Development of IoT-based Machine Learning Application for Data Anomaly Detection With in a Smart Manufacturing Plant

T.G Kukuni[1],E. Markus[1*], B. Kotze[1#], A.M. Abu-Mahfouz[2]
[1]Department of Electrical, Electronic and Computer Engineering
Central University of Technology, Free State
Bloemfontein, South Africa
[1]tgkukuni@gmail.com,[1*]emarkus@cut.ac.za, [1#]bkotze@cut.ac.za,
[2]Council for Scientific and Industrial Research, Gauteng
Pretoria, South Africa
[2]aabumahfouz@csir.co.za

Abstract

The application of machine learning in resolving complex cyber-security challenges in smart manufacturing plant is growing. Network intrusion and anomaly detection is posing high risks in sensory data integrity and optimisation of processes leading to high efficiency and high profits within smart manufacturing plants. This research paper makes use of interquartile range algorithm for the detection of anomalies. The data is collected within a 5-hour period and is transmitted to Google sheets via WiFi connectivity. However, the data transfer requires the user to permit access to the google account for this process to take place. After the addition of errors for every 11th entry, the file is resent back to Raspberry PI for the execution of interquartile range algorithm. Once the results are obtained, the results file is transmitted via WiFi connectivity to the output monitor. This research results demonstrates that if the data collected is higher or lower than the required threshold (11-14^{o}C for temperature and 45-50% for humidity) the system will automatically detect and flag the anomaly. This paper therefore concludes that the use of interquartile range algorithm for anomaly detection based on sensory data is relevant and efficient for such an investigation.

Keywords: *Anomaly, Intrusion Detection System, Machine Learning, Cyber-security, IoT, Sensory Data, Smart Manufacturing Plant.*
DOINumber:10.14704/nq.2022.20.10.NQ55352 **NeuroQuantology2022;20(10):3637-3648**

1. Introduction

In recent years, the advancement in internet connectivity and communication has increased rapidly. This has led to the rise in Information and Communication Technologies (ICT) devices becoming

necessary for daily activities and for conducting business and solving daily challenges. However, despite the necessity for data transmission across different networks, the networks themselves become vulnerable to exploitation and intrusion and become prone to cyber-attacks [1]. Internet of Things (IoT) is denoted as a system of interrelated computing devices, mechanical and digital machines, objects, etc. that are provided with unique identifiers and can transfer data over a network without any human intervention [2]. However, despite the intelligence that IoT framework brings, the anomalies or intrusion cannot always be categorised as intrusion as other anomalies are not necessarily harmful. However, since manual marking of defects is extremely time-consuming, this has led to anomaly detection within manufacturing plants gaining traction amongst researchers [3]. In contrast, intrusion detection involves intelligent monitoring of patterns in a network traffic or log files some of which are based on if-then rules [4]. However, to mitigate this network threat, an Intrusion Detection System (IDS) is required toaid with the discovery and identification of harmful sensory data within the smart manufacturing plant.

The traditional IDS methods are less effective for IoT-based platforms due to factors such as limited energy and data ambiguity. Additionally, Machine Learning (ML) based algorithms have gained a lot of traction and credibility in providing IoT-based solutions for automated anomaly detection. Hence the need to explore and investigate the feasibility for the development of such a model using machine learning approach. Furthermore, ML-based approaches require an input dataset (images or sensory) that can be trained to conduct automated anomaly detection and advance the network security for the manufacturing plant.

The remainder of this paper is as follows: Section 2 outlines the problem statement identified forthe investigation of the development of ML-base anomaly detection model. Section 3 discusses the related work around IDS-based approaches. Section 4 presents the system application modellingand development approach with Section 5 presenting the results and the paperconcluding in Section 6.

2. Problem statement

Despite the large sum of money paid in optimising and securing computer networks and sensitive data utilising approaches such as IDS, there is still a risk for unknown attacks such as data anomalies and malware detection accurate classification. The evaluation of accurate sensory data transfer within amanufacturing setup is very key as it contributes to job sustainability and revenue as well as the safety of the employees. However, due to high rate of cyber-attacks, there is a need for the development of an advanced IDS system that will be used to secure the sensory data and detect any data anomalies within the network model. It is therefore, against this background, that there is a need for the development of ML-based model with the capability for anomaly detection without any human intervention to improve cybersecurity within an industrial setup needs to be investigated.

3. Related work

There are several methods and research based on the advancements in network-based intrusion security. There

have been several approaches such as the use of Keras library on top of a Tensorflow framework using deep learning for enabling seamless exhibition [5]. Keras is an API designed for human beings, and not machines and offers best practices for reducing cognitive load and consistency. This study focuses mainly on machine learning intrusion-based solutions and the reason for that is the advancement of technology both on the intruder's site and cybersecurity point of view. Ashiku et al. [6] proposed a deep learning classification architecture coupled with semi-dynamic hyperparameter tuning for the detection and classification of network attacks. In his approach, the author used UNSW-NB15 datasets primarily for its representation of real-world network traffic. However, in this study, the author`s approach only focused on the classified attacks per day with no data handling results for zero-day attacks. This is important for the stability of the system in terms of correct intrusion exposures and zero-day behavioral features for risk reduction. Furthermore, Kiran et al. [7] presents an IoT-based platform that was built utilising Node MCU ESP8266, DHT11 sensor, ThinkSpeak platform and a wireless router. However, Kiran`s study had challenges with data reliability and high-quality training datasets and as a result of the inconsistent data flow, the system is prone to data interception. Therefore, a model optimisation technique needs to be designed and developed using machine learning algorithms to maintain reliable and high-quality datasets before the datasets can be added to the system.

Saranya et al [8] present a review comparison of different Machine Learning models that are used for intrusion detection systems in various environments. However, the datasets utilised in this research study, are not real-time datasets which opens a window for scrutiny and debate around the legitimacy of the results. In retrospect, Kukuni et al [9]presents a method arrangement for obtaining sensory data by positioning it in a SCADA database and transferring it via a Wemo Microcontroller to an Augmented Reality application using the TCP/IP protocol. However, Kukuni`s method did not consider the data intrusion aspect but rather the accuracy comparison of the data packet between the SCADA data and the AR application sensory data output. Subblakshmi et al [10] present a four-layered classification approach to detect four types of attacks utilising the KDD datasets. However, it is noted that there is a misclassified error in the dataset that the authors did not report. Edith et al [11]presents the use of asymmetric support vector to find different types of attacks in the network. However, despite the author citing the asymmetric support vector as an accurate algorithm, this algorithm is deemed very expensive. Subsequently, Pajouh et al [12] proposed a two-tier approach making use of the NSL-KDD datasets. Pajouh's study focuses on the detection of two specific classified anomalies namely, 1) User to Root (U2R) and 2) Remote to Local (R2L) attacks. In this paper, theresults demonstrate that using the NSL-KDD datasets increases detection performance hence the analogy thatPajouh`s model outperforms the currently existing models.

The use of KDD dataset is one of the most commonly used datasets in IDS systems.However, it is argued that this dataset contributes to the misclassification error of attacks. Additionally, this dataset is associated

with the complexity and accuracy difficulty hence the need to optimise this dataset by reducing the dataset features. In contrast, Salman et al [13] argues that the KDD dataset has been widely used for anomy detection and classification models using machine learning techniques. However, despite Salman`s proposition supported by Subbalakshmi, their comment is generic and doesn't provide any sound technical backing on the misclassification error and whether the features are a contributing factor or not. To expand on the misclassification challenge, Yin adds that not only features might contribute to this misclassificationchallenge, but also the noise in the dataset can also dramatically reduce the classification accuracy and increase the complexity [14].

4. Methodology

The security in a network system is a very pivotal aspect hence the introduction of approaches such as IDS modelling.This paper thus makes mention of use of an IDS based approach for the development of data anomaly detection within a smart manufacturing plant. The model development comprises of the following components:

- DHT11 humidity and temperature sensor – measure both temperature and humidity.
- Raspberry PI – processing of both sensory data and machine learning algorithm.
- Google sheets – is used to store the sensory data.
- Computer – display the output results.

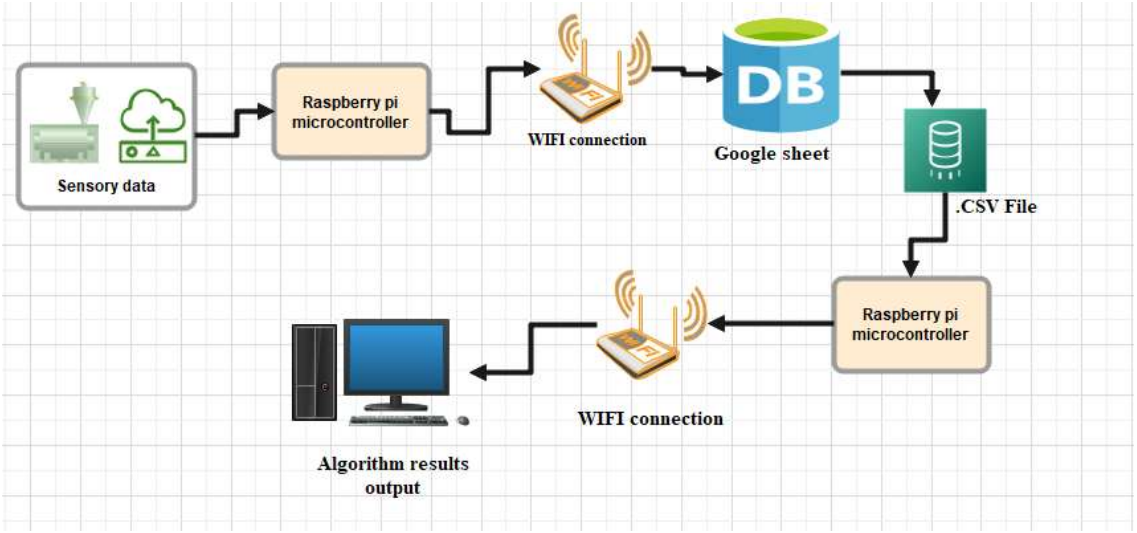


Figure 1. Proposed architectural IoT-based data anomaly detection model.

Figure 1 depicts the proposed model utilisinginterquartile range algorithm for the detection of sensory data anomalies within a smart manufacturing plant network. The sensory data is collected for a period of a day. However, to ensure due to the stability and consistency of the sensory data collected the model was based only on five-hour dataset. The collected sensory data was then

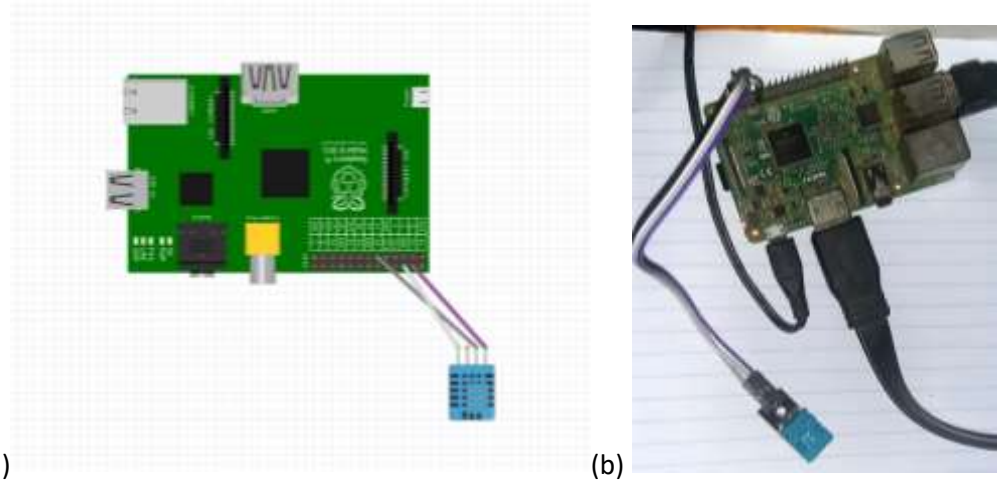
transmitted and uploaded to Google sheets via WiFi connectivity. This data was then converted into a .CSV file to ensure a standardised formatting to be used by the algorithm. The .CSV file was then sent backto the Raspberry PI microcontroller for processing of the algorithm and detection of data anomalies. In this model, the interquartile range algorithm wasused. Once the interquartilerange algorithm is complete, the processed data including the output graphs are send via WiFi connectivity to the results output monitor.

4.1. Systemmodellingdesign

The system modelling was conducted by following a set of processes namely, 1) Data collection; 2) Data processing; 3) Create a model; 4) Evaluate and test the model and 5) Model deployment.

4.1.1. System modelling procedure

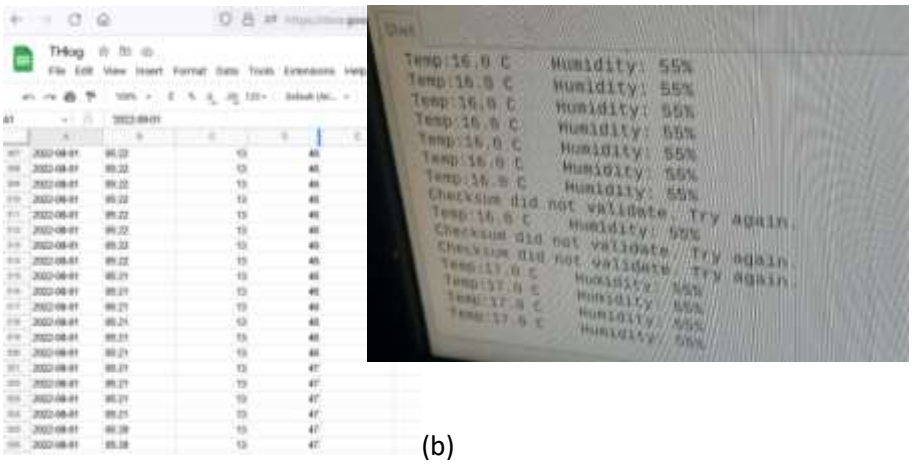
Data collection comprises of the DHT11 humidity and temperature that is connected to the Raspberry PI. Figures 2(a) and (b) depicts the connection model.



(a) (b)

Figure 2.a) Model connection on Proteus software; b) Physical model connection.

Figures 2 (a) and (b) depicts the simulation connection as well as the physical connection. The model connection is programmed to take reading simultaneous for both temperature and humidity every second for a duration of 5 hours. Once this data has been collected and stored, isthen transferred via WiFi connectivity to Google sheets as depicted in Figure 3.



(a) (b)

Figure 3.(a) Google cloud sheet; (b) anomaly detection.

Figure 3 (a) depicts Google cloud sheet that is used as a database for the storage of the .CSV file. However, due to sensory data collection model, anomalies that occurred during the test were not uploaded to the cloud. The reason for this was to process the algorithm with a fixed set of anomalies to test its effectiveness. Furthermore, Figure 3 (b) depicts the errors obtained during the data collection process.

On average, the data collected in a minute is 10 entries with 5 seconds between entries plus 1 second of network lag. However, for datasets where there are less than 10 entries were collected within a minute, other entries are deemed and added to the dataset with the timestamp as highlighted in Figure 4.

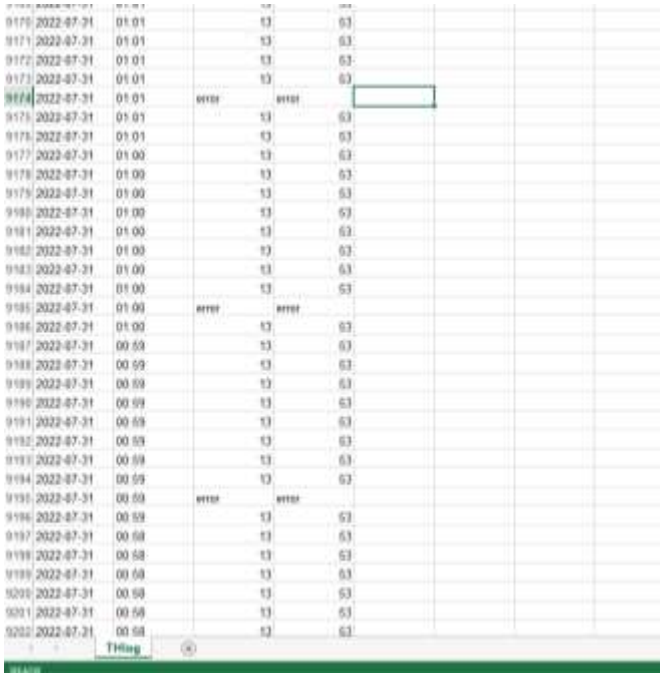


Figure 4. Dataset file with anomalies.

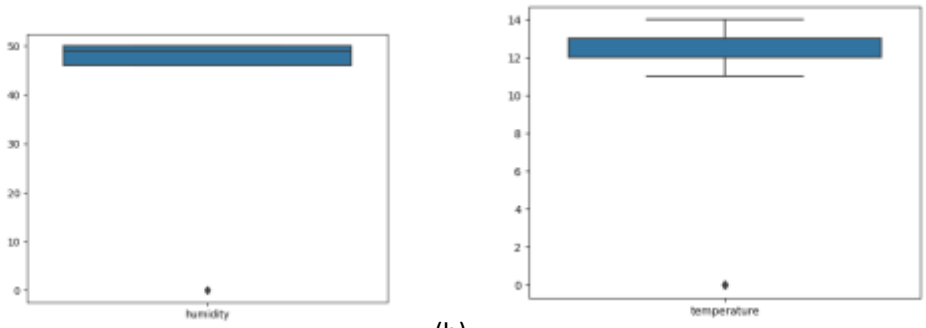
Figure 4 depicts the file with anomalies for every 11th entry.

4.2. Algorithm modelling

The interquartile range algorithm was modelled by first preparing data as outlined in Figure 4. The .CSV file was then placed into the algorithm. However, it is pivotal when using this algorithm to first calculate the quartile which in this case only Q1 (25th lower quartile) and Q3 (upper quartile) were calculated. Upon obtaining the quartiles, both the upper and lower bounds are calculated and the points are filtered as indicated below.

```
lower_data = q1_data - (1.5 * iqr_pc1)
upper_data = q3_data + (1.5 * iqr_pc1)
anomolies = ((data.max() >upper_data) | (data.min() <lower_data)).astype('int')
```

Figures 5 (a) and (b) depicts the graphic view anomalies as outliers.



(a) (b)

Figure 5.(a) Humidity anomaly as an outlier; (b) Temperature anomaly as an outlier.

Figures 5(a) and (b) presents the anomalies as outliers based on the 5 hours collected data. Humidity anomaly occurs when the humidity reading is lower than 45% but greater than 50%, while temperature anomaly any reading lower than 11^{o}C but greater than 14^{o}C.

5. Results& discussion

Anomaly detection model is a very crucial and a very much needed model mainly in the manufacturing sector to reduce cost of production but increase efficiency, profit and job sustainability. In most manufacturing plants the gradual adoption of smart and interconnectivity within devices is experienced. However, this transition is prone to cyber-security threads hence the need for the development of anomaly detection systems. Figures 6 and 7depicts the detected anomalies on sensory data collected within a 5-hour period.

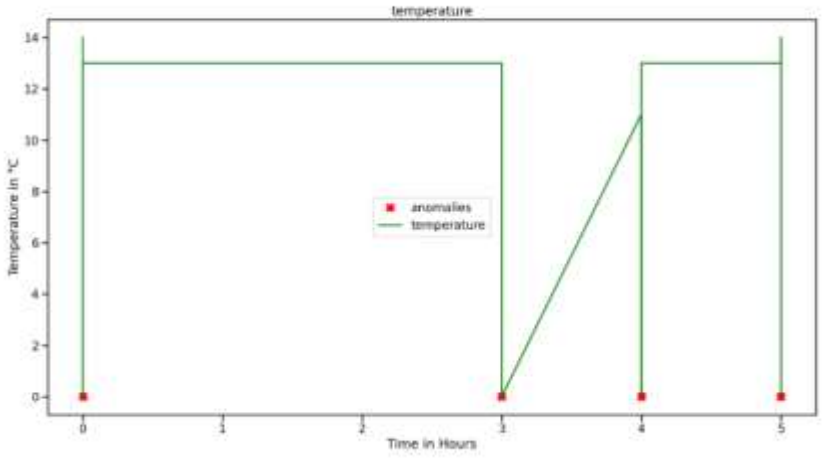


Figure 6. Temperature detected anomalies.

Figure 6 depicts the anomalies detected in a 5-hour data collection period. The collected data was extracted between 5am in the morning and 10am. However, this might contribute to the abnormal anomaly detected at 0th hour

(initial test). In the first three hours, the results demonstrate the system consistency without any anomaly detected with the temperature reading at 13°C. However, for at the third hour, the temperature dropped to below 1°C and it was flagged as anomaly. Between 3rd and 4th hour, the temperature was gradually increasing to almost 10.6 °C and it was also flagged as an anomaly as the required temperature is between 11 °C and 14 °C. Subsequently, the temperature raised to 13°C between the 4th and 5th hour of testing and no anomaly was detected. However, on the 5th the temperature dropped again below the required temperature and the anomaly was detected.

Figure 7 depicts the anomalies detected in the same period but focusing on the humidity reading.

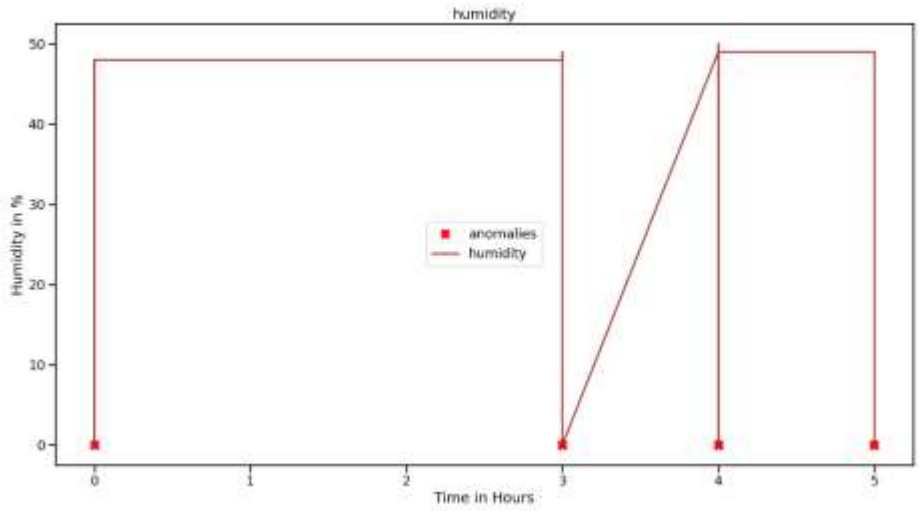


Figure 7. Humidity detected anomalies.

Figure 7 depicts the anomalies detected in a 5-hour data collection period. The collected data was extracted between 5am in the morning and 10am. However, this might contribute to the abnormal anomaly detected at 0th hour (initial test). In the first three hours, the results demonstrate the system consistency without any anomaly detected with humidity reading at 50%. However, for at the third hour, the humidity dropped to below 1% and it was flagged as anomaly. Between 3rd and 4th hour, humidity was gradually increasing to 45%. Subsequently, for the 4th and 5th hour of testing and no anomaly was detected. However, on the 5thhumidity dropped again below the required humidity measurement and the anomaly was detected.

6. Conclusion

The main goal of this research paper was to develop an anomaly detection model utilising machine learning with a smart manufacturing plant. The model was developed by measuring both temperature and humidity utilising DHT11 sensor for a period of 5 hours. Once the 5-hour period was completed, the data was stored transmitted to Google sheets via WiFi connectivity. The file was then converted to .CSV file and errors were manually placed at every 11th entry of the

data collected. Upon completion of this process, the .CSV was transmitted to Raspberry PI microcontroller for the execution of the algorithm and the results were transmitted via WiFi connectivity to the results output monitor.

The results thus demonstrate the feasibility of such a model and its effectiveness. The results presented further demonstrates that when the either the temperature or humidity reading goes below or high the outliers an anomaly is recorded on the model.

7. References

[1] D. Solane and M. Omar, "Using data mining algorithms for developing a model for intrusion detection system," pp. 46-51, 2015.

[2] A. Gillis, "IoT Agenda," TechTarget, August 2021. [Online]. Available: https://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT. [Accessed 19 January 2022].

[3] M. Gamal, A. Donkol, A. Shaban, F. Constantino, G. Di Gravio and R. Patriarca, "Anomalies Detection in Smart Manufacturing Using Machine Learning and Deep Learning Algorithms," in *International Conference on Industrial Engineering and Operations Management*, Rome, italy, 2021.

[4] N. B. Idris and B. Shanmugam, "Artificial intelligence techniques applied to intrusion detection," *INDICON, Annual IEEE,* pp. 52-55, 2005.

[5] Keras, "Keras," [Online]. Available: https://keras.io/. [Accessed 19 January 2021].

[6] L. Ashiku and C. Dagli, "Network Intrusion Detection System using Deep Learning," in *Complex Adaptive Systems Conference Theme: Big Data, IoT, and AI for a Smarter Furture* , Malvern, Pennsylvania, 2021.

[7] S. Kiran, K. Devisetty, P. Kalyan, K. Mukundini and R. Karthi, "Building a Intrusion Detection System for IoT Environment using Machine Learning Techniques," in *Third International Conference on Computing and Network Communications*, 2020.

[8] T. Saranya, S. Sridevi, C. Deisy, T. D. Chung and A. Khan, "Performance Analysis of Machine Learning Algorithms in intrusion Detection System: A Review," *Third International Conference on Computing and Network Communication,* vol. 171, pp. 1251-1260, 2020.

[9] T. Kukuni and B. Kotze, "Industrial Augmented Reality as an Approach for Device Identification within a Manufacturing Plant for Property Alteration Purpose," Bloemfontein, 2021.

[10] T. Subbulakshmi and A. Afroze, "Multiple learning based classifiers using layered approach and feature selection for attack detection," *In Emerging Trends in Computing, Communication and Nanotechnology,* pp. 308-314, 2013.

[11] J. Edith and A. Chandrasekar, "Layered architecture to detec attacks using asymmetric support vector machine," *Applied Security Research ,* vol. 9, pp. 133-149, 2014.

[12] H. Pajouh, R. Javidan, R. Khayami, A. Dehghantanha and K. Choo, "A two-layer dimension reduction and two-tier classification model for anomaly-based intrusion detection in IoT backbone networks," pp. 1-11, 2019.

*e*ISSN1303-5150

[13] T. Salman, D. Bhamare, A. Erbad, R. Jain and M. Samaka, "Machine Learning for Anomaly Detection and Categorization in Multi-Cloud Environments," in *2017 IEEE 4th International Conference on Cyber Security and Cloud Computing*, 2017.

[14] H. Yin and H. Dong, "The problem of noise in classification: past, current and future work," in *IEEE 3rd Internation Conference on Communication Software and Networks*, Xian, China, 2011.

Authors

DrKukuniTshepo Godfrey: Managing Director: Rea Thusa Consulting Engineers, Pretoria, Gauteng, 0087, South Africa.
tgkukuni@gmail.com,0834230412

Tshepo Godfrey Kukuni holds Masters and Doctoral degrees in Electrical Engineering and he has worked both in industry and tertiary education. He is currently employed as an Open Innovation Specialist at The Innovation Hub Management Company (TIHMC) and his also a Director at Rea Thusa Consulting Engineers (PTY) LTD. DrKukuni has published papers in different journals both nationally and internationally. DrKukuni is also currently working on research on Network Intrusion Detection applications, Augmented Reality Applications, Free-Piston Engines, Image Processing, Computer vision, Machine Learning, Design and modelling, and Machine Vision. DrKukuni is also a registered member of the South African Association of PhDs (SAAPhDs) and his also on the Advisory committee on Innovation at Institute for the Study of Legislature in South Africa (ISLSA).

Prof. Elisha Markus

Prof Markus holds a Doctoral degree in Electrical Engineering. His research interests are in Robotic control, smart networks, WSNs, IoTs, machine learning, assistive mobility technologies and electromagnetic fields. He is currently working on cooperative robotic systems in health care and mining applications at the Center for sustainable smart cities at the Faculty of Engineering, Built Environment and Information Technology (FEBIT). He is employed in the Faculty of Engineering, Built Environment and Information Technology at the Central University of Technology, Free State. He is also actively involved in the Centre for Sustainable Smart Cities.

Dr Ben Kotze: Assistant Dean at Central University of Technology, FS Bloemfontein, Free-State, 9300, South Africa:
bkotze@cut.ac.za,

Ben Kotze holds a Masters and Doctoral degrees in Electrical Engineering.
He is professionally registered with the Engineering Council of South Africa (ECSA) and several associations of which the oldest South African Institute of Electrical

Engineers (SAIEE est. 1904) where he is a fellow.

With seven years, industry experience and over thirty years tertiary education experience in Electrical Engineering of which he lectured more than 23 subjects. He is still actively involved with industry and work integrated learning (WIL).

He is currently doing research on vision, several different AGV's, renewable energy systems, simulation and control, augmented reality systems, IoT security, smart farming, and prediction methods. Several undergraduate students, masters and doctoral passed by his mentorship. He attended several international and national conferences in engineering and education and have published in accredited journals on these topics. He is currently the assistant dean teaching and learning in the Faculty of Engineering, Built Environment and Information Technology at the Central University of Technology, Free State. Also actively involved in the Centre for Sustainable Smart Cities.

Prof. Adnan M. Abu-Mahfouz

Adnan M. Abu-Mahfouz ((Senior Member, IEEE) received the M.Eng. and Ph.D. degrees in computer engineering from the University of Pretoria, Pretoria, South Africa, in 2005 and 2011, respectively. He is currently a Chief Researcher and the Centre Manager of the Emerging Digital Technologies for 4IR (EDT4IR) Research Centre, Council for Scientific and Industrial Research, Pretoria; an Extraordinary Professor with University of Pretoria; a Professor Extraordinaire with the Tshwane University of Technology, Pretoria; and a Visiting Professor with the University of Johannesburg, Johannesburg, South Africa. His research interests are wireless sensor and actuator network, low power wide area networks, software-defined wireless sensor network, cognitive radio, network security, network management, and sensor/actuator node development.

Prof Abu-Mahfouz is a Section Editor-in-Chief with the Journal of Sensor and Actuator Networks, an Associate Editor at IEEE INTERNET OF THINGS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON CYBERNETICS, IEEE ACCESS and FRONTIERS IN PLANT SCIENCE, and a member of many IEEE technical communities