DOI: https://doi.org/10.52756/ijerr.2023.v30.019 *Int. J. Exp. Res. Rev.*, Vol. 30: 219-227 (2023)

Review Article | Peer Reviewed | Open Access

International Journal of Experimental Research and Review (IJERR)

ISSN: 2455-4855 (Online)
www.iaph.in

Examining a generic streaming architecture for smart manufacturing's Big data processing in Anomaly detection: A review and a proposal

Milton Samadder[1*], Anup Kumar Barman[2] and Alok Kumar Roy[3]

[1]Department of Computer Science & Engineering, Central Institute of Technology, Kokrajhar, Assam-783370, India; [2]Department of Computer Science & Engineering Central Institute of Technology, Kokrajhar, Assam-783370, India; [3]Department of Computer Science & Engineering, Bankura Unnayani Institute of Engineering, Subhankar Nagar, Pohabagan, Bankura, West Bengal, India

E-mail/Orcid Id:

MS, ph19cse1908@cit.ac.in, https://orcid.org/0009-0008-5261-9047; *AKB,* ak.barman@cit.ac.in; *AKR,* aloke.roykgp@gmail.com

Article History:

Received: 30th Jan., 2023
Accepted: 11th Apr., 2023
Published: 30th Apr., 2023

Keywords:
Big data processing, data streaming, Industry 4.0, message streaming, smart manufacturing, stateful computation

Abstract: The smart manufacturing industry has witnessed a rapid increase in data generation due to the integration of sensors, IoT devices, and other advanced technologies. With this huge amount of data, the need for efficient data processing methods becomes critical for identifying anomalies in real-time. With the rise of Industry 4.0 practices, digitally enabled manufacturing units are shifting their focus towards Smart Manufacturing paradigm for better productivity, throughput and increased business volume. Traditionally digital manufacturing units have considered different AI approaches like Neural Network, Statistical Methods, Deep Learning etc. to detect and predict anomalies in their production lines. But with the Smart Manufacturing ecosystem, a manufacturing unit must integrate manufacturing intelligence in real-time across entire production lines through sensor data of IOT devices. Hence the traditional anomaly detection systems fall short to respond well, under the changed scenario, where large volumes of unstructured and varied types of data are being generated at high velocity, to be processed at (soft) real time. The article reviews the current state-of-the-art in big data processing for anomaly detection in smart manufacturing. The review covers various aspects such as data collection, data processing, anomaly detection, and real-time monitoring. The current paper also proposes a novel stateful data streaming computational model for big data processing in smart manufacturing units which conceptually lays the foundation on top of which any discrete anomaly detection engine would be able to work. The proposed architecture has several benefits, including its ability to handle the large volume, velocity, and variety of data generated in smart manufacturing. The architecture can be applied to various smart manufacturing applications, including predictive maintenance, quality control, and supply chain optimization. It is expected that this proposed architecture will pave the way for the development of more efficient and effective smart manufacturing systems in the future.

Introduction

Anomaly detection is essential methods which are utilizing to recognize fraud, suspicious activity, network intrusion, and other unexpected occurrences that may be of considerable importance but are hard to spot (Goldstein and Uchida, 2016). The importance of anomaly detection lies in the process' ability to transform data into crucial information that can be used for action and to reveal insightful information in a variety of application domains (Chandola et al., 2009). Historically, since pre-pre-digital era, any anomaly in the production line in manufacturing facilities has attracted a huge production cost resulting in poor growth of overall business. With IT adoption, the manufacturing sector started introducing parameter control systems in the production lines but in a discrete manner. With the advancement of different predictive analysis models and statistical tools, in the next phase of IT industrialization,

*Corresponding Author: ph19cse1908@cit.ac.in

individual production line's anomalies were detected and monitored.

But, with the rise of IIOT (Industrial IOT) and other pervasive computing technologies, machines within composite production lines are now largely driven by sensors and actuators. The sensor data reveal the efficiency and preciseness of the manufacturing process and set a benchmark for parametric data accountable for optimum production cost. As the individual machines in the production lines are connected, it forms a formidable and composite ecosystem where parametric data could be accessed and processed in real time to detect (and predict) any drift from the normal and optimum sensor data range.

AI based and statistical anomaly detection systems have been in the market for the last few years, but they are considered to be working on data at rest. However, as governed by Industry 4.0 standards, IIOT sensors produce a mix of unbounded and bounded streams of big measurement data that demand specialized processing methodology which needs to be fast, scalable, data driven and precise. Traditional intelligent and statistical anomaly detection systems in manufacturing sectors are classically known to be working with discrete data levels having considerably larger response time. These algorithmic processes are task or event driven and require an intermediate set up time to achieve scalability. Hence there exists a research gap and the urgency of designing a fast, data driven, stream processing architecture for anomaly detection has motivated the current research to be carried out with utmost care.

In the production pipeline of manufacturing facilities, Anomalies are described as drift of functional parameters from their usual or typical values. Early Detection in anomalies can predict and rationalise any future malfunction saving time and finance, which are true crucial factors in any manufacturing units.

In order to offer a fundamental understanding of the numerous ways for anomaly detection, Agrawal and Agrawal (2015) explore various anomaly detection strategies. Zheng et al. (2022) presented a novel model to detect anomalies where a deep neural network to extract low-dimensional features from the background space, and subsequently separates these features using a hypersphere. This enables the model to distinguish between normal and anomalous classes. Patcha and Park (2007) conducted a very thorough study on anomaly detection systems and hybrid intrusion detection systems. They were able to identify areas where these systems can improve and the difficulties they may face. As shown in the report (Kamat and Sugandhi, 2020) the evolution of maintenance of the manufacturing industry has come across a long way. It started with the style of "Reactive Maintenance" in pre-digital era, and after crossing certain phases like "Preventive Maintenance" in early IT adoption days, and "Rule-based Predictive Maintenance" in structured data driven analytics models, currently, the manufacturing industries are passing through the phase of "Predictive Maintenance" where anomaly detections play a critical role in maintaining equilibrium within the production pipeline. It has also formalized the definition of Anomaly Detection and has described certain challenges towards Anomaly Detection. The representation of a dataset with multiple dimensions presents an arduous task and a myriad of scientific investigations have been aimed at tackling the mounting complexity of dimensionality (Tatu et al., 2012). There are many surveys that have been done by different researchers to address the challenges and issues related to anomaly detection and dealing with complex data (Jindal and Liu, 2007; Pathasarathy, 2007; Tamboli et al., 2016; Spirin et al., 2012). Since the last few years, researchers are working on different statistical and AI based models to implement a robust and reliable anomaly detection methodology for a comprehensive predictive maintenance system.

Liu et al. (2018) has proposed a Structured Neural Network Model (under supervised learning paradigm) to detect anomalies. It focuses on the struterization of Neural Networks on the basis of Event Ordering Relationships. Tang et al. (2018) has conceptualized a unique model of anomaly detection where the anomaly patterns have been described through Convolutional Neural Networks (CNN). A semi-supervised deep learning approach was undertaken for early detection and classifications of anomaly. Pittino et al. (2020) introduced a Statistical Learning Method for automatic detection of anomalies on In-Production machines with realistic data of wet wafer processing machines of Semiconductor manufacturing units. They have emphasized on designing a machine learning algorithm which works pretty well for Control Charting mechanism with varied classification schemes. Lindemann et al. (2019) has reported the comparative analysis of two data driven Self Learning mechanisms for Automatic Anomaly Detection in discrete manufacturing processes. It has used real data set from the metal formation process and demonstrated K-means based approach with sliding windows and LSTM based approach using autoencoder structure for anomaly detection. The report concludes that the second approach has better sensitivity to predict machine faults at early stages. Li et al. (2020) have used

DOI: https://doi.org/10.52756/ijerr.2023.v30.019

data augmentation techniques for anomaly detection. Their studies have proved that individual augmentation works superior with high accuracy over the combination of augmentation. DeLaus (2019) has undergone a deep study on the field of machine learning approaches and in order to detect anomalies in semiconductor manufacturing units, he has introduced a hybrid model of cluster analysis & time series forecasting. Zope et al. (2019) studied seven AI techniques and demonstrated them in semi supervised mode to introduce a new Automatic Anomaly Detection model, capable of anomaly diagnosis as well. Żabiński et al. (2019) proposed a platform that monitors intelligent conditions for detection of ideal novelty within a production line parametric data and utilizes those conditions to detect the anomalies later on with training data sets. The reports mentioned above, have studied the problem of automatic detection of anomalies in the manufacturing sector from different contexts and perspectives. However, it is noteworthy that the efforts were all towards data at rest having discrete values at different temporal micro batch timings. To the best of our belief, there is serious insufficiency of computation platforms for anomaly detection under real time data streaming mode which is now obvious with the onset of Big Data and Industry 4.0 standards. Big data may produce real-time answers to problems in all spheres of life (Phuyal et al., 2020). An online real-time data anomaly detection framework and a cutting-edge anomaly detection method were introduced by Corizzo et al. (2019) to solve real time problems.

In this context, next section describes the impact of Big Data and Industry 4.0 standards in the manufacturing sectors that has brought huge changes in the nature of the generation and distribution of parametric data. These changes certainly motivate the current study for proposing a generic data streaming platform automatic anomaly detection.

Manufacturing industry, as a whole, is experiencing a positive transformation from raw parametric values of digital data of machines, to data centric manufacturing intelligence that emerges out of real time streams of continuous data. This transformation is due to the insurge of IIOT (Industrial IOT), CPS (Cyber Physical Systems) and similar upcoming technologies and their embedment into the physical process of the manufacturing facilities. These intelligently embedded processes generate data from all across the facility and the data generated are characterized by considerable volume, a great generation speed and complex, as well as varied unstructured types. All these notions confirm prominent Big Data technologies to be used to analyse those data. The utilization of bigdata data analytics and stream processing technologies constitutes an indispensable prerequisite for the effective implementation of prognostic maintenance solutions (Ferreiro et al., 2016; Wang, 2016).

Industry 4.0 standards represent the fourth industrial revolution which optimizes the digitization process that took place during adoption of industry 3.0 standards. The optimization is considered to be in the form of integrated intelligence derived from a huge real time stream of data generated by the sensors and other devices introduced in the production line during matured time lines of industry 3.0 standards. Hence there exists a prominent paradigm shift of mode of data processing that happens during adoption of industry 4.0 standards which enforces a real time big data ecosystem within the manufacturing facilities.

A digitized manufacturing process line equipped with state-of-the-art IT infrastructure lets the entire facility to produce the stream of real time data that is not possible to be consumed by traditional data processing frameworks for effective and meaningful decision-making processes. As a result, Real Time Big Data processing frameworks come into picture to process these new kinds of data sets for extracting meaningful information out of it, thereby empowering the manufacturing unit to transform to a Smart manufacturing unit.

In the recent past, O'Donovan et al. (2015a) reported a systematic mapping study of impact of Big Data in the manufacturing industry. A higher abstraction level conceptual framework, capable of handling Big Manufacturing Data in cloud infrastructure, has been presented by Gökalp et al. (2016). While Yan et al. (2017) has discussed the challenges and issues related with implementing Big and Multi Sourced Heterogeneous Spatial Data in Industry 4.0 environment with respect to predicting maintenance jobs (Gölzer et al., 2015; Latinovic et al., 2019; O'Donovan et al., 2015b), on the other side, have presented some use cases of applications of Big Data in Industry 4.0 ecosystem with elaborate discussions on generic requirements of data processing framework and suitability of Big Data for innovative solutions in Industry 4.0 state of the art. A more realistic study has been carried out by Rivetti et al. (2017), which propose a distributed platform to detect anomalies in the manufacturing sector as an application of Apache Flink, a market level analytic standard for Real Time Big Data computation.

All the efforts, as mentioned in the previous paragraph, have established the fact that Big Data technologies have a natural promise to work better for smart manufacturing processes, as the data there, are real

DOI: https://doi.org/10.52756/ijerr.2023.v30.019

time, distributed and make a continuous stream as opposed to being centrally located, traditional and discrete.

On the basis of on-going discussion, the following section takes a closer look of the problem domain and formalizes specific requirements for designing a Big Data processing platform for Automatic Detection of Anomalies in the manufacturing sector.

Materials and Methods

Requirement Analysis for the Proposed Architecture

Any Industry 4.0 manufacturing set up is connected with IOT sensors for sensing the finest parametric values of different electro mechanical machines and with preliminary requirement ensures the capability of a Supervised Learning environment, where the current parametric sensor data would be examined and anomaly detection would be carried out on the basis of historic data sets.

Hence, the requirement analysis of generic streaming data processing architecture for anomaly detection in manufacturing pipeline within an Industry 4.0 setup could be formalized as follows-

A. The proposed architecture must have capabilities for processing unbounded data streams

B. It should have definite protocols for running specific algorithms for bounded data stream with dynamic data structures

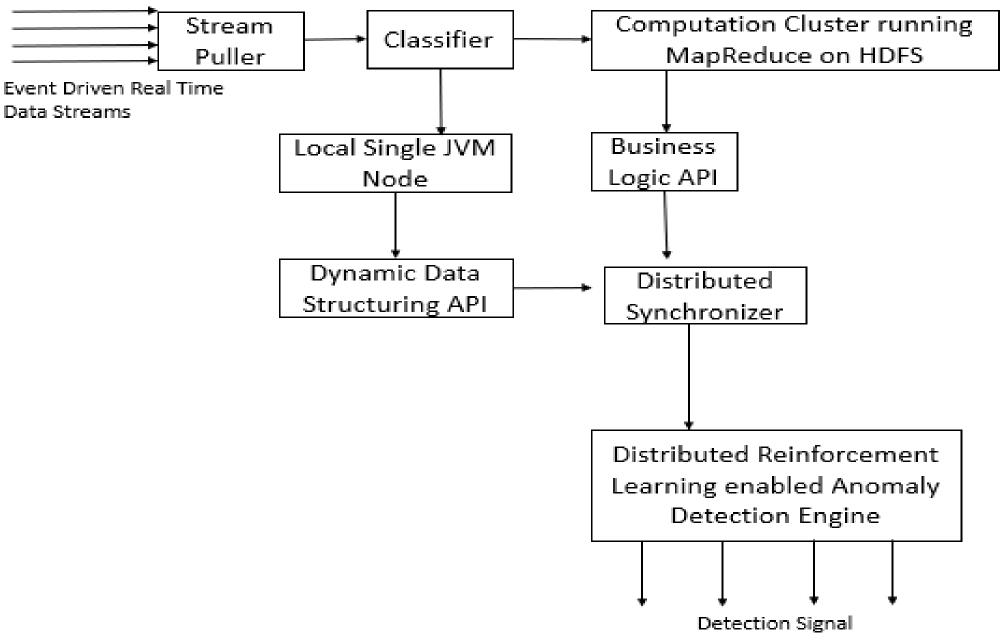

Figure 1. Functional Data Model of the Proposed Architecture

historic fault reports with current machine status. While these sensors give rise to unbounded data streams, the historic log reports generate a bounded data stream. Unbounded data streams are required to be processed with minimum latency as soon as they are ingested and the operation results are to be modified every time the processing is done over the specified section of unbounded data streams. However, the processing of bounded data streams will require to be ingested fully before any data processing computation takes place. Hence bounded data streaming computation can be carried out with micro batch processing mode.

While considering a generic robust streaming architecture for automatic anomaly detection in Industry 4.0 environment, both the types of data streams (unbounded and bounded) must be processed within the same architecture so that processing of bounded data streams can provide support to the processing of unbounded data streams. Identification of this

C. Unbounded data streams must be processed with low latency and at large scale

D. Bounded data streams must be processed in memory for most of the case, unless it exceeds the predefined memory size.

E. Proposed architecture must exhibit a prominent fault tolerance principle

Proposal of New Architecture

On the background of the previous section which discusses the basic requirements of a data streaming architecture for anomaly detection at Industry 4.0 environment, the current section proposes a new data streaming architecture which is inspired from Apache Flink (https://flink.apache.org/flink-architecture.html) along with certain modifications and introduction of new functional elements. It is noteworthy that the proposed platform does not describe any new Anomaly Detection Algorithm, rather it presents a generic architecture for running any Anomaly Detection algorithm to work efficiently over Industry 4.0 standards.

Current section further discusses the nature of the Anomaly Detection Algorithms that are best suited for the proposed generic architecture.

Functional Data Model of the Proposed Architecture

Figure 1 below describes the data model of the proposed architecture where the components are logically connected to reinforce computations on the unbounded and bounded data streams on the same physical platform.

Functional element Stream Puller pulls the unbounded data stream from any distributed Message Queuing systems like Kafka or RabbitMQ. It ensures the pulling of bounded data streams from any other NOSQL databases of the Hadoop family like HBase.

Classifier is a Map Reduce agent which segregates the data streams on the basis of their data signature and routes unbounded streams to distributed clusters on HDFS, whereas the bounded data streams are routed to Single JVM node.

Distributed Hadoop Cluster running on HDFS runs business logic MapReduce APIs for preparation of conditional sensor data. This API takes care of how the unbounded data streams are partitioned and how the parametric sensor data are to be consumed based upon the functional weightage as per detection algorithm. These weightage flows back from the detection algorithm section to the Master node of the cluster. However, Single JVM node runs an agent program to structure bounded data and runs a program to sort and cluster the same in memory. If the memory runs out, it can use optional disk space for storing the result.

Processed unbounded and bounded data are synchronized with any distributed scheduler or synchronizer like Apache Zookeeper and ultimately fed to the Anomaly Detection Engine which is again a Hadoop cluster running any Supervised Learning algorithm on Map Reduce framework to generate the detection signal for the outside world. The said engine feeds the learning experience back to the unbounded data stream processing cluster so that it can partition and prepare sensor parametric data with increasing efficiency. On the background of the functional data model of the proposed framework, the following subsection describes its logical data model.

Logical Data Model of Proposed Architecture

The logical data model of the proposed Architecture is based on a parallel and distributed computing paradigm. The incoming unbounded streams are split into stream partitions and these partitions are placed to some programs which run in parallel.

These programs represent unit computations of the business logic of handling sensor parametric data or the distributed Supervised Learning engine. On a directional data flow graph, these programs act as nodes and streams act as edges. The partitioned data streams transport data between one program to the other in one to one or one to many models.

As shown in Figure 2 above, the unit programs for a

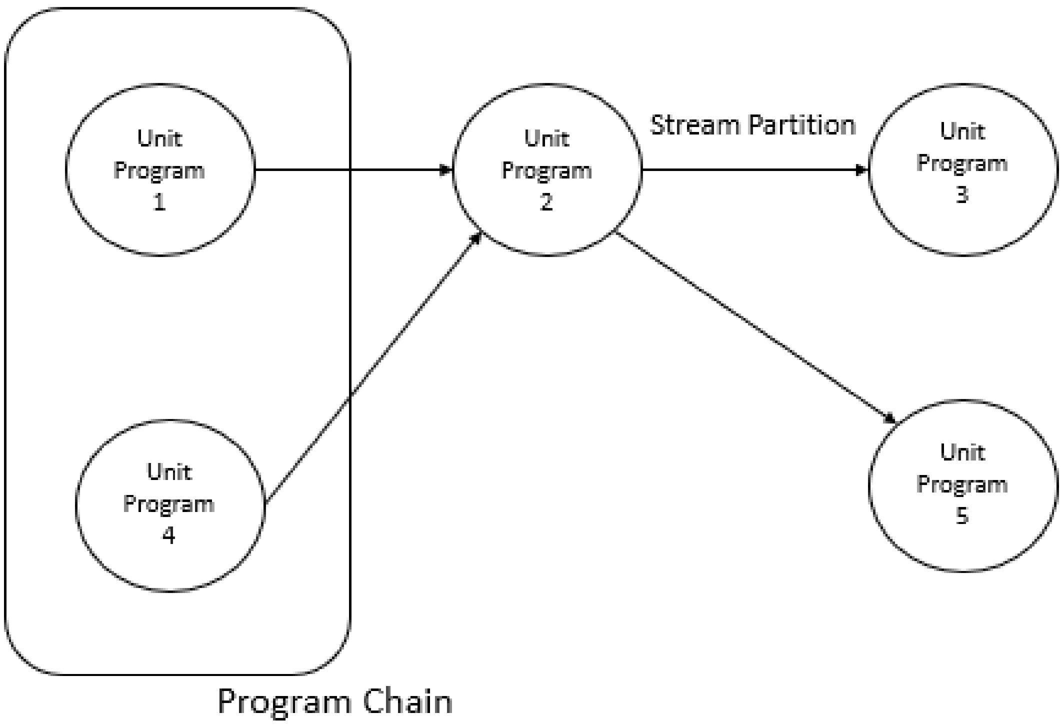

Figure 2. Logical Data Model of the Proposed Architecture

DOI: https://doi.org/10.52756/ijerr.2023.v30.019

condensed business logic or reinforced learning path are clubbed together to form a program chain. These program chains are run in slave nodes of the computation clusters. Moreover, the unit programs within the program chain run in parallel in different threads of the same slave node. In the current context, following subsection describes the physical data model of the proposed architecture-

Physical Data Model of Proposed Architecture

Physical data Model represents how the logical elements like data streams, unit programs and program chains are hosted into the physical infrastructure.

As shown in Figure 3, the process of the runtime of the proposed architecture is divided into two types. Master Process runs within Master Node of the distributed cluster and takes care of the distributed execution model of the entire framework. The Master Process schedules tasks for the slave machines and coordinates some distributed functions like recovery on failure, check point management etc. Task scheduling is done through a supervisor daemon process.

both the Master Process and Slave Processes are initialized within the same file system.

Dynamic Data Structuring and Supervised Learning in the Proposed Architecture

The proposed architecture computes unbounded and bounded data streams. While unbounded data streams are represented by parametric values of the attached sensors of the manufacturing pipeline, bounded data streams, on the other hand, are represented by labelled historical data stating usual and anomaly state of the machine.

The proposed framework supports an in-memory dynamic data structuring provision for this bounded data stream. The length of bounded data streams is supposed to vary from one machinery infrastructure to another and hence the architecture has to prepare an in-memory dataset through API calling. If the size of the bounded dataset grows larger than the fixed in-memory size, it can be stored on disk. Architecture uses Java Generics with proper API hierarchy to utilize proper collection framework and the bounded data set for each machine of

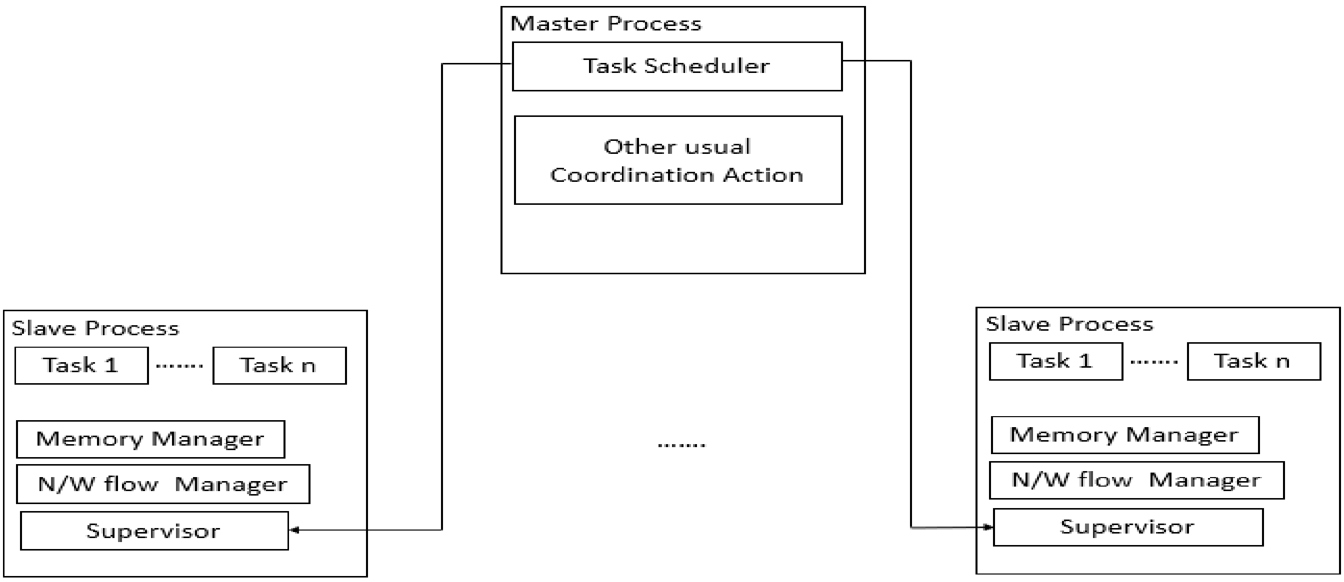

Figure 3. Physical Data Model of the Proposed Architecture

The slave process on the other side runs in slave machines and forks new threads within the slave machine. The slave process represents the program chain and the threads represent the unit programs. The physical data model described above are employed for unbounded data streams flowing through distributed hadoop clusters as in sensor data value computation section, as well as in Supervised Learning section. In this context, unit programs could be abstracted as Map Reduce tasks running on a Hadoop Distributed File System.

However, the physical data model for bounded streams on a single JVM node can be instantiated as the special version of the distributed logical model where

the pipeline are stored into Hash Table for access with low latency. The entire labelled data is processed in memory and fed to the master node of the distributed hadoop cluster running specific supervised learning with distributed features.

Supervised learning in distributed environments can be carried out through any of the popular algorithms like Support-Vector Machines, Linear Regression, Logistic Regression, Naive Bayes, Linear Discriminant Analysis, K-Nearest Neighbor and Neural Networks etc. However, the intrinsic distributed nature of the proposed architecture demands the traditional supervised learning algorithms to be modified to reach high scalability, lower latency and better throughput. Various research efforts

DOI: https://doi.org/10.52756/ijerr.2023.v30.019

(Bajo and Rodríguez, 2017; Navia-Vazquez et al., 2006; Japa and Shi, 2020; Zuo et al., 2021) could be considered as potential candidates to be plugged into the proposed algorithm.

Conclusion and Scope of Further Research

In conclusion, this review and proposal have highlighted the importance of big data processing in anomaly detection for smart manufacturing. Current paper also reports a design proposal for a generic, distributed and streaming architecture for Anomaly Detection in the manufacturing sector. In the background of different existing anomaly detection principles and Big Data solutions in Industry 4.0 ecosystem, proposed architecture works both for bounded and unbounded real time data streams and processes data analytics query in real time. The concept proposed in the current report could further be physically modelled and tested with static and dynamic industrial data from multiple domains. While detecting anomalies within Industry 4.0, the choice of distributed supervised learning algorithm would be the most critical issue. As the current report proposes a generic design of streaming architecture, it does not specify any distributed supervised algorithm, hence experimenting the proposed work with different distributed supervised learning algorithms and their comparative study will certainly produce novel insights of the problem domain and its solutions. Overall, this review and proposal offer valuable insights into the design and implementation of scalable, efficient, and reliable streaming architectures for smart manufacturing's big data processing in anomaly detection.

Acknowledgements

The authors are grateful to Mr. Utpal Madhu, Assistant Professor, Department of Mechanical Engineering, Swami Vivekananda Institute of Science & Technology, Kolkata, India for his valuable support during preparation of the manuscript.

Conflict of interest statement

The author declares that in this work there is none conflict of interest.

References

Agrawal, S., & Agrawal, J. (2015). Survey on anomaly detection using data mining techniques. *Procedia Comput Sci*., *60*, 708–713. https://doi.org/10.1016/j.procs.2015.08.220

Apache flink. https://flink.apache.org/flink-architecture.html

Bajo, J., & Rodríguez, J.M.C. (2017). Neural Networks in Distributed Computing and Artificial Intelligence. *Neurocomputing*, pp. 272. https://doi.org/10.1016/j.neucom.2017.06.022

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: a survey. *ACM Comput. Surv.*, *41*(3), 15. https://doi.org/10.1145/1541880.1541882

Corizzo, R., Ceci, M., & Japkowicz, N. (2019). Anomaly detection and repair for accurate predictions in geo-distributed big data, *Big. Data Res.*, *16*, 18–35. https://doi.org/10.1016/j.bdr.2019.04.001

DeLaus, M.D. (2019). Machine Learning for Automated Anomaly Detection in Semiconductor Manufacturing - Master's thesis at Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Ferreiro, S., Konde, E., Fernández, S., & Prado, A. (2016). Industry 4.0: predictive intelligent maintenance for production equipment. *European conference of the Prognostics and Health Management Society*, pp. 1–8.

Gökalp, M. O., Kayabay, K., Akyol, M. A., Eren, E. P., & Koçyigit, A. (2016). Big Data for Industry 4.0: A Conceptual Framework. 2016 *International Conference on Computational Science and Computational Intelligence* (CSCI). *IEEE*, pp. 431-434. https://doi.org/10.1109/CSCI.2016.0088

Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PLoS ONE*, *11*(4), 0152173. https://doi.org/10.1371/journal.pone.0152173

Gölzer, P., Cato, P., & Amberg, M. (2015). Data Processing Requirements of Industry 4.0 - Use Cases for Big Data Applications. Twenty-Third European Conference on Information Systems (ECIS), Münster, Germany, 2015. *ECIS 2015 Research-in-Progress Papers*, pp. 61. http://aisel.aisnet.org/ecis2015_rip/61

Japa, A., & Shi, Y. (2020). Parallelizing the Bounded K-Nearest Neighbors Algorithm for Distributed Computing Systems". *2020 10th Annual Computing and Communication Workshop and Conference* (CCWC), pp. 0038-0045. https://doi.org/ 10.1109/CCWC47524.2020.9031198

Jindal, N., & Liu, B. (2007). Review spam detection. In: Proceedings of the 16th International Conference on World wide web. *ACM*., pp. 1189–90. https://doi.org/10.1145/1242572.1242759

Kamat, P., & Sugandhi, R. (2020). Anomaly Detection for Predictive Maintenance in Industry 4.0- A survey. *E3S Web of Conferences 2020*. https://doi.org/10.1051/e3sconf/202017002007

Latinovic, T., Preradović, D., Barz, C. R., Vadean, A. P., & Todić, M. (2019). Big Data as the Basis for the innovative Development Strategy of the Industry 4.0. *IOP Conference Series: Materials Science and Engineering*, *477*, 012045. https://doi.org/10.1088/1757-899X/477/1/012045

Li, X., Zhang, W., Ding, Q., & Sun, J. Q. (2020). Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation. *Journal of Intelligent Manufacturing*, *31*, 433–452. https://doi.org/10.1007/s10845-018-1456-1

Lindemann, B., Fesenmayr, F., Jazdi, N., & Weyrich, M. (2019). Anomaly Detection in Discrete Manufacturing Using Self-Learning Approaches. *Benjamin Lindemann et al. Procedia CIRP*, *79*, 313–318. https://doi.org/10.1016/j.procir.2019.02.073

Liu, J., Guo, J., Orlik, P.V., Shibata, M., Nakahara, D., Mii, S., & Takac, M. (2018). Anomaly Detection in Manufacturing Systems Using Structured Neural Networks. *IEEE World Congress on Intelligent Control and Automation 2018*, pp. 1-6. http://dx.doi.org/10.1109/WCICA.2018.8630692

Navia-Vazquez, A., Gutierrez-Gonzalez, D., Parrado-Hernandez, E., & Navarro-Abellan, J.J. (2006). Distributed Support Vector Machines. *August 2006 IEEE Transactions on Neural Networks*, *17*(4), 1091-1097. https://doi.org/10.1109/TNN.2006.875968

O'Donovan, P., Leahy, K., Bruton, K., & Dominic T. J. O. (2015a). Big Data in Manufacturing: A Systematic Mapping Study. *Journal of Big Data*, *2*, 20. https://doi.org/10.1186/s40537-015-0028-x

O'Donovan, P., Leahy, K., Bruton, K., & O'Sullivan D. T. J. (2015b). An Industrial Big Data Pipeline for Data-driven Analytics Maintenance Applications in Large-scale Smart Manufacturing Facilities. *Journal of Big Data*, *2*, 25. https://doi.org/10.1186/s40537-015-0034-z

Parthasarathy, S., Ghoting, A., & Otey, M. E. (2007). A survey of distributed mining of data streams. In: Data streams. Springer, pp. 289–307. https://doi.org/10.1007/978-0-387-47534-9_13

Patcha, A., & Park, J. M. (2007). An overview of anomaly detection techniques: existing solutions and latest technological trends. *Comput. Netw.*, *51*(12), 3448–3470. https://doi.org/10.1016/j.comnet.2007.02.001

Phuyal, S., Bista, D., & Bista, R. (2020). Challenges, opportunities and future directions of smart manufacturing: A state of art review, *Sustain. Futures*, 2. https://doi.org/10.1016/j.sftr.2020.100023 100023

Pittino, F., Puggl, F., Puggl, M., Moldaschl, T., & Hirschl, C. (2020). Automatic Anomaly Detection on In-Production Manufacturing Machines Using Statistical Learning Methods. *Sensors*, *20*, 2344. https://doi.org/10.3390/s20082344

Rivetti, N., Busnel,Y., & Gal, Y. (2017). Grand Challenge: FlinkMan – Anomaly Detection in Manufacturing Equipment with Apache Flink. *11th ACM International Conference on Distributed and Event-based Systems, Jun 2017, Barcelone, Spain*. pp. 274-279. https://doi.org/10.1145/3093742.3095099

Spirin, N., & Han, J. (2012). Survey on web spam detection: principles and algorithms. *ACM SIGKDD Explor. Newsl.*, 13(2), 50–64. https://doi.org/10.1145/2207243.2207252

Tamboli, J., & Shukla, M. (2016). A Survey of outlier detection algorithms for data streams. In: Computing for sustainable global development (INDIACom), *2016 3rd International Conference on. IEEE*, pp. 3535–3540.

Tang, Z., Chen Z., Bao, Y., & Li, H. (2018). Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring. *Structural Control Health Monitoring*, *26*, e2296 http://dx.doi.org/10.1002/stc.2296

Tatu, A., Maaß, F., Färber, I., Bertini, E., Schreck, T., Seidl, T., & Keim, D. (2012). Subspace search and visualization to make sense of alternative clusterings in high-dimensional data. *In: Visual analytics science and technology (VAST), 2012 IEEE Conferenc*e, pp. 63–72. https://doi.org/10.1109/VAST.2012.6400488

Wang, K. (2016). Intelligent Predictive Maintenance (IPDM) system – Industry 4.0 scenario. *WIT Trans Eng. Sci*., 113, 259–268. https://doi.org/10.2495/IWAMA150301

Yan, J., Meng, Y., Lu, L., & Li. L. (2017). Industrial big data in an industry 4.0 environment: challenges, schemes, and applications for predictive maintenance. In *IEEE Access*, *5*, 23484-23491. https://doi.org/10.1109/ACCESS.2017.2765544

Zheng, J., Li, J., Liu, C., Wang, J., Li, J., & Liu, H. (2022). Anomaly detection for high dimensional space using deep hypersphere fused with probability approach. *Complex & Intelligent Systems*, 8, 4205–4220. https://doi.org/10.1007/s40747-022-00695-9

Zope, K., Singh, K., Nistala S. H., Basak, A., Rathore, P., & Runkana, V. (2019). Anomaly Detection and Diagnosis in Manufacturing Systems: A Comparative Study of Statistical, Machine Learning and Deep Learning Techniques. *Annual Conference of the PHM Society, Arizona USA, 2019*. https://doi.org/11. 10.36001/phmconf.2019.v11i1.815

Żabiński, T., Mączka, T., Kluska, J., Madera, M., & Sęp, J. (2019). Condition monitoring in Industry 4.0 production systems - the idea of computational intelligence methods application. *Procedia CIRP*, *79* (3-4), 63-67. https://doi.org/10.1016/j.procir.2019.02.012

Zuo, L., Zhang, H., & Wang, H. (2021). Optimal subsample selection for massive logistic regression with distributed data. *Comput. Stat*., *36*, 2535–2562. https://doi.org/10.1007/s00180-021-01089-0

How to cite this Article:

Milton Samadder, Anup Kumar Barman and Alok Kumar Roy (2023). Examining a generic streaming architecture for smart manufacturing's Big data processing in Anomaly detection: A review and a proposal. *International Journal of Experimental Research and Review*, *30*, 219-227.

DOI : https://doi.org/10.52756/ijerr.2023.v30.019