Hindawi
Security and Communication Networks
Volume 2021, Article ID 3320436, 9 pages
https://doi.org/10.1155/2021/3320436

Research Article

Privacy-Aware Data Forensics of VRUs Using Machine Learning and Big Data Analytics

Muhammad Babar,[1] **Muhammad Usman Tariq**,[2] **Ahmed S. Almasoud**,[3] **and Mohammad Dahman Alshehri**[4]

[1]*Department of Computer Science, Allama Iqbal Open University, Islamabad, Pakistan*
[2]*Abu Dhabi School of Management, Abu Dhabi, UAE*
[3]*College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia*
[4]*Department of Computer Science, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia*

Correspondence should be addressed to Muhammad Babar; muhammad.babar@aiou.edu.pk

Received 27 September 2021; Revised 4 November 2021; Accepted 12 November 2021; Published 28 November 2021

Academic Editor: Farhan Ullah

The present spreading out of big data found the realization of AI and machine learning. With the rise of big data and machine learning, the idea of improving accuracy and enhancing the efficacy of AI applications is also gaining prominence. Machine learning solutions provide improved guard safety in hazardous traffic circumstances in the context of traffic applications. The existing architectures have various challenges, where data privacy is the foremost challenge for vulnerable road users (VRUs). The key reason for failure in traffic control for pedestrians is flawed in the privacy handling of the users. The user data are at risk and are prone to several privacy and security gaps. If an invader succeeds to infiltrate the setup, exposed data can be malevolently influenced, contrived, and misrepresented for illegitimate drives. In this study, an architecture is proposed based on machine learning to analyze and process big data efficiently in a secure environment. The proposed model considers the privacy of users during big data processing. The proposed architecture is a layered framework with a parallel and distributed module using machine learning on big data to achieve secure big data analytics. The proposed architecture designs a distinct unit for privacy management using a machine learning classifier. A stream processing unit is also integrated with the architecture to process the information. The proposed system is apprehended using real-time datasets from various sources and experimentally tested with reliable datasets that disclose the effectiveness of the proposed architecture. The data ingestion results are also highlighted along with training and validation results.

1. Introduction

In a recent technological globe, data are mounting rapidly, and humans are mostly relying on data. Besides the pace at which the data rise, it is becoming impracticable to stock up the data into any specific server. Today the planet holds an enormous quantity of data that persists to grow exponentially at very high speed and is insecure [1]. Moreover, the entire globe has gone online with the invention of the web, and every single action we do puts down a digital map out that is prone to vulnerability [2]. With the rise of big data and machine learning, the notion of improving accuracy and enhancing the efficacy of AI projects is also gaining importance and is largely recognized [3]. Some of these factors of the evolution of data are the enhancement of technology, social media, and Internet of Things (IoT). IoT is one of the latest concepts in the current age that is mostly applicable in traffic controlling and monitoring applications. The future of this globe is secure IoT that will be going to alter today's world objects into intelligent and smart objects [4]. Smart systems include IoT devices, such as sensors and actuators, process input connectivity, and people. Sensors and

actuators are acting as a backbone for any emerging system. The interactions among all these components create a new type of smart application and service. With the rise of IoT devices, the idea of edge computing is also gaining prominence and is broadly recognized. Machine learning solutions provide improved guard safety in hazardous traffic circumstances in the context of traffic applications [5–7].

As several new-fangled and ground-breaking technologies pledge benefits through enhanced optimization of traffic community systems, "Smart" traffic system development chooses the best of these techniques and services to resolve traffic most imperative confronts [8–10]. Hence, the smart traffic trend going towards the higher side. There are many aspects of urban from transportation management to building blueprint and community safety, which are examined as grown for reinvention. Besides, some cutting-edge and imperative technologies such as cloud computing, robotics, artificial intelligence, machine learning, big data, and particularly, machine learning seems progressively more within the reach [3, 11]. The overall big data analytics process goes through several stages to serve the purpose [12]. These stages include identification of the problem, designing the data requirements, preprocessing, data loading, performing processing and analytics, and data visualization. Firstly, all the problems are needed to be identified accurately which are required to be addressed using big data analytics. Then, all the data requirements are designed to provide a logical solution to be executed in the later stages. Big data are usually very chaotic, messy, incoherent, incomplete, and inconsistent [13].

Therefore, proper preprocessing is required to be done before processing the big data. Consequently, the next phase is the data loading before the data processing and analytics of big data. The smart traffic environment is based on IoT devices and objects generating gigantic data (Big Data) which requires efficient aggregation, processing, and analysis to achieve optimal results for decision-making [14, 15]. Efficient exhaustive analysis of such data is not possible through traditional data analytics techniques. On the contrary, some big data analytics methods are also found in the past other than traditional methods; however, there is no all-inclusive, common, and effective resolution proposed to aggregate and process the big data produced in an IoT-based smart traffic environment [16–18]. The existing solutions are based on traditional or classical Hadoop framework. Moreover, the data ingestion or data loading performance of big data files into Hadoop is overlooked in the existing solutions, which is one of the major factors affecting the overall processing [19, 20]. Big data analytic involves smart management of the data to give real-time monitoring of the data population of the VRUs which has drastically expanded everywhere throughout the world. The solutions using IoT big data are proposed for the VRUs' information management along with traffic management. This research prefers the customization of the YARN parallel and distributed framework. However, to comprehend the reliability of the smart traffic, many challenges are required to be addressed where privacy is one of the most brutal between the imperative challenges.

2. Literature Review

A malevolent hit on the services of users can be extremely costly in the context of the trustworthiness of edge computing [21, 22]. Hence, this article presents a secure architecture for data supervision to deal with data security challenges in smart traffic applications. The work related to the proposed architecture about data analytics and machine learning for smart traffic data management is very significant. The key problems decorated in the architecture are the use of traditional MR cluster, inadequate data piling, intangible structure, and only specific dataset [10, 23]. A scheme was discussed in detail in the context of V2X connections [24]. The bog data analytics approaches are also taken into consideration including Tiers that are accountable for various steps and activities of the data analytics. Though it is a complete four-tier architecture consisting tiers from data collection to data analysis usage, it causes processing delay [25, 26], and a classical map-reduce framework is used that slows down the performance. Moreover, data aggregation before data loading is focused while data loading competence is overlooked. The data aggregation of results is preferred and data loading before analysis is overlooked in this architecture.

An approach is proposed for reducing the conflict between VRUs and automated vehicles [27]. This proposal is only focusing on automated vehicles. It does not support big data processing in general. The key issue is the data ingestion performance in this model. It takes a lot of time to insert the big data into the system for processing. Some researchers proposed a model based on data analysis that promotes the notion of smart traffic and utilizes the big data to be processed but overlooks the data loading efficiency. A framework is presented to overcome the VRUs' issues, but these researchers also overlook the data loading and ingestion into a distributed environment. There are some models proposed to deal with the same problem of Big Data analysis in the smart environment [14, 28], but this solution is the utilization of the conventional cluster resource management scheme and insufficient data loading to the Hadoop server. Moreover, architecture is proposed to investigate the data in a transport environment that is more accessible and efficient [29]. However, it causes an additional delay in processing, and the said scheme is only tested for transportation datasets, and data load efficiency is overlooked while loading Big Data to Hadoop server as well. The additional delay affected the overall performance of the big data analytics.

On the contrary, a scheme is proposed using a parallel processing approach. Though a YARN-based solution is offered, the data loading efficiency is still overlooked in this architecture. The standard practice of traditional data analytics techniques is to analyze the limited data only, which generates an open area of errors and biases in the Big Data scenario. Another challenge that needs to be addressed is insufficient data loading into the traditional cluster management framework, e.g., Hadoop. The traditional data loading challenges are time-consuming, more storage required, commands are difficult, no append, and no partial ingestion. Similarly, Hadoop processing based on traditional

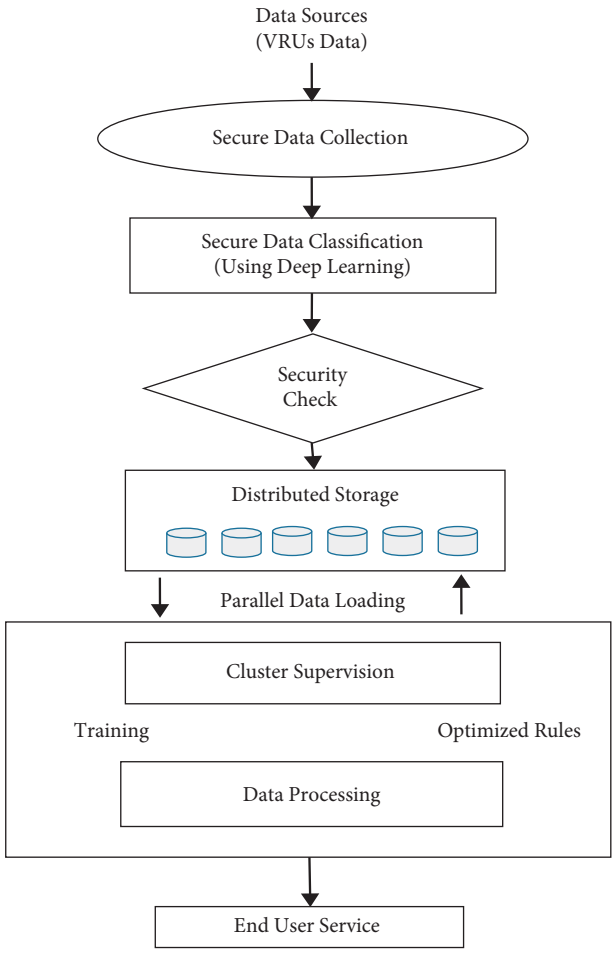

Figure 1: Overview of the proposed framework.

cluster management challenges includes scheduling issues, inefficient load balancing, scalability issues, NameNode availability, and responsibility unification. The objective of this research is to propose a framework based on edge intelligence to process enormous data efficiently and overcome the data loading and processing issue. IoT gathers data and directs the driver to follow the free lanes. A specific proposal is designed to realize the map-reduce paradigm integrated with Apache Spark for real-time data processing comprehension of big data. Spark deals with hasty computation and allows reusability. Effective data ingestion into the distributed storage mechanism is missing in the loading and storage process efficiently.

The current work has deficiencies in the big data storage and processing for IoT-enabled intelligent transportation. Furthermore, model parallelism is also missing for effective extrapolation and decision-making. The proposed research will propose a framework to overcome the existing challenges. Trust and privacy in the smart traffic application, particularly considering the VRUs, is a prejudiced experience that brings complexity in recognizing the attacks. The insecure VRUs in the smart traffic applications could cause a breakdown in the transportation monitoring and controlling services. Therefore, to enhance security, we need to evaluate the level of insecurity in an application first. This study proposed a secure architecture based on machine learning in the smart traffic domain that evaluates the privacy level of the VRUs.

3. Proposed Architecture

The proposed architecture based on machine learning connects the smart community departments (e.g., traffic monitoring and control department). The data sources are comprised of traffic monitoring and controlling big data. The workflow of the proposed parallel and the distributed scheme is depicted in Figure 1. Data gathering is done by the respective units collected from various traffic control sources (e.g., sensors and cameras). To devise effective parallel and distributed architecture, the data must be scrutinized before computation. The data are generated by different devices such as environmental sensors, security monitoring sensors, traffic cameras, and transportation monitoring sensors. The data are properly collected by the various departments such as the traffic-controlling authorities. This process is known as secure data collection. The data are classified using the machine learning approach.

The data are given to the proposed parallel and distributed architecture to process using proposed modules. The number of parallel changes is balanced using the fixed block size of the chunk. The default block size of the utility is time-consuming and has less parallelism. The default size is optimized and modified to improve the data loading efficiency. This data collection is a part of a distributed system. It involves overall data management that includes aggregation, collection, and storage. The data are also preprocessed before injecting into the proposed scheme to remove noise and anomalies for speeding up the processing activities. Afterward, the data are divided into different chunks for parallel processing at the edge level. The distributed storage mechanism is also taken into consideration to assist the parallel processing. The YARN parallel and distributed platform for big data analytics is preferred because the cluster management is dealt with separately by the resource manager that is a part of the YARN. Premediated algorithms are applied while processing the data in the cluster.

The processed results are sent for decision-making to the concerned smart society services' providers that are finally forwarded to the users. Following filtration, the Hadoop processing unit is used to process the data which are stored in the distributed storage mechanism. Lastly, the analyzed data are operated for community planning. The data are collected from the departments, and the decisions are sent back to the community development departments. The objective is to realize a smart traffic scheme to perform processing and keep the data private. The said-community departments are the data sources for the proposed system and a mediator between the system and the user. Architecturally, the anticipated solution consists of 3 modules that are data security, organization, and processing, which are shown in Figure 2.

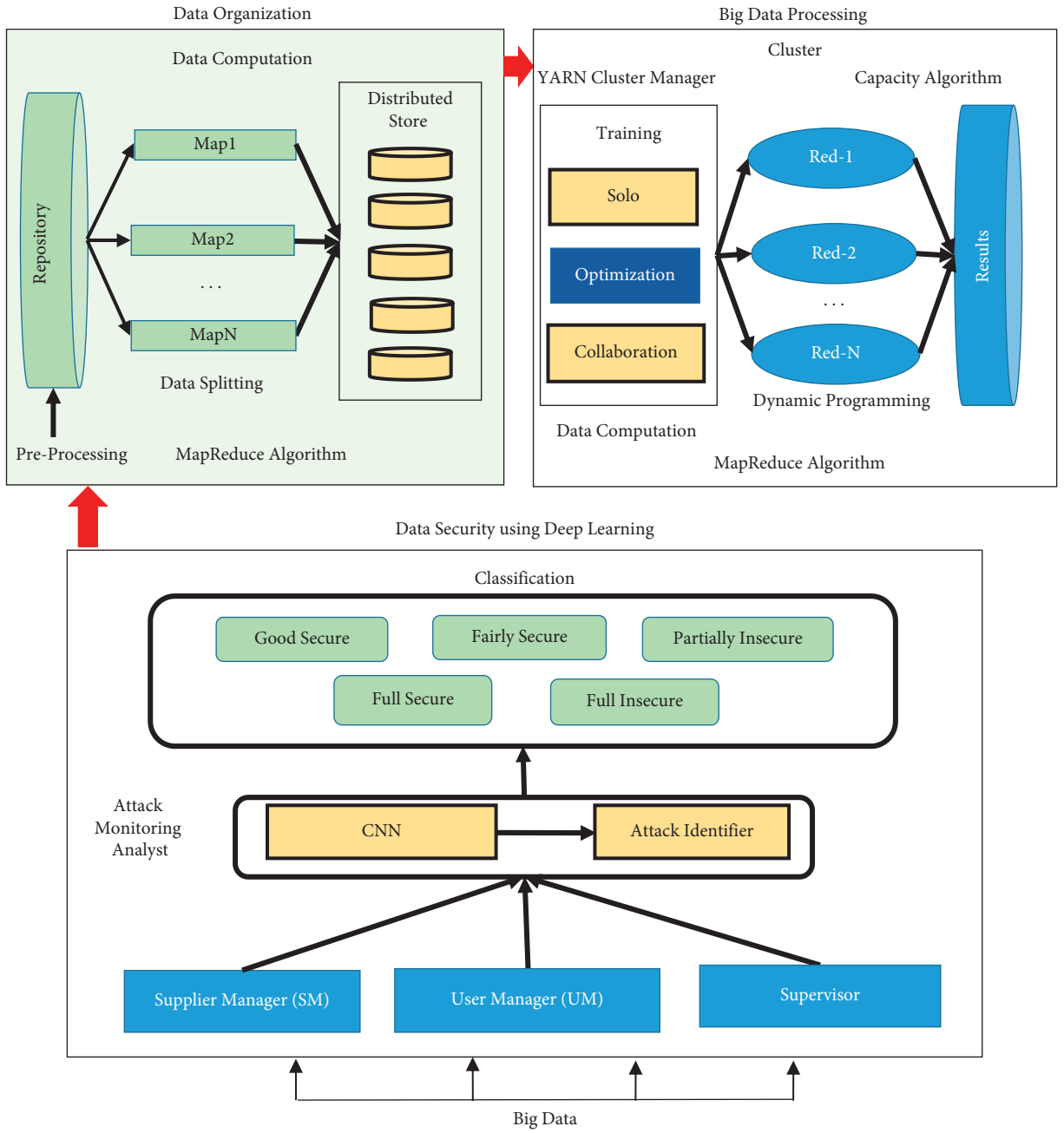

Figure 2: Proposed architecture.

3.1. Data Security Layer. The proposed structural design has a security layer for keeping secure the VRUs' data from attacks. It is a part of smart traffic architecture. It recommends flexibility in opposition to the attacks. The supplier manager (SM), user manager (UM), and superviser are the components' security layer. The SM and UM pay attention to the supplier and the user, while the supervisor applies the algorithm of machine learning. The CNN DL technique is integrated that classifies secure or insecure data. The SM is accountable for the profile maintenance of every supplier, and the UM is accountable for the profile maintenance of users. The supervisor is trained using the classifier. The level of security is predicted using special classes that are highly secure (HS), fine secure (FS), moderately secure (MS), highly insecure (HIS), and partly insecure (PIS). Equation (1) is used to compute the security score:

$$\text{Lev\_of\_Security} = \begin{cases} \text{HS}, & \text{if Sec} = 1, \\ \text{FS}, & \text{if Sec} \geq 0.76 \text{ and } S < 1, \\ \text{MS}, & \text{if Sec} \geq 0.51 \text{ and } S < 0.76, \\ \text{HIS}, & \text{if Sec} \geq 0.26 \text{ and } S < 0.51, \\ \text{PIS}, & \text{if Sec} = 0 \text{ and } S < 0.26, \end{cases} \tag{1}$$

where Sec is computed using

$$\text{Sec} = \frac{1}{n}\sum_{k=0}^{n} Fk. \quad (2)$$

Equation (1) is used to calculate the various levels of security. The purpose of the different security levels is to give the particular score to the candidate user for the prospect. The major purpose of the multiclassification is to identify the watch list of the risks in the future. It helps identify the intruders with less score to be analyzed further for future investigation.

3.2. Big Data Organization. The big data organization system involves the overall data management including aggregation, collection, and storage. The data are distributed across various nodes for computation to get a load from the central server or cloud. Intelligent applications are supported by acquiring data via the Internet from various local devices. Several devices that include sensors, cameras, and object-mounted devices record the information of the environment in the different domains. This data are later utilized for analysis to get insights and produce intelligent decisions. It is the first layer that is accountable for assembling the data from different community departments that are used to manage the smart community development services. A practical community does not only hold large data only but also includes versatile and wide-ranging processing areas. The smart community implementation is dependent on all forms of data processing due to their heterogeneous nature. Data collection is used to transform signals that are assessed in practical circumstances and converts outcomes to the digital form for processing. The collection is done by a special system that converts data from analog to digital form. The smart traffic centers pull out the data using various sensors in the community to gather real-time data. The data organization layer further contains the data aggregation, where the data are grouped based on the identification of the connected devices. This aggregation process is implemented due to the data size because the data are very massive and required to be assembled for efficient processing. The aggregation improves the modularity and processing.

3.3. Big Data Processing. This unit is the main processing part that preprocesses the raw data initially including the irrational data combination, missing values, and values beyond the range which are integrated before processing. If the data are not inspected for such problems, there could be misleading results during decision-making. Hence, the transformation is also done to scale the data to a specific scale. Then, the data are taken by a parallel processing unit that is the backbone of the proposed architecture. The proposed architecture is based on a parallel computing model called MapReduce that is utilized. MapReduce is introduced to realize big data analytics. This programming paradigm is composed of Map and Reduce functions. It is a useful model that exploits huge datasets and processes them in parallel. It executes processes in a distributed manner and offers high availability. The underlying system also manages machine failures, performance issues, and efficient communications. Task distribution in the cluster is carried out using the YARN distributed cluster management framework. The YARN is equipped with dynamic programming for task distribution and cluster management. The previous platforms such as MapReduce paradigm are only responsible for the processing. The YARN is preferred because the cluster management is dealt with separately by the resource manager that is a part of the YARN. The fair algorithm is integrated with YARN to perform scheduling. Besides, interleaving is possible between map and reduce phases; therefore, the reduced phase might begin before the map phase finishes.

4. Results and Discussion

The proposed scheme is implemented using the parallel and distributed platform of Hadoop version 3.0. The Hadoop is equipped with Apache Spark module. The reliable datasets are utilized. The pyspark library is utilized in Python 3.8. Similarly, the resilient agent evaluation is carried out using a detailed setting with a machine learning classification module. The machine learning library is also utilized and implemented in Python 3.8. The comparative analysis of the proposed design is provided with current proposals. The experimental results and comparison disclose the effectiveness of the proposed design. The discussion about the results is provided in this section. Results are produced using various reliable datasets to assess the proposed architecture based on parallel and distributed paradigms using premeditated algorithms. We performed a noise and anomalies removal process on data on top of our proposed architecture. The anomalies are removed using the min-max normalizations and Kilman algorithm. The data ingestion is achieved using the map-only algorithm.

The traditional YARN cluster management framework is customized with improved capacity and a fair algorithm of scheduling. We applied the dynamic algorithm to set the parameters of the YARN framework dynamically. The processing is performed using MapReduce algorithms. We also optimize the MapReduce algorithm for edge computing to utilize at every edge. Thus, notable efficiency is achieved in the processing time. The proposed architecture implemented using the Hadoop parallel and distributed framework along with optimized algorithms. These datasets are preferred due to the utilization of this dataset in the literature. We deliberately executed almost the same queries to compare the processing time and throughput of proposed edge-enabled IoT architecture using customized MapReduce and YARN for parallel processing.

4.1. Data Security Results. The results and experiments of the security layer include the required training of the dataset using an ML classifier. The model is trained using secure and insecure interaction with the proposed architecture. The assessment of the security layer is performed in a specific setting. Initially, the model was trained on $365 * 925$ matrices. The training process of the Naïve Bayes classifier is shown in Figure 3.

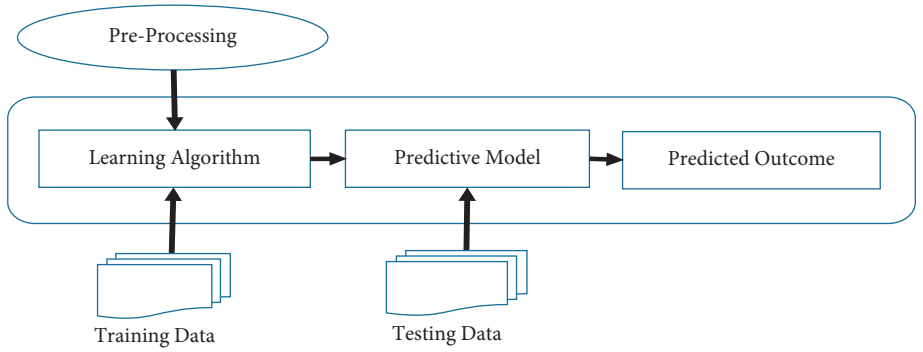

Figure 3: Model training.

Table 1: Confusion matrix.

		Predicted	
		Secure	Insecure
Labeled	Secure	TP	FN
	Insecure	FP	TN

Table 2: Proposed model results using the confusion matrix.

		Predicted	
		Secure	Insecure
Actual	Secure	96.3%	3.7%
	Insecure	4.1%	95.9%

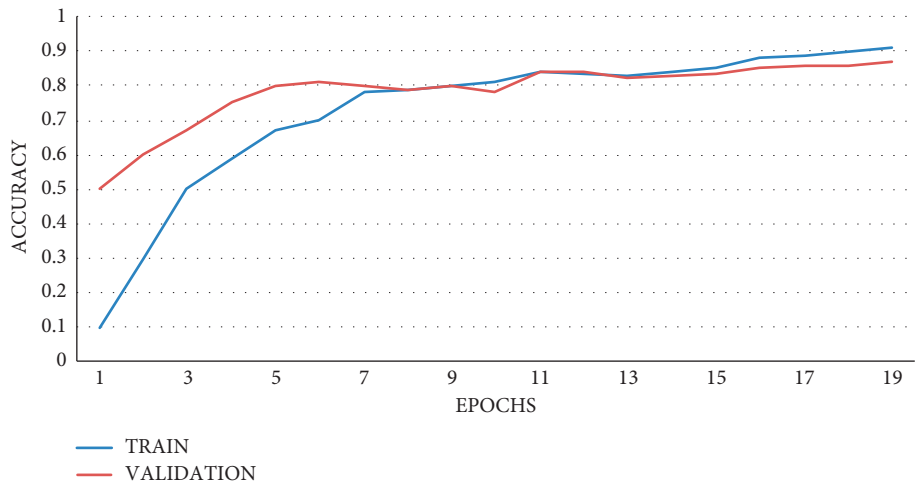

Figure 4: Accuracy.

The proposed resilient agent evaluation is also performed by setting a specific environment where the proposed model is trained using the proposed model. To assess the proficiency of the proposed model, the confusion matrix is exploited, as depicted in Table 1. To measure the effectiveness of the classifier, the confusion matrix is utilized concerning two classes (e.g., secure and insecure), as shown in Table 2. The value is considered secure if it is greater than 0.5; otherwise, it is considered insecure. The performance measures are applied to the ML technique utilized for a resilient agent. The accuracy of the technique is expressed in the form of percentages in Table 2. The specific value of percentage of each confusion matrix value is also highlighted in Table 2.

4.2. Training and Validation Results. Figure 4 is the confirmation of the enhanced accuracy of the validation and training. The enhanced level of accuracy in training and validation is a result of the enlarged number of epochs (e.g., 200 epochs). Likewise, Figure 5 reveals the proposed model's validation and training loss that is the indication of minimal loss. The reduction in the loss is a result of the enlarged number of epochs (e.g., 200 epochs).

4.3. Data Ingestion Results. It gets nearly no time to load the dataset into Hadoop when the dataset size is small. There are not quite time differences of data loading either manually or

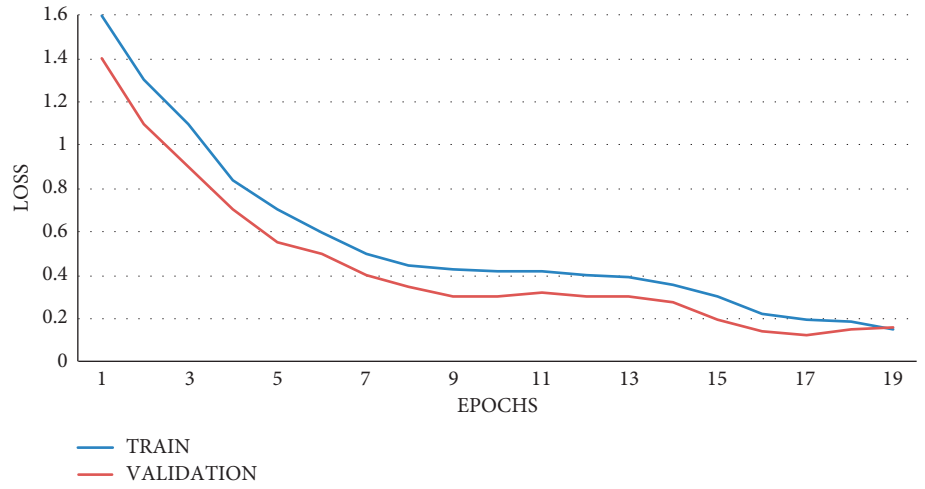

Figure 5: Loss.

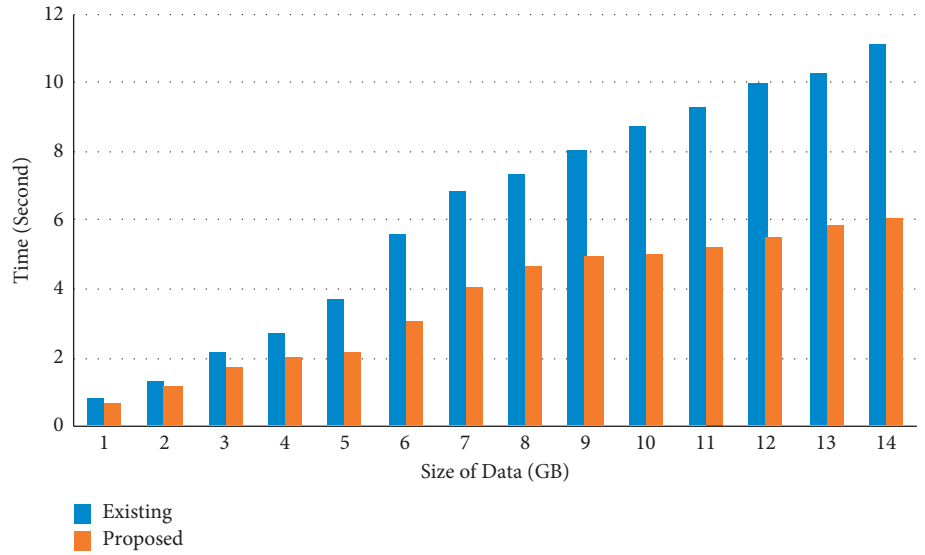

Figure 6: Overall data loading efficiency.

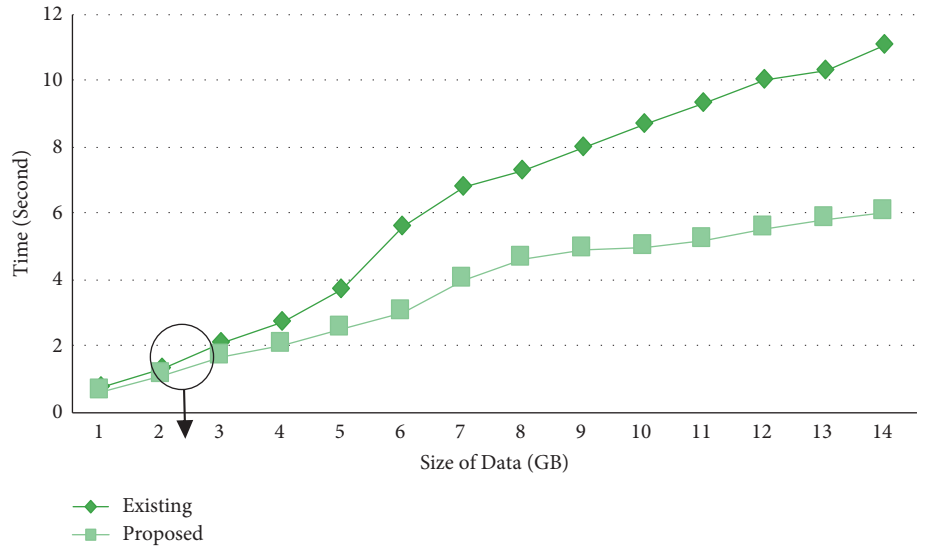

Figure 7: Dataset size threshold (overall).

using the specific utility. It has been experimentally proved that it gets nearly no time to load the dataset into Hadoop when the dataset size is small.

Overall, the proposed system efficiency including all the parameters' modification of data loading is shown in Figure 6. In the same way, Figure 7 demonstrates the threshold for all the parameters' modification of data loading using the proposed system. The threshold is the alarming set value that highlights the focal point where the difference between existing and proposed schemes starts. The proposed scheme is manual in the context of data ingestion and automated in the context of classification and processing.

5. Conclusion

A smart traffic application is considered by the extensive expansion of IoT-connected devices particularly with the rise of Big Data and machine learning. Machine learning solutions provide efficient results in the context of efficiency and accuracy of the machine learning models. However, it becomes challenging to tackle the privacy of the users in the smart traffic management and surveillance of the users because that produces an enormous amount of big data to be processed and analyzed efficiently. In this study, an architecture is proposed based on machine learning to process big data efficiently in a secure environment considering user privacy. The proposed architecture is a layered framework with a parallel and distributed module using machine learning on big data to achieve secure big data analytics. A specific privacy layer is proposed that classifies the dishonest entities using machine learning. The proposed system is apprehended using real-time datasets from various sources and experimentally tested with reliable datasets that disclose the effectiveness of the proposed architecture. The data ingestion results are also highlighted along with training and validation results. This study proposes an architecture based on machine learning to process big data efficiently in a secure environment considering user privacy. The proposed design is the optimization of the existing parallel and distributed framework to achieve efficient processing. The current proposals lack efficient parallel data ingestion and efficient mechanism for communication overhead. Therefore, the security challenges using machine learning are explored in this paper. This paper proposes a separate secure and resilient module to overcome the privacy issue of the users. The proposed architecture is equipped with a resilient agent using an ML classifier. A stream processing unit is also integrated with the architecture to process the information produced by edge devices.

Data Availability

The data used to support the findings of the study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Taif University Researchers Supporting (project no. TURSP-2020/126), Taif University, Taif, Saudi Arabia.

References

[1] M. I. Razzak, M. Imran, G. Xu, and G. Xu, "Big data analytics for preventive medicine," *Neural Computing & Applications*, vol. 32, no. 9, pp. 4417–4451, 2020.

[2] N. Paltrinieri, L. Comfort, and G. Reniers, "Learning about risk: machine learning for risk assessment," *Safety Science*, vol. 118, pp. 475–486, 2019.

[3] S. K. Maurya and A. Choudhary, "Deep learning based vulnerable road user detection and collision avoidance," in *Proceedings of the 2018 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, pp. 1–6, IEEE, Madrid, Spain, 2018 September.

[4] M. Ahmad, T. Younis, M. A. Habib, R. Ashraf, and S. H. Ahmed, "A review of current security issues in internet of things," *Recent Trends and Advances in Wireless and IoT-enabled Networks*, Springer, Singapore, pp. 11–23, 2019.

[5] M. Garcia-Venegas, D. A. Mercado-Ravell, and C. A. Carballo-Monsivais, "On the safety of vulnerable road users by cyclist orientation detection using Deep Learning," 2020, https://arxiv.org/abs/2004.11909.

[6] M. Goldhammer, S. Köhler, S. Zernetsch, K. Doll, B. Sick, and K. Dietmayer, "Intentions of vulnerable road users—detection and forecasting by means of machine learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 7, pp. 3035–3045, 2019.

[7] Z. Ahmed and R. Iniyavan, "Enhanced vulnerable pedestrian detection using deep learning," in *Proceedings of the 2019 International Conference on Communication and Signal Processing (ICCSP)*, pp. 0971–0974, IEEE, Chennai, India, April 2019.

[8] M. Babar and F. Arif, "Real-time data processing scheme using big data analytics in internet of things based smart transportation environment," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 10, pp. 4167–4177, 2019.

[9] E. Carter, P. Adam, D. Tsakis, S. Shaw, R. Watson, and P. Ryan, "Enhancing pedestrian mobility in smart cities using big data," *Journal of Management Analytics*, vol. 7, pp. 1–16, 2020.

[10] W. Tabone, J. de Winter, C. Ackermann et al., "Vulnerable road users and the coming wave of automated vehicles: expert perspectives," *Transportation Research Interdisciplinary Perspectives*, vol. 9, Article ID 100293, 2021.

[11] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: a deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865–873, 2014.

[12] A. Sharma, G. Singh, and S. Rehman, "A review of big data challenges and preserving privacy in big data," in *Advances in Data and Information Sciences*, pp. 57–65, Springer, Singapore, 2020.

[13] S. Boubiche, D. E. Boubiche, A. Bilami, and H. Toral-Cruz, "Big data challenges and data aggregation strategies in wireless sensor networks," *IEEE Access*, vol. 6, pp. 20558–20571, 2018.

[14] D. Nallaperuma, R. Nawaratne, T. Bandaragoda et al., "Online incremental machine learning platform for big data-driven

smart traffic management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 12, pp. 4679–4690, 2019.

[15] A. Ahmad, M. Babar, S. Din et al., "Socio-cyber network: the potential of cyber-physical system to define human behaviors using big data analytics," *Future Generation Computer Systems*, vol. 92, pp. 868–878, 2019.

[16] M. Elkhodr, B. Alsinglawi, and M. Alshehri, "Data provenance in the internet of things," in *Proceedings of the 2018 32nd international conference on advanced information networking and applications workshops (WAINA)*, pp. 727–731, IEEE, Krakow, Poland, May 2018.

[17] S. G. Farrag, N. Sahli, Y. El-Hansali, E. M. Shakshuki, A. Yasar, and H. Malik, "STIMF: a smart traffic incident management framework," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 1, pp. 85–101, 2021.

[18] A. Ahmad, M. Khan, A. Paul et al., "Toward modeling and optimization of features selection in big data based social internet of things," *Future Generation Computer Systems*, vol. 82, pp. 715–726, 2018.

[19] J. Yang, Y. Han, Y. Wang, B. Jiang, Z. Lv, and H. Song, "Optimization of real-time traffic network assignment based on IoT data using DBN and clustering model in smart city," *Future Generation Computer Systems*, vol. 108, pp. 976–986, 2020.

[20] N. Cárdenas-Benítez, R. Aquino-Santos, P. Magaña-Espinoza, J. Aguilar-Velazco, A. Edwards-Block, and A. Medina Cass, "Traffic congestion detection system through connected vehicles and big data," *Sensors*, vol. 16, no. 5, p. 599, 2016.

[21] J. Zhang, B. Chen, Y. Zhao, X. Cheng, and F. Hu, "Data security and privacy-preserving in edge computing paradigm: survey and open issues," *IEEE Access*, vol. 6, pp. 18209–18237, 2018.

[22] C. Johnsson, A. Laureshyn, and T. De Ceunynck, "In search of surrogate safety indicators for vulnerable road users: a review of surrogate safety indicators," *Transport Reviews*, vol. 38, no. 6, pp. 765–785, 2018.

[23] A. Siulagi, J. F. Antin, and L. J. Molnar, S. Bai, S. Reynolds, O. carsten, and R. greene-roesel, Vulnerable road users: how can automated vehicle systems help to keep them safe and mobile?" in *Road Vehicle Automation*, vol. 3, pp. 277–286, Springer, Cham, 2016.

[24] N. Dasanayaka, K. F. Hasan, C. Wang, and Y. Feng, "Enhancing vulnerable road user safety: a survey of existing practices and consideration for using mobile devices for V2X connections," 2020, https://arxiv.org/abs/2010.15502.

[25] H. S. A. J. J. Sun and R. Bie, "Internet of Things and big data analytics for smart and connected communities," *IEEE Access*, vol. 4, pp. 766–773, Mar. 2016.

[26] R. Tönjes, P. Barnaghi, M. Ali et al., "Real time iot stream processing and large-scale data analytics for smart city applications," in *Proceedings of the European Conference on Networks and Communications (poster session)*, Bologna, Italy, 2014.

[27] J. M. Owens, R. Greene-Roesel, A. Habibovic, L. Head, and A. Apricio, "Reducing conflict between vulnerable road users and automated vehicles," in *Road Vehicle Automation*, vol. 4, pp. 69–75, Springer, Cham, 2018.

[28] A. Ahmad, A. Paul, M. M. Rathore, and H. Chang, "Smart cyber society: integration of capillary devices with high usability based on Cyber–Physical System," *Future Generation Computer Systems*, vol. 56, pp. 493–503, 2016.

[29] S. Shukla, K. Balachandran, and V. S. Sumitha, "A framework for smart transportation using Big Data," in *Proceedings of the 2016 International Conference on ICT in Business Industry & Government (ICTBIG)*, pp. 1–3, IEEE, Indore, India, November 2016.