapidly generated voluminous amount of industrial data. The proposed model's main attention is on Industrial internet of things (IIoT), so we can apply it on any use case which has more than one machine due to its aggregation characteristic. Here, we mainly discuss about two use cases. One is remote maintenance application, the proposed model will provide a solution of handling the rapid generation of data from different sources and their interconnection. In the second use case, the network traffic and the log information of networks can be handled. Author in his article, the researchers can label the problems of rapid generation of data and also they can build or develop algorithms for effective detection of misbehaviors. The major concerns to adopt IIoT is illustrated in figure.4.



**Figure 4. Barriers to implement IIoT**

Guangjie Han, Wenhui Que, GangyongJia, Wenbo Zhang proposes a resource-utilization-aware energy efficient server consolidation algorithm(RUAEE) allows multiple resource utilization by minimizing the virtual machine migrations to reduce energy consumption. Now-a-days IIoT generates a lot of data from production phase, to store or process such massive data a cloud data centers have to be

extended with the rapid growth of computation and storage. With this, lot of problems occurred with greenhouse gas emission. So, RUAEE provides feasible and effective solution for reducing the resource utilization. Researchers can enhance the proposed algorithm by analyzing different types of virtual machines and also by reducing their migrations.

AthulJayaram [11] come up with a Lean Six Sigma Approach for Global Supply Chain Management using Industry 4.0 and IIoT. Global Supply chain management will manage the supply chain firms which consists of suppliers, vendors etc. which are located globally. Quality control strategies, for example lean and Six Sigma are proposed to enhance productivity and distribution by global supply chain management.

Lean Process will detect and eliminates the unused processes which are not required to perform the operations of a business; it reduces the cost and also the wastage of resources. This can make the customers to develop a desired product. Six sigma follows Define, Measure, Analyze, Improve and Control (DMAIC) a standard approach and it is a method which contains fault tolerance technique that detects and removes the faults in every step. These two approaches are more advantageous in increasing the revenue of a business, efficiency and cost reduction. The proposed algorithm works effectively and efficiently when you integrate the Industry 4.0, which are connected to the internet to share the information automatically between the supply chain and the logistics with IIoT to get highly optimized resources as well as fault tolerance.

Benjamin Weber [13], Jens Butschan and Sven Heidenreich describes how the heterogeneous competences will affect the usage capacity of IIoT in the organization. To increase effective and efficient high production, a new capabilities or components are required that being so will increase the usage of IIoT in the firm to handle the obstacles of digital transformation. The competence will influence the IIoT digital data transformational process. Now-a-days digitalization of data becomes prominent, to identify the data easily which was stored in globally and also for the long-term success of a firm or organization. The suitable competence model has to be noticed, a concrete competence has been proposed to access the human capital in IIoT and it have 3 parts. Initially we have to analyze the environment of innovations, how it was influencing by digitalization. Secondly, select appropriate competence model for digital transformation and finally it analysis the results of competence.

The authors, in this article explain the effects of different competences such as social, cognitive, social and processual competences on the usage of IIoT in the organization. Among all, the person,

who has high cognitive competence can easily adopt new technologies in any satiation without problem. The cognitive competence helps the usage of IIoT for digital transformation of IIoT data.

Ralph Rio [12] proposed Industrial IoT-enabled remote Monitoring technique to improve the performance of Original Equipment Manufacturer (OEM) service. If the machineries go down and having any internal problem, then a two-pass repair solution has been given to solve the problem. In the first pass, a person can inspect the machinery to detect the problem. While in another pass, actual process i.e., repair can be performed. This technique took more time to solve. In order to reduce the down time and mean time to repair (MTTR), an IIoT enable remote monitoring has been developed. By this, OEM will check the performance of its machineries frequently, whenever the problem occurs to the machinery then it automatically sends the request, they analyze the problem through web and fix it. This technique possess reliability and increases the mean time to repair. A glimpse of reliable works in this area enables us to identify the researchers interested in this promising area along with their research studies that include novel ideas, algorithms, frameworks and approaches with the expanding scope of problems in this area.

## **6. Conclusion and Future Work**

There is a drastic increase in the data production over the past few decades with the large number of sensor and smart devices acquired from distributed data sources. Industrial Internet of Things endow with a convenient and a viable service to develop applications in the context of building intelligent industry. Massive data acquired from IIoT deployment, results in the wastage of resources during the data transmission process. Data obtained from IIoT deployment will have features like dynamic heterogeneity and polymorphism. This article address the problem of adopting IIoT technology and presents a comprehensive review of various studies that address the stature of IIoT deployment across various application industries. Moreover, this article provides the insights about the potential barriers of implementing IIoT for intelligent industry. There is need for deriving more accurate analytical mechanisms at the edge of the network and it could be a prominent area for further research.

## **7. References**

- [1] J. Lee, B. Bagheri and H.-A. Kao, "A cyber-physical systems architecture for industry 4.0-based manufacturing systems," *Manufacturing Letters*, vol.v3, pp. 18-23, 2015
- [2] H. Hirsch-Kreinsen, "Digitization of industrial work: Development paths and prospects,"

Journal for Labour Market Research, vol. 49, no. 1, pp. 1–14, 2016.

[3] D. M. Truxillo, D. M. Cadiz, and J. R. Rineer, *The Aging Workforce: Implications for Human Resource Management Research and Practice*, Oxford: Oxford University Press, 2014

[4] D. Gorecky, M. Schmitt, M. Loskyll and D. Zühlke, "Human-machineinteraction in the industry 4.0 era," in 2014 12th IEEE International Conference on Industrial Informatics (INDIN), 2014

[5] J. Lee, H.-A. Kao and S. Yang, "Service innovation and smart analytics for industry 4.0 and big data environment," *Procedia CIRP*, vol. 16, pp. 3-8, 2014

[6] Cugola, G., Margara, A.: *Processing Flows of Information: From Data Stream to Complex Event Processing*. In: *ACM Computer Surveys* (2012)

[7] Jiang, G., Cybenko, G.: *Temporal and Spatial Distributed Event Correlation for Network Security*. In: *Proceedings of the American Control Conference* (2004)

[8] Shu-Ting Deng and Cong Xie, "Research Based on Data Processing Technology of Industrial Internet of Things", *ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2017* F. Chen and Y. Luo (Eds.): *Industrial IoT 2017*, LNICST 202, pp. 53–60, 2017.

[9] Simon Duque Antón, Daniel Fraunholz, Janis Zemitis, Frederic Pohl, and Hans Dieter Schotten, "Highly Scalable and Flexible Model for Effective Aggregation of Context-based Data in Generic IIoT Scenarios", 9th ZEUS Workshop, ZEUS 2017, Lugano, Switzerland, 13-14 February 2017.

[10] M Shamim Hossaina, Ghulam Muhammad, "Cloud-assisted Industrial Internet of Things (IIoT)- enabled framework for health monitoring", *Computer Networks* (2016), doi: 10.1016/j.comnet.2016.01.009

[11] Athul Jayaram, "Lean Six Sigma Approach for Global Supply Chain Management using Industry 4.0 and IIoT", 2nd International Conference on Contemporary Computing and Informatics (ic3i),2016.

[12] Ralph Rio," Industrial IoT-Enabled Remote Monitoring Improves OEM Service Performance", ARC ,3 Allied Drive , Dedham, MA 02026 USA 2016 , 781-471-1000.

[13] Benjamin Weber, "Tackling hurdles to digital transformation – The role of competencies for successful IIoT implementation", 2017 IEEE Technology & Engineering Management Conference (TEMSCON).